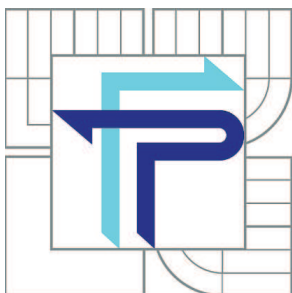


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA PODNIKATELSKÁ
ÚSTAV INFORMATIKY

FACULTY OF BUSINESS AND MANAGEMENT
INSTITUTE OF INFORMATICS

ANALÝZA CHOVÁNÍ UŽIVATELŮ INTERNETOVÝCH SOCIÁLNÍCH SÍTÍ Z HLEDISKA BEZPEČNOSTI

ANALYSIS OF SOCIAL NETWORK USERS' BEHAVIOUR FROM THE SECURITY ASPECT

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

MAREK WALDER

VEDOUCÍ PRÁCE

SUPERVISOR

doc. RNDr. JIŘÍ KROPÁČ, CSc.

BRNO 2010

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Walder Marek

Manažerská informatika (6209R021)

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách, Studijním a zkušebním řádem VUT v Brně a Směrnicí děkana pro realizaci bakalářských a magisterských studijních programů zadává bakalářskou práci s názvem:

Analýza chování uživatelů internetových sociálních sítí z hlediska bezpečnosti

v anglickém jazyce:

Analysis of Social Network Users' Behaviour from the Security Aspect

Pokyny pro vypracování:

Úvod

Vymezení problému a cíle práce

Teoretická východiska práce

Analýza problému a současné situace

Vlastní návrhy řešení, přínos návrhů řešení

Závěr

Seznam použité literatury

Přílohy

Seznam odborné literatury:

KROPÁČ, J. Statistika B. 2. vyd. Brno : Fakulta podnikatelská, 2009. ISBN 978-80-214-3295-6.

CIPRA, T. Analýza časových řad s aplikacemi v ekonomii. 1. vyd. Praha : SNTL/ALFA, 1986.

CYHELSKÝ, L., aj. Základy teorie statistiky pro ekonomy. 1. vyd. Praha : SNTL/ALFA, 1979.

SEGER, J., aj. Statistika v hospodářství. 1. vyd. Praha : ETC Publishing, 1998. ISBN 80-86006-56-5.

Vedoucí bakalářské práce: doc. RNDr. Jiří Kropáč, CSc.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2009/2010.

L.S.

Ing. Jiří Kříž, Ph.D.
Ředitel ústavu

doc. RNDr. Anna Putnová, Ph.D., MBA

V Brně, dne 28.05.2010

Abstrakt

Předmětem bakalářské práce je statistická analýza chování uživatelů internetových sociálních sítí z hlediska bezpečnosti na základě vytvořeného dotazníku. První část bakalářské práce se zabývá teoretickými východisky z oboru statistiky, sociologického výzkumu a bezpečnostních zásad na internetu. Druhá část praktickým provedením, a to jak sociologického řešení a vytvoření dotazníku, tak statistickým zpracováním nasbíraných dat, včetně příslušných závěrů a doporučení.

Abstract

Theme of my bachelor's thesis is statistic analysis of social network users' behaviour from the security aspect, based on self created question-form. First part of this bachelor's thesis deals with theoretical resources of statistics, sociological research and security internet principles. Second part contains applicative execution of sociological solutions, question-form creation and statistic processing of gathered fact, including relevant conclusions and recommendations.

Klíčová slova

Statistika, statistický analýza, sociologický výzkum, internetový dotazník, sociální sítě, bezpečnost, internet

Key words

Statistics, statistic analysis, sociological research, internet question-form, social networks, security, internet

Bibliografická citace

WALDER, M. *Analýza chování uživatelů internetových sociálních sítí z hlediska bezpečnosti*. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, 2010. 65 s.

Vedoucí bakalářské práce doc. RNDr. Jiří Kropáč, CSc.

Čestné prohlášení

Prohlašuji, že předložená diplomová práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu Zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 31. května 2010

.....

Poděkování

Tímto bych chtěl poděkovat svému vedoucímu práce, panu doc. RNDr. Jiřímu Kropáčovi, CSc za odborné vedení a také Mgr. Kamilu Malinkovi za konzultaci a cenné rady.

Obsah

Úvod.....	9
Cíle a postupy	9
1 Teoretická východiska	11
1.1 Statistika.....	11
1.1.1 Základní soubor, statistická jednotka, statistický znak.....	11
1.1.2 Výběrový soubor.....	11
1.1.3 Datový soubor	12
1.1.4 Základní charakteristiky datových souborů.....	12
1.1.5 Intervalové odhady parametrů	13
1.1.6 Testování statistických hypotéz	14
1.1.7 Dvourozměrné datové soubory kvantitativních znaků.	15
1.1.8 Dvourozměrné datové soubory kvalitativních znaků.	16
1.2 Sociologický výzkum	18
1.2.1 Typy výzkumů	19
1.2.2 Etapy sociologického výzkumu	20
1.2.3 Metody sociologického výzkumu	22
1.2.4 Techniky sociologického výzkumu	23
1.2.5 Otázky; druhy a zásady.....	23
1.2.6 Technika dotazníku.....	25
1.2.7 Velikost výběrového souboru, neboli vzorku.....	26
1.2.8 Zkreslení korelace.....	26
1.3 Internetové sociální sítě a bezpečné chování na internetu.....	27
1.3.1 Historie a současnost internetových sociálních sítí	27
1.3.2 Problematika bezpečného chování na internetu.....	28
1.3.3 Sdílení dat prostřednictvím internetu.....	29
1.3.4 Chyby nejen českých uživatelů, ohrožené skupiny	30
2 Praktická část	32
2.1 Formulace hypotéz.....	32
2.2 Rozhodnutí o populaci a vzorku	33
2.3 Pilotní studie	33

2.4	Rozhodnutí o technice a sběru informací	33
2.5	Konstrukce dotazníku	33
2.6	Předvýzkum	36
2.7	Terénní šetření:	36
2.8	Statistické zpracování	37
2.8.1	Autentizace - základní charakteristiky	37
2.8.2	Autentizace - Tabulky a grafy:	39
2.8.3	Dostupnost dat a důvěřivost uživatelů – charakteristiky	41
2.8.4	Dostupnost dat a důvěřivost uživatelů - tabulky a grafy	42
2.8.5	Citlivost údajů – charakteristiky	43
2.8.6	Citlivost údajů - Tabulky a grafy:	44
2.8.7	Osobní údaje - charakteristiky	45
2.8.8	Osobní-údaje - tabulky a grafy	46
2.8.9	Intervalové odhady věkových charakteristik	47
2.8.10	Testy hypotéz:	47
2.8.11	Motivační otázky a znaky mimo kategorie:	54
3	Zhodnocení	57
3.1	Shrnutí nejvýraznějších charakteristik	57
3.1.1	Autentizace	57
3.1.2	Dostupnost dat a důvěřivost uživatelů	58
3.1.3	Citlivost údajů	58
3.1.4	Osobní údaje	58
3.1.5	Ostatní	58
	Závěr a návrhy	59
	Použitá literatura	61
	Seznam obrázků a grafů	62
	Seznam tabulek	63
	Přílohy	65

Úvod

Sociální sítě zažily v poledních dvou letech v České Republice obrovský rozmach. Tedy alespoň ty zahraniční. Před světovými velikány, jako jsou dnes Twitter nebo Facebook, byly doménami českých sociálních sítí servery Spolužáci.cz, Libimseti.cz nebo projekt společnosti seznam Lidé.cz. České sociální sítě možná ztratily s (očekávaným) příchodem zahraničních ledoborců značnou část uživatelů, nicméně rozvířené vody českého internetu brázdí stále.

Prudký vývoj, který započal v roce 2004 příchodem nové vývojové etapy webových aplikací, tzv. Web 2.0, na sebe rozhodně nenechal dlouho čekat a obzvláště v posledních době okouší uživatelé nebyvalou záplavu nových technologií. Možná, že už někoho unavuje neustále omílaná fráze prudkého vývoje technologií a internetu, ale jako skutečnost ji nemůže opomíjet. Existenci a vývoj sociálních sítí typu Facebook (nebo, nedej Bože, samotného internetu jakožto sociální sítě největší) nelze ignorovat.

Pozornost všech (nejen) ekonomických subjektů se zaměřila na tyto přístavy plné uživatelů, jejich dat a nových možností týkajících se komunikace, šíření informací, sdílení dat a zábavy. Avšak tyto přístavy poskytují také nové možnosti v oblastech marketingu, personalistiky a v neposlední řadě i v oblasti bezpečnosti informačních technologií.

Právem tyčí varovné prsty všichni ti, kterým nejsou lhostejné tak velké koncentrace dat. Otázky týkající se bezpečnosti sociálních sítí netrápí dnes málokoho, a to hlavně bezpečnosti ze strany uživatelů.

Jak silná používají zabezpečení? Jak citlivá data ve svých profilech sdílí? Jsou si uživatelé sociálních sítí vůbec vědomi možných hrozeb? Právě proto jsem se rozhodl vypracovat sociologický průzkum uživatelů sociálních sítí z hlediska bezpečnosti jejich chování, ve kterém se pokusím zjistit do jaké míry jsou, či nejsou ohroženi a zda existují skupiny uživatelů samy sobě nebezpečné více než skupiny jiné.

Cíle a postupy

Cílem mé práce bude sociologický průzkum, který má ukázat do jaké míry jsou uživatelé ohroženi, a to ze tří hledisek: hlediska autentizace, z hlediska důvěřivosti

uživatelů (dostupnosti jejich dat) a z hlediska citlivosti údajů sdílených na sociálních sítích.

Práce bude definovat teoretická východiska zpracovávat tři oblasti: statistiky jakožto nástroje zpracování dat, sociologického průzkumu jakožto nástroje pro získání těchto dat a v nespolední řadě z oblasti sociálních sítí a internetové bezpečnosti. Zdroj všech grafů a tabulek je vlastní.

1 Teoretická východiska

1.1 Statistika

Statistika se obvykle rozděluje na statistiku *deskriptivní* a statistiku *induktivní*. Deskriptivní statistika, je statistika historicky starší a zabývá se sběrem a zpracováním hodnot statistických znaků, popisuje tedy vybraná data. Induktivní statistika rozšiřuje získané závěry z datového souboru na základní soubor, z něhož byl datový soubor vybrán, vyvozuje tedy závěry o vlastnostech celku ze kterého byla původně data vybrána. (CHROBOČEK, 2009, s. 13)

1.1.1 Základní soubor, statistická jednotka, statistický znak

Základní soubor, někdy také *populace*, je množina prvků, které podrobujeme statistickému zkoumání. Jedná se o soubor objektů o nichž nemáme úplné informace, a které tedy chceme poznat. Jednotlivé prvky základního souboru se pak nazývají statistickými jednotkami. Typickým rysem základního souboru je množina vlastností, které jsou společné všem prvkům základního souboru. *Statistickým znakem* pak nazýváme vyjádření určité vyšetřované vlastnosti statistické jednotky.

Statistické znaky dělíme na znaky *kvantitativní* a *kvalitativní*.

Kvantitativní znaky nabývají číselných hodnot a dělí se dále na *diskrétní*, nabývající oddělených číselných hodnot (např. počet dětí, počet reklamací) a *spojité*, nabývající všech hodnot z intervalu (např. teplota, váha).

Kvalitativní znaky lze vyjádřit slovně a dělí se dále na znak *ordinální* a znak *nominální*. Ordinální znaky můžeme s jistotou uspořádat, např. je seřadit od nejmenšího po největší nebo dle významnosti. Dají se tedy vyjádřit i pomocí čísel. Podle počtu variant kterých může kvalitativní znak nabýt je dělíme ještě na znaky *alternativní* - nabývající pouze dvou variant (např. odpovědi na otázky typu ano, ne) a znaky *množné* - nabývající více než dvou variant. (KROPÁČ, 2009, s. 1-5), (NOVOVIČOVÁ, 1999, s. 10)

1.1.2 Výběrový soubor

Jako *výběrovým soubor* označujeme vybranou podmnožinu prvků (statistických jednotek) základního souboru.

Způsoby, jakými je možné vybrat prvky základního souboru a vytvořit tak soubor výběrový jsou dva:

Náhodný výběr - Náhodným výběrem chápeme takový výběr, kdy může být každý prvek základního souboru vybrán se stejnou pravděpodobností. Tento způsob výběru nazýváme také jako tzv. *nezávislý výběr*.

Reprezentativní výběr – Při reprezentativním výběru se výběrový soubor podobá souboru základnímu, vystihuje nejlépe jeho charakteristiky. Takový výběrový soubor může podávat mnohem lepší informaci o zkoumaných vlastnostech základního souboru, ale na druhou stranu může dojít i k situaci opačné, kdy je tento výběr proveden úmyslně k oklamání zadavatele statistického šetření nebo k oklamání veřejnosti. Takový výběr samozřejmě ztrácí svou reprezentativnost. (KROPÁČ, 2009, s. 1-5) (CHROBOČEK, 2009, s. 18)

1.1.3 Datový soubor

Zkoumáme-li statistický znak na prvcích výběrového souboru, pak množina hodnot tohoto znaku je nazývána datovým souborem.

Počet těchto hodnot se nazývá rozsah datového souboru.

Zkoumáme-li u každé statistické jednotky pouze jediný statistický znak, jedná se o *jednorozměrný datový soubor*. V případě, že zkoumáme u každé statistické jednotky dva nebo více statistických znaků hovoříme o *vícerozměrném datovém souboru*. (KROPÁČ, 2009, s. 5-19)

1.1.4 Základní charakteristiky datových souborů

Empirickými charakteristikami kvantitativních znaků jsou výběrový průměr, výběrový rozptyl a výběrová směrodatná odchylka

Výběrový průměr:

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

Výběrový rozptyl:

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

Výběrová směrodatnou odchylku pak vypočteme jako odmocninu z výběrového rozptylu.

Dalšími významnými charakteristikami jsou tzv. kvantily. Kvantil je číslo na ose reálných čísel, které označujeme \tilde{x}_p . Kvantil rozděluje datový soubor na dvě části. $100p\%$ hodnot datového souboru je menších nebo rovno číslu \tilde{x}_p a zbytek, tedy $100(1-p)\%$ hodnot datového souboru je větších než číslo \tilde{x}_p .

Kvantily mají velký význam pro vypovídající hodnotu statistického šetření. Mnohdy se ve veřejných statistikách setkáváme pouze s výběrovým průměrem, který má z pohledu laika největší vypovídající hodnotu. Problém tkví v citlivosti výběrového průměru na extrémně vysoké nebo nízké hodnoty, který pak podává zkreslenou představu o datovém souboru. Z tohoto pohledu zastupuje nejvýznamnější úlohu pro zlepšení vypovídající hodnoty tzv. *medián*, což je kvantil $\tilde{x}_{0,5}$, který datový soubor rozděluje na dvě stejné poloviny, kde je 50% hodnot menších nebo rovno mediánu a 50% je vyšších než medián. V případě, že má *uspořádaný datový soubor* lichý počet prvků, je hodnota mediánu rovna hodnotě prostředního prvku, avšak v případě, kdy má datový soubor sudý počet prvků, medián vypočteme jako aritmetický průměr hodnot posledního (nejvyššího) prvku z dolní poloviny (nižších) hodnot a prvního (nejnižšího) prvku z horní poloviny (vyšších) hodnot.

Dalšími kvantily se kterými se můžeme setkat je *dolní (první)* a *horní (třetí)* kvartil $\tilde{x}_{0,25}$ a $\tilde{x}_{0,75}$. (KROPÁČ, 2009, s. 5-19), (CHROBOČEK, 2009, s. 17-18),

1.1.5 Intervalové odhady parametrů

Intervalové odhady vzniknou následujícím postupem: z datového souboru $x = (x_1, x_2, \dots, x_n)$ určíme interval ohraničený čísly $b_D = f_D(x)$ a $b_H = f_H(x)$, kde je b_D levou hranicí a b_H pravou hranicí. Tento interval pokrývá daný parametr s určitou přesností, resp. spolehlivostí, ta je dána *koeficientem spolehlivosti* $1-\alpha$ a interval (b_D, b_H) je *dvoustranným* $100(1-\alpha)\%$ -ním *intervalem spolehlivosti*. Interval (b_D, b_H) tedy pokrývá daný parametr s $100(1-\alpha)$ procentní spolehlivostí. Nejčastěji se používá 95% interval spolehlivosti, $\alpha = 0,05$.

Výpočet intervalů spolehlivosti se s různými druhy pravděpodobnostního rozdělení liší.

Dvoustranný $100(1-\alpha)\%$ -ní interval spolehlivosti pro μ normálního rozdělení

$$N(\mu, \sigma^2) \text{ je roven: } \left(\bar{x} - t_{1-\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}} \right) \quad (3)$$

Kde $t_{1-\frac{\alpha}{2}}(n-1)$ je $100(1-\frac{\alpha}{2})\%$ -ní kvantil tzv. Studentova rozdělení.

Dvoustranný $100(1-\alpha)\%$ -ní interval spolehlivosti pro σ^2 normálního rozdělení

$$N(\mu, \sigma^2) \text{ je roven: } \left(\frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}; \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}^2(n-1)} \right), \quad (4)$$

Kde $\chi_{1-\frac{\alpha}{2}}^2(n-1)$ a $\chi_{\frac{\alpha}{2}}^2(n-1)$ jsou kvantily Pearsonova rozdělení. Dvoustranný

$100(1-\alpha)\%$ -ní interval spolehlivosti pro směrodatnou odchylku σ normálního rozdělení $N(\mu, \sigma^2)$ získáme odmocněním hranic dvoustranného $100(1-\alpha)\%$ -ního intervalu spolehlivosti pro σ^2 .

(KROPÁČ, 2009, s. 21-28), (NOVOVIČOVÁ, 1999, s. 77-92)

1.1.6 Testování statistických hypotéz

Podmínkami pro užití testů jsou:

- Měřený znak je náhodnou veličinou normálním rozdělením $N(\mu, \sigma^2)$, kde hodnoty μ a σ^2 neznáme.
- Ze základního souboru vytvořeného datového souboru vypočteme charakteristiky \bar{x} a s .
- Test posuzuje vztah mezi zvoleným číslem a neznámou hodnotou parametru.

Při testování statistické hypotézy postupujeme tak, že nejprve vyřkneme tzv. *nulovou hypotézu* H_0 a k ní *alternativní hypotézu* H_1 .

Pro testování hypotézy H_0 použijeme náhodnou veličinu, jenž je funkcí náhodného výběru X , tu nazýváme *testovým kritériem* a značíme G . Vypočítáme její realizovanou hodnotu g z datového souboru. Pomocí čísla α , nazývaného *hladina významnosti*, určíme tzv. *kritický obor* W_α . Kritický obor, je obor hodnot, ve kterém se, v případě že hypotéza H_0 platí, realizuje nejvýše $100\alpha\%$ hodnot testového kritéria G . Pokud je

realizovaná hodnota g z oboru hodnot kritického oboru, tj. $g \in W_\alpha$, nulovou hypotézu H_0 zamítáme a přijmeme alternativní hypotézu H_1 . Pokud hodnota g není prvkem kritického oboru, přijmeme nulovou hypotézu. Přijmutí hypotézy není důkazem její pravdivosti, znamená to pouze, že nebyla vyvrácena. V této souvislosti se setkáváme s termínem chyba prvního a druhého druhu. Chyba první druhu nastává, je-li hypotéza zamítnuta, přestože platí. Chyba druhého druhu nastává, není-li hypotéza zamítnuta, přestože platí. V této souvislosti je důležité při interpretaci výsledků nezapomenout na jaké hladině významnosti, tedy s jakou pravděpodobností chyby prvního druhu, výsledky podáváme. (KROPÁČ, 2009, s. 29-46), (NOVOVIČOVÁ, 1999, s. 95-118)

1.1.7 Dvourozměrné datové soubory kvantitativních znaků.

Dvourozměrný datový soubor je dán n dvojicemi hodnot (x_i, y_i) , $i=1,2,\dots,n$, kvantitativního typu, kde nazýváme *rozsahem datového souboru*. U dvourozměrných datových souborů můžeme samozřejmě také vypočítat jejich základní charakteristiky jsou výběrové průměry nebo výběrové rozptyly.

Stejně jako u jednorozměrných datových souborů získáme *výběrové směrodatné odchylky* odmocněním výběrových rozptylů.

Nejdůležitější vlastností dvourozměrných datových souborů, je samozřejmě vztah, závislost, resp. rozdělení měřených znaků. Existuje-li mezi dvěma měřenými znaky nějaká funkční závislost, ať již silnější, nebo slabší, označujeme tyto znaky za *korelované*. Nejjednodušším způsobem, jak získat představu o tom, zda mají znaky mezi sebou nějaký silnější nebo slabší vztah, je vytvoření *dvourozměrného bodového grafu*, neboli *korelačního diagramu*.

Korelaci mezi složkami náhodného vektoru (výběru) (X,Y) můžeme v datovém souboru měřit pomocí *výběrové kovariance* a *výběrového koeficientu korelace*.

$$\text{Výběrová kovariance:} \quad C_{XY} = \frac{1}{n-1} \left[\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right] \quad (5)$$

$$\text{Koeficient korelace:} \quad r_{XY} = \frac{C_{XY}}{s_X s_Y} \quad (6)$$

Náhodné veličiny X,Y jsou nekorelované, není mezi nimi lineární závislost, je-li výběrová kovariance rovná nule. Náhodné veličiny X,Y jsou korelované, existuje mezi nimi lineární závislost, je-li výběrová kovariance různá od nuly. Z toho vidíme, že

výběrová kovariace nám nevyjádří velikost vzájemné lineární vazby. K tomu slouží výběrový koeficient korelace, jehož vlastnosti jsou: $r_{XY} = r_{YX}$ a $|r_{XY}| \leq 1$. Náhodné veličiny X a Y jsou nekorelované, je-li koeficient korelace roven nule, je-li vyšší nuly, jsou náhodné veličiny X, Y kladně korelovány, je-li nižší nuly, jsou záporně korelovány.

Závislost vyjádříme slovně pomocí absolutní hodnoty koeficientu korelace a to:

Jestliže se $|r_{XY}|$ blíží nule, říkáme, že závislost mezi náhodnými veličinami X,Y je *velmi slabá*. Jestliže se $|r_{XY}|$ blíží jedné polovině, říkáme, že závislost mezi náhodnými veličinami X,Y je *průměrná*. Jestliže se $|r_{XY}|$ blíží jedné, říkáme, že závislost mezi náhodnými veličinami X,Y je *velmi silná*. Korelaci nemusíme znázorňovat pouze korelačními diagramy, ale třeba i korelační tabulkou, znázorňující vztah mezi náhodnými veličinami X,Y obdobně jako diagram, avšak včetně hodnot a jejich četností.

Koeficient korelace r_{XY} , resp. korelace ρ nám o znacích výběrového souboru, resp. základního souboru, v případě že je roven nule, neříká, že jsou znaky stoprocentně nezávislé, nýbrž nezávislé s určitou pravděpodobností, tzv. *stochasticky nezávislé*. Pro testování nezávislosti dvou kvantitativních znaků používáme testování hypotézy se

$$\text{statistikou } T = \frac{r_{XY} \sqrt{n-2}}{\sqrt{1-r_{XY}^2}} \quad (7)$$

, která má Studentovo rozdělení o $n-2$ stupních volnosti a $n \geq 3$.

Kritické obory a hypotézy pro testování nezávislosti dvou kvantitativních znaků uvádí příslušná literatura, využívá se statistika se Studentovým rozdělením.

Vzhledem k použité statistice se Studentovým rozdělením, jsou samozřejmě $t_{1-\frac{\alpha}{2}}(n-2)$

a $t_{1-\alpha}(n-2)$ kvantily Studentova rozdělení, viz. kapitola 1.1.6 Testování statistických hypotéz. (KROPÁČ, 2009, s. 54-64)

1.1.8 Dvourozměrné datové soubory kvalitativních znaků.

Dvourozměrný datový soubor je dán n dvojicemi hodnot (x_i, y_i) , $i=1,2,\dots,n$, kvalitativního typu, kde n nazýváme rozsahem datového souboru. Znaky kvalitativního

typu nemůžeme charakterizovat bodovými charakteristikami jako je rozptyl nebo směrodatná odchylka. Můžeme je ale vyřadit dle jejich variant, tomuto procesu se říká také *kategorizace* a proto tyto znaky někdy také nazýváme *kategoriální*. Nejznámějším způsobem jakým můžeme hodnotit znaky kvalitativního typu je tzv. *kontingenční tabulka*. Kontingenční tabulka obsahuje v záhlaví varianty prvního znaku a v prvním sloupci varianty druhého znaku, jejím obsahem jsou tzv. *simultánní četnosti* n_{ij} , které značí kolikrát nastala varianta X_i znaku X a zároveň varianta Y_j znaku Y . V posledním sloupci, resp. řádku jsou sloupcové $n_{\bullet j}$, resp. řádkové $n_{i\bullet}$ součty j -tých variant znaku Y , resp. i -tých variant znaku X , v pravém dolním rohu je tedy celkový počet pokusů, variant nebo četností n . Tyto součty nazýváme *marginálními četnostmi*.

X\Y	Y_1	Y_2	...	Y_s	$n_{i\bullet}$
X_1	n_{11}	n_{12}	...	n_{1s}	$n_{1\bullet}$
X_2	n_{21}	n_{22}	...	n_{2s}	$n_{2\bullet}$
...
X_r	n_{r1}	n_{r2}	...	n_{rs}	$n_{r\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$...	$n_{\bullet s}$	n

Tabulka 1; Kontingenční tabulka

Simultánní nebo marginální četnosti vydělené počtem opakování pokusu n nazýváme simultánními nebo marginálními pravděpodobnostmi:

Marginální pravděpodobnost:
$$\hat{p}_{i\bullet} = \frac{n_{i\bullet}}{n}; \hat{p}_{\bullet j} = \frac{n_{\bullet j}}{n} \quad (8)$$

Simultánní pravděpodobnost:
$$\hat{p}_{ij} = \frac{n_{ij}}{n} \quad (9)$$

I u kvalitativních znaků má význam hodnotit jejich stochastickou závislost. A testovat nezávislost dvou kvalitativních znaků. Podmínkami pro testování nezávislosti dvou kvalitativních znaků jsou:

- Musejí být měřeny dva kvalitativní znaky.
- Nejdříve testujeme hypotézu, zda jsou tyto znaky na sobě stochasticky závislé.
- Pokud je hypotéza přijata, testujeme sílu jejich závislosti

Statistika testového kritéria hypotézy existence závislosti mezi dvěma kvalitativními znaky má Pearsonovo rozdělení. A vypočteme ji jako

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}} \quad (10)$$

Jelikož tímto testem opět nezjistíme velikost závislosti, je vhodné uvést tzv. *Cramérův*

koeficient kontingence

$$V = \sqrt{\frac{\chi^2}{n(m-1)}} \quad (11)$$

, kde za m dosadíme menší z čísel r a s . Při úplné nezávislosti je Cramérův koeficient kontingence roven nule, v případě úplné závislosti jedné.

Speciálním případem kontingenční tabulky je stav, kdy jsou znaky X a Y *alternativního typu* viz. 1.1.1 Základní soubor, statistická jednotka, statistický znak. Tabulku pak nazýváme čtyřpolní tabulkou a doporučený výpočet statistiky testového kritéria, při $n > 40$ je

$$\chi^2 = n \frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\bullet}n_{\bullet 1}n_{2\bullet}n_{\bullet 2}} \quad (12)$$

Pokud je $n < 40$ doporučuje se Fischerův (exaktní) faktoriální test.

(KROPÁČ, 2009, s. 64-73), (PECÁKOVÁ, 2007, s. 61-65)

1.2 Sociologický výzkum

Sociologie je věda sociální. Předmětem sociologie jsou *sociální jevy*, tzn. veškeré jevy mezi lidmi. Sociologie je věda o zákonitostech, tendencích a pravidelnostech vývoje a pohybu sociálních systémů. (NOVOTNÁ, 2008, s. 7)

Sociologii dělíme na sociologii teoretickou a sociologii empirickou. Součástí empirické sociologie je právě sociologický výzkum. Každý empirický výzkum musí být založen na šesti předpokladech:

1. Výzkumník se nesmí nechat ovlivnit subjektivními postoji. – Objektivnost
2. Výzkum má respektovat proměnlivost jevů a změny v jejich vývoji – Dynamičnost jevů
3. Zkoumané jevy jsou součástí vyššího celku. – Relativní autonomnost jevů
4. Odhalování vnitřních vztahů mezi jevy a jejich vzájemné ovlivňování – Vnitřní rozpornost

5. Respektování složitosti sociálních jevů. – Složitost
6. Uvědomění si existence celé řady přímých i nepřímých vlivů – Vzájemná podmíněnost[10] (ZICH, 1976)[2] (JEŘÁBEK, 1993)

Někteří autoři k těmto předpokladům přidávají i další (ČIHOVSKÝ 2006, s. 5-6), k serióznímu empirickému šetření patří :

7. Stanovení teoretického východiska (paradigmatu)
8. Vědomí možnosti multiparadigmatického přístupu, tzn. vše může mít více podstat a principů.
9. Určitá míra gnoseologického skepticismu při prezentaci výsledků.
10. Komplementární přístup, nikoliv komplexní.

1.2.1 Typy výzkumů

Typologie výzkumů závisí, jako každá typologie, na úhlu pohledu, proto typů výzkumů může být mnoho. V zásadě rozdělujeme sociologické výzkumy do dvou zásadních skupin, a to kvantitativní a kvalitativní.

Kvantitativní výzkum je numerické šetření, podávající informace o určitém počtu zkoumaných prvků, může nám říci jak zobecnitelné jsou výsledky výzkumu, nebo jaká je pravděpodobnost omylu. Jeho cílem je potvrdit, nebo vyvrátit naše hypotézy.

Kvalitativním výzkumem rozumíme nenumerické šetření a interpretaci sociální reality, jeho cílem je odkrýt význam podkládaný sdělovaným informacím.(DISMAN, 2009, s. 284-287)

Čímž se rozumí ta skutečnost, kterou neumíme vyjádřit v kvantitativním výzkumu – tj. lidský pohled na věc; kvantitativní výzkum nám může odpovědět „jak moc“, resp. „kolik“, ale neumí říci, co to pro zkoumaný vzorek znamená ze sociologického hlediska. Kvalitativní výzkum slouží k vytváření teorií a hypotéz. Podává nám velké množství informací o zpravidla malé skupině jedinců, kdežto kvantitativní výzkum nám podává omezené množství informací o velké skupině jedinců. Kvalitativní výzkum často předchází samotnému kvantitativnímu výzkumu, bývá považován za součást předvýzkumu, přestože jsou jeho výsledky naprosto nezávislé. Pro kvantitativní výzkum je důležité, že nám kvalitativní výzkum dává své poznatky – informace (data) o zkoumaném jevu, popisuje nalezené vazby, hledá struktury, vytváří teorie a hypotézy tolik důležité pro kvantitativní výzkum, který umí tato data zvalidovat.

Dále předkládám příklady některých dalších možných pohledů na typologii sociologických výzkumů :

Podle cíle

- orientační
- popisný
- vysvětlující

Dle frekvence opakování

- jednorázový
- opakovaný
- dlouhodobý

Podle zaměření výsledků

- základní
- aplikovaný
- operativní

Dále můžeme typy výzkumů dělit z hlediska rozsahu zkoumaného vzorku, měřitelnosti nebo komplexnosti. Pohledy na časové dělení a i jiné pohledy se často u různých autorů liší. Z časového hlediska jsou výzkumy (REICHEL, 2009) děleny na krátkodobé, střednědobé, dlouhodobé a longitudinální (extrémně dlouhodobé). Mezi základní pohledy patří i hledisko tématického záběru, kde rozlišujeme výzkumy extenzivní (velké množství zkoumaných problémů, jevů, nezabírá se jimi příliš do hloubky) a intenzivní (méně objektů, zaměřen na konkrétní sociální jev, zkoumá jej do hloubky). (ČIHOVSKÝ 2006, s. 8)

1.2.2 Etapy sociologického výzkumu

1. Přípravná fáze: je etapa ve které je veškerá pozornost věnována studiu dané problematiky, hledání literárních i elektronických zdrojů, zkoumání výsledků dřívějších šetření, popř. konzultacím se zadavatelem. Je to etapa časově nejnáročnější a taky nejvýznamnější, neboť chyby, kterých bychom se v této fázi dopustili, nelze v průběhu šetření napravit. Součástí této fáze bývá někdy i tzv. *předvýzkum*. Výstupem přípravné fáze by měl být *zadávací projekt*.
2. Terénní šetření: nebo-li sběr informací, je sběr dat a informací pomocí výzkumných technik (viz. 1.2.4 *Techniky sociologického výzkumu*).

Součástí terénního šetření je rovněž kontrola a eliminace nepoužitelných zdrojů informací.

3. Statistické zpracování: Aplikování matematické statistiky k dalšímu třídění a popisování jevů a vlastností. Statistika je mocným nástrojem sociologických výzkumů, pomáhá nejen zpracovávat data, ale zejména umožňuje interpretovat výsledky.
4. Interpretace výsledků: znamená charakterizování zkoumaného vzorku, vysvětlení jednotlivých otázek, zodpovězení hypotéz a vyvození závěrů a příslušných doporučení.
5. Závěrečná etapa: Zde patří obhajoba práce, její předání a případná diskuse. (ČIHOVSKÝ 2006, s. 9), (JEŘÁBEK, 1993)

Postup výzkumu lze rozepsat do těchto praktických kroků (DISMAN, 2009, s. 120):

1. Formulace (teoretického nebo praktického) sociálního problému,
2. Formulace teoretické hypotézy
3. Formulace pracovních hypotéz
4. Rozhodnutí o populaci a vzorku
5. Pilotní studie
6. Rozhodnutí o technice a sběru informací
7. Konstrukce nástrojů pro tento sběr
8. Předvýzkum
9. Sběr dat
10. Rozbor dat
11. Interpretace, závěry, teoretická zobecnění

Prvních osm bodů je de-facto detailnější rozepsání přípravné fáze, neboť přípravná fáze je právem ta nejdůležitější a leckdy i nejnáročnější ze všech etap sociologického výzkumu.

Pilotní studie v přípravné fázi má za cíl zjistit, zda informace, které budeme požadovat od zkoumaného vzorku, nám je schopen vzorek poskytnout a tedy, zda jsme vůbec cílenou problematiku schopni zkoumat. Je prováděna na malém vzorku a má nejčastěji podobu kvalitativního výzkumu v podobě nestandardizovaného rozhovoru.

Předvýzkum se od pilotní studie liší obvykle jen o něco málo větším vzorkem, jeho cílem je testovat nástroje, které jsme pro výzkum zvolili tzn. zejména testování srozumitelnosti a jednoznačnosti otázek.

1.2.3 Metody sociologického výzkumu

Metodami sociologického výzkumu rozumíme uspořádání a způsob organizace výzkumu. Termín *metody* bývá někdy nahrazen termínem *procedury*.

Při sociologickém zkoumání nám žádná metoda neumožňuje úplnou jistotu a spolehlivost, stejně jako vypovídající hodnota induktivních statistických šetření nemůže být nikdy stoprocentní. Předpokladem pro dosažení co nejvyšší spolehlivosti informací a nejkvalitnějších výsledků je volba vhodné metody.

Metody sociologického výzkumu lze provádět dvojím způsobem, a to jako metodu *terénního šetření*, která spočívá v získávání dat a informací z terénu. Příkladem může být získávání informací *anketěrem* (tazatelem) od skupiny respondentů

Druhá metoda se nazývá *metodou introspektivní*, ta se oproti metodě terénního šetření zakládá na zkušenostech výzkumníka, jeho znalostech prostředí, popř. praxi v nějakém společenském jevu. Z toho je jasné, že většina konkrétních šetření, probíhá v obou úrovních, tzn. výzkum je prováděn samotným terénním šetřením, zejména v příhodně pojmenované fázi terénního šetření, ale také je ovlivňován zkušenostmi výzkumníka, které nabral především v přípravné fázi.

Konkrétní metody sociologického výzkumu pak jsou:

Experimentální metoda pracuje s tzv. *experimentálním (kontrolním) souborem*, jenž je podrobován různým změnám situací. Tato Metoda je však ze všech nejsložitější, neboť lze jen velmi obtížně navodit odpovídající laboratorní podmínky bez nežádoucích vnějších vlivů.

Monografická metoda zkoumá podrobně pouze jeden jediný případ, jev. Veškeré závěry vynáší právě jen a pouze pro daný jev nebo případ, nelze je zobecňovat.

Historická metoda zkoumá vývoj sledovaného jevu, jeho historii. Popisuje nejen samotný vývoj, ale i změny, které ve vývoji nastaly a zákonitosti těchto změn. Na základě této metody můžeme pak vynášet prognózy, nebo předpovídat trendy dalšího krátkodobého vývoje.

Typologická, neboli srovnávací metoda, hledá a sleduje původ odlišností na základně srovnávání vlastností jednotlivých případů. V tomto případě musí jít o konkrétní a

specifickou situaci, ve které lze metodu aplikovat, neboť je vzhledem k proměnlivosti sociálního prostředí časově omezená.

Statistická metoda umožňuje jako jediná z metod pracovat s velkým počtem případů, zbylé metody pracují s malým počtem procesů, jevů a objektů zkoumání, (monografická dokonce pouze s jedním). Matematická statistika slouží právě v situacích, kdy je základní soubor nekonečný nebo příliš velký. Zkoumání základního souboru induktivními statistickými metodami skrze soubor výběrový je navíc mnohem levnější než výzkumu podrobovat celou populaci (základní soubor) viz. kapitola *1.1 Statistika*. (ČIHOVSKÝ 2006, s. 9-11), (JEŘÁBEK, 1993)

1.2.4 Techniky sociologického výzkumu

Termínem technik sociologického výzkumu chápeme způsob, jakým jsou získávány informace. Metodami si tedy výzkumník zvolí rámec organizace a uspořádání, techniky jsou nástrojem pro samotnou realizaci.

Výběr vhodné techniky je součástí přípravné fáze, klíčovými aspekty výběru jsou; stanovené cíle, hypotézy, jaký je předmět zkoumání. Z těchto aspektů vyplývá rámec a celková koncepce výzkumu. Výzkumník při volbě techniky hledá odpověď na otázky; jaký je nejvhodnější způsob praktického uskutečnění, jaká může být finanční a časová náročnost a zejména jaká bude efektivita realizované techniky. Samotná sociologie jako věda, nám těchto technik nabízí hned několik; k nejčastěji používaným a obecně známým patří *dotazník, anketa, pozorování a interview*, ale můžeme se setkat např. s tzv. *časovým snímkem, sociometrickým testem, nebo analýzou textů*. Někteří autoři () sem řadí další techniky jako je *brainstorming, brainwriting* nebo *delfskou metodu*. Techniky lze samozřejmě kombinovat, nejčastěji se kombinují s technikou pozorovací, neboť je částečně obsažena již ve zmiňované fázi příprav. Našemu účelu postačí seznámit se všeobecně s problematikou otázek a zejména s technikou dotazníku. (ČIHOVSKÝ 2006, s. 13-17), (REICHEL, 1993, 93-96)

1.2.5 Otázky; druhy a zásady

Existuje osm druhů otázek:

Zavřené: Nabízejí alternativy odpovědí. Dle nabízených alternativ dělíme zavřené otázky na dichotomické, kde jsou alternativy odpovědí ve formě ano, ne nebo muž,

žena a polytomické, kde je alternativ odpovědí více, ať už se jedná o výběr odpovědí, výčet nebo odpovědi uspořádané ve stupnici.

Otevřené: Nenabízejí alternativy odpovědí. Záleží ryze na respondentovi jak odpoví.

Polootevřené: Jsou kombinací otázek zavřených a otevřených, nejčastěji mají podobu „ano, ne? Jestliže ano, napište proč.“

Filtrační: Úkolem filtračních otázek je vyloučit nevhodné respondenty.

Projektivní: Jsou otázky nepřímé, nejsou orientovány přímo na respondenta. Nejčastěji mají tvar „co si myslíte o...“. Klasickým případem projektivních otázek je tzv. *bublinový test*, který má podobu komiksového obrázku do kterého má respondent vepsat, co si zobrazované objekty myslí nebo říkají.

Kontrolní: Kontrolní otázky jsou otázky které se ptají na to samé jako otázky předešlé, ale jiným způsobem. Slouží ke kontrole vypovídající hodnoty získaných odpovědí a slouží tedy kvalitě informací.

Úvodní: Otázky v úvodu dotazníku by měly být jednoduché a respondenta motivovat k zodpovězení celé sady (tzv. *baterii*) otázek.

Sociodemografické: Slouží k třídění odpovědí, resp. respondentů do demografických skupin. Jsou to otázky zaměřené např. na věk, pohlaví, bydliště, vzdělání atp.

Platí, že veškeré otázky by měly být stručné, jazykově a zároveň obsahově srozumitelné. V opačném případě hrozí nejen, že respondent odpoví špatně, nebo vůbec, ale zároveň respondenta odrazují od zodpovězení dalších otázek. S tím také souvisí, že by otázky neměly být věcně náročné a vnucovat respondentovi jiný názor než jeho vlastní. Poslední zásadou je, že by otázky neměly začínat slovem proč; vypovídající schopnost takových dotazů ztrácí význam.

Pohledu výzkumníka musí být otázky relevantní a empiricky smysluplná. Otázka se musí vztahovat jak k pojmům testované teorie tak k praktickým pojmům zkoumaného prostředí. Častými chybami jsou výčet odpovědí přímo v otázce a mlhavost. Další chybou anketiéra bývá skutečnost, že se jednou otázkou ptá na dvě i více věcí (*dvojitá, všeobjímající otázka*), nebo se ptá respondenta na názor někoho jiného. Otázka také nesmí být hypotetická nebo předpokládat stav či skutečnost, která u respondenta nemusí nastat. (JEŘÁBEK, 1993) (ČIHOVSKÝ, 2006, s. 13)

1.2.6 Technika dotazníku

Dotazník sestavujeme na základě toho, co chceme vědět. To bychom si měli ujasnit již ve fázi příprav. Na základě toho, co chceme zjistit, formulujeme otázky. Otázky je dobré roztřídit do bloků, dle jejich tématického zaměření. Přechody mezi různě orientovanými bloky otázek by měli být hladké. K tomu nám můžou posloužit různé spojovací věty nebo „oddychové“ odstavce, které respondenta seznámí s tím, co ho čeká a dají mu možnost odpočinku. Zde se klade důraz na schopnost výzkumníka vcítit se do role potenciálního respondenta. Jako první by měl být zařazen blok obsahující otázky neosobní a motivující, měl by respondenta zaujmout a přimět k vyplnění celého dotazníku. Otázky osobního charakteru, otevřené otázky zařazujeme až na závěr dotazníku. Na úplný konec dotazníku se, pokud je to nutné, zařazují otázky které by mohly respondenta naprosto odradit, tzn. otázky choulostivé, citlivého osobního rázu. Jeřábek (JEŘÁBEK, 1993) pro tyto otázky používá výstižného termínu „zeptej se a uteč“. Střední část dotazníku by se měla věnovat zbylým tématickým blokům. Samotnou podobu dotazníku před jeho distribucí je třeba tzv. pilotovat; ověřit srozumitelnost otázek, jejich správnou formulaci, včetně vhodnosti odpovědí.

Dotazník by ve svém úvodu měl obsahovat *průvodní dopis*, v něm by se měl respondent dozvědět účel výzkumu, kdo jej provádí, popř. financuje, jakým způsobem byl respondent k výzkumu vybrán. Samozřejmostí je prosba o spolupráci a příslib anonymity. V případě sběru osobních údajů, který je právně ošetřen zákonem č. 480/2004 Sb je nutné se navíc registrovat u Úřadu pro ochranu osobních údajů. Do průvodního dopisu patří i instrukce k vyplnění dotazníku a může v něm být připojena i nabídka zaslání výsledků výzkumu respondentovi.

Dotazník je nejpoužívanější výzkumnou technikou nejen pro svou relativní časovou a finanční nenáročnost, ale také pro svou nenáročnost na přípravu a počet anketiérů. Poskytuje respondentovi čas rozmyslet si dobře odpověď, umožňuje přímý i nepřímý kontakt a lze jej snadno opakovat. Má samozřejmě i své nevýhody, kterými jsou zejména možnost, že respondent odpovídá něco jiného, než si ve skutečnosti myslí. Také se může stát, zvláště při nepřímém kontaktu s respondentem, že respondent je úplně někdo jiný, než za koho se vydává. Při šíření a realizaci dotazníku elektronickou cestou, zejména tedy internetem, může samozřejmě dojít k situaci, že respondent zodpoví dotazník vícekrát, (ačkoliv se tento problém týká zejména různých zájmových

anket, anketních žebříčků atp.). Při statistickém šetření takových případů, kdy se může stát, že jeden respondent může odpovědět vícekrát se jedná o tzv. náhodný *výběr s opakováním*. V tomto případě, pokud je šetřený základní soubor velký je statistický rozdíl oproti situaci, kdy by každý respondent opravdu mohl zodpovědět dotazník jen jednou, zanedbatelný. Stoprocentní eliminaci těchto problémů nemůžeme uskutečnit ani v případě přímého kontaktu s respondentem a samozřejmě tomu nezabrání ani žádné technické řešení, (např. nemožnost hlasovat z jedné IP adresy vícekrát). Proto by měl anketiér mít vždy na paměti, že kvalita sbíraných dat závisí jen na něm a do stádia šetření patří rovněž kontrola a eliminace nepoužitelných zdrojů informací.

1.2.7 Velikost výběrového souboru, neboli vzorku

Odpověď na ideální velikost zkoumaného vzorku (tj. výběrového souboru) je celkem jednoduchá. Nejlepší vzorek je (co se týče množství) je samozřejmě ten, který obsahuje všechny prvky, nebo-li celý základní soubor. To je však často u sociologických výzkumů naprosto nemožné a proto odpověď na ideální velikost výběrového souboru zní: Abychom mohli z výběrového souboru předpovídat chování základního souboru, je nutné aby struktura výběrového souboru co nejlépe imitovala strukturu populace (souboru základního). (DISMAN, 2009, s. 92-115)

Ze hlediska statistické teorie je tedy ideální *soubor náhodný*, který je současně *reprezentativní*. Viz. kapitola 1.1.2.

1.2.8 Zkreslení korelace

V sociologii, jakožto ve vědě o přirozeném systému, který nelze nikdy úplně popsat, je důležité hledět kriticky i na validní data, získaná statistickými metodami. Nejčastějšími chybami při interpretaci závislosti dvou jevů (statistických znaků) jsou:

1. Nepravá korelace

Je korelace, k jejímuž zkreslení přispěl třetí nepozorovaný, nebo neanalyzovaný znak, který ovlivňuje oba zdánlivě závislé znaky.

2. Dvojitá příčina

Je případ kdy je jeden znak zdánlivě závislý na druhém, ale ve skutečnosti je závislý i na třetím, nebo je třetím tato závislost posilována

3. Chybějící střední člen

Mezi závislostí znaků X a Y ve skutečnosti stojí ještě třetí znak Z, tedy: $X \Rightarrow Z \Rightarrow Y$

4. Vývojová sekvence

Proměnná X ovlivňující proměnnou Y má předcházející proměnnou Z , která ji ovlivňuje.

V reálném světě nastávají některá zkreslení závislosti vždy, např. vše má svou příčinu a tato příčina má jinou příčinu (Vývojová sekvence), většina věcí má více než jednu příčinu (Dvojitá příčina), s nepravou korelací se setkáváme jako s nejčastější chybou výzkumníka a její odhalení závisí na výzkumníkově přehledu, nadhledu a důslednosti při analyzování zkoumané problematiky. Příklad chybějícího středního členu se pak často vyskytuje záměrně pro potvrzení různých „zajímavých“, nejčastěji konspiračních nebo xenofobních teorií. (DISMAN, 2009, s.20-24)

1.3 Internetové sociální sítě a bezpečné chování na internetu

Ekvivalentem sociální sítě jsou termíny komunita, komunitní, či společenská síť. Sociální síť je systém, jehož prvky tvoří lidé nebo organizace a vazby jsou tvořeny vztahy mezi nimi.

Internetové sociální sítě vytvářejí virtuální komunity na základě společných zájmů, různých sociálních vazeb, rodinných vazeb, profesí atp.

Důležité je uvědomit si, že internetovou sociální sítí je internet sám.

1.3.1 Historie a současnost internetových sociálních sítí

Dříve měly podobu spíše menších lokálních skupin lidí v různých internetových diskuzích, někdy třeba i s možností prezentace a vzájemného hodnocení umělecké, nebo profesní tvorby, nebo existovaly jako chaty (čti „čety“) věnující se různým problematikám. Webové aplikace umožňující interakci takovýchto komunit postupně rozšiřovaly nejen portfolio svých služeb a možností, ale také množství oblastí společných zájmů a tedy i množství uživatelů – už se nevěnovaly pouze omezené zájmové oblasti nebo konkrétní skupině lidí či organizací. To vedlo ke vzniku dnešních internetových sítí tak, jak je pod tímto termínem známe dnes; jako webové aplikace sdružující skupiny lidí nebo organizace, umožňující jejich společný styk a kontakt, kde portfolio služeb které internetové sociální sítě uživatelům nabízejí zahrnuje nejčastěji písemnou, ale také zvukovou i vizuální komunikaci, sdílení informací, souborů nebo jakýchkoliv jiných dat. Typickým znakem každé internetové sociální sítě jsou nástroje

pro vzájemnou dohledatelnost jejich uživatelů, které umožňují vznik vazeb mezi prvky těchto sociálních systémů.

Int. Soc. sítě fungují tak, že se uživatel zaregistruje, vytvoří si svou identitu zahrnující nejen data o něm ale i o jeho vazbách s ostatními prvky. Pak už jen vyhledává další uživatele nebo skupiny, přičemž je sám vyhledáván a vytváří komunikační a datové toky - vazby.

První internetovou sociální sítí byl tedy internet sám, resp. jeho předchůdce ARPA NET, jehož uživateli (uzly) byly čtyři americké univerzity. Poté přišlo Dalším významným krokem bylo objevení Internet Relay Chat – komunikace po internetu v reálném čase. Za první sociální sít' jako takovou je považována classmates.com, jenž vznikla v roce 1995. tyto webové stránky umožňovaly (a stále umožňují) registrovat se, hledat a udržovat vztahy mezi spolužáky a studenty. Na počátku 21. století následoval boom obřích sociálních sítí LinkedIn (2002), MySpace (2003), Facebook (2004), Twitter (2006) a mnoha dalších.

Na českém internetu dnes figurují zejména projekt společnosti Seznam Lidé.cz a asi největší Líbímseti.cz.

1.3.2 Problematika bezpečného chování na internetu

Internetová kriminalita je rozsáhlé téma, v této kapitole se zaměřím pouze na důvody a způsob, jak by se měli lidé na internetu, resp. na sociálních sítích chovat. Ačkoliv si mnozí lidé internetovou kriminalitu stále představují jako poškození počítače virem nebo jako hackerský útok za účelem vloupání se do jejich internetového bankovníctví, faktem zůstává, že největším nebezpečím jsou uživatelé sami sobě.

Pro naše pojetí bezpečnosti chování na internetových sociálních sítích je důležité vysvětlit několik termínů z oblasti nejčastěji používaných technik internetové kriminality zaměřené zejména na neoprávněný přístup k údajům, získání utajovaných informací nebo jiných informací o osobách, činnostech apod.:

Sociální inženýrství: ovlivňování, manipulace a navádění uživatelů k poskytnutí informací. Asi nejslavnější hacker a zároveň průkopník sociálního inženýrství Kevin Mitnick sám ve své slavné knize *Umění klamu* uvádí, že k činům, za které strávil ve vězení 5 let, neužíval zdaleka tak tvé znalosti IT jako právě sociálního inženýrství.

Sociotechnika: je ovlivňování a přesvědčování lidí s cílem oklamat je tak, aby uvěřili, že sociotechnik je osoba s totožností, kterou předstírá a kterou si vytvořil pro potřeby manipulace.

Phishing: bývá česky překládán jako „rybaření“, sociotechnická metoda, která prostřednictvím emailu podstrkuje uživateli email sloužící k vylákání informací jako jsou přístupové údaje a hesla. V ČR se stal nejslavnějším phishingovým útokem email odeslaný v roce 2006 zejména klientům České spořitelny, vyzívající k aktivaci nového bezpečnostního systému.

Útok hrubou silou: Anglicky známý jako *brute force attack* je pokus o dešifrování bez jakékoliv znalosti klíče šifry. Má podobu systematického testování všech, nebo omezené podmnožiny všech kombinací. Využíván velmi často obzvláště v případě, kdy má útočník alespoň částečné znalosti o přihlašovacích údajích - např. zná jméno účtu nebo délku hesla.

Mimo kybernetickou kriminalitu stojí stále častěji skloňování termín z oblasti personalistiky: *headhunting*. Jedná se o vyhledávání profesionálů na pozice nejčastěji vyššího a top managementu dle potřeb zadavatele. Headhunter přitom jen výjimečně oslovuje nezaměstnané kandidáty. Cílem naopak bývají ti nejuspěšnější pracující špičky ve svém oboru, metoda headhuntingu velmi často využívá výzkum na hraně zákona, kdy se snaží zjistit, často pod falešnými záminkami, interní informace z organizace, ve které „lovec mozků“ loví. Dnes snad každý personalista využívá (i v ne tak extrémních případech) internetu a sociálních sítí k získání jakýchkoliv informací o zaměstnancích potenciálních i stávajících.

Internet a sociální sítě jsou výborným prostředkem k získání takových informací. Je důležité mít na paměti, že bezpečné chování na internetu nechrání pouze před kriminálními činy, ale předchází jakékoliv možnosti eticky či neeticky využít data uživatelů a organizací.

1.3.3 Sdílení dat prostřednictvím internetu

Každé kybernetické kriminalitě předchází sběr dat, ať už účelný nebo ne, a právě v tom spočívá nebezpečí i krása internetu – nepředstavitelná množství dat, která se v této síti nachází někde s větší, jinde s menší ochranou a často bez jakéhokoliv byť jen pokusu o ochranu.

Problém nastává v momentě, kdy se jedná o data zneužitelná. Nejčastěji jsou takovými daty osobní údaje, nebo fotky, čísla účtů, přístupové údaje atd.

Bezpečnostní nevýhodou sociálních sítí a důvodem proč se tak často stávají cílem kybernetické kriminality je skutečnost, že se jedná o uzly s obrovskou koncentrací dat. A když už se nejedná o útok s cílem získat „množství dat“, jsou sociální internetové sítě výborným nástrojem k získání údajů o cílených osobách, samozřejmě za předpokladu, že se jedná o uživatele důvěřivého, nebo neznalého prostředí.

V poslední době se velice často upozorňuje na možné zneužívání dat sbíraných velkými internetovými společnostmi, přičemž nemusí jít o klasické sociální sítě. V tomto případě je nejčastěji kritizovaný Google.

Jde o fakt, že o poskytovaných službách a užitelných si Google sbírá data, která může využít třeba pro zkvalitnění svých služeb, ale třeba taky k něčemu úplně jinému. Stačí si uvědomit jaký monopol Google je, jaké služby nabízí a jaká data uživatelé Googlu „dávají“; od hledaných výrazů, přes email (Gmail), fotky (Picasa) až po dokumenty (Google Docs), videa (YouTube), geografická poloha mezi přáteli (Latitude) a spousty dalších. Ne málokomu při tom vysvitne myšlenka na „Velkého Bratra“.

1.3.4 Chyby nejen českých uživatelů, ohrožené skupiny

Uživatelé podceňují rizika, a provozovatelé webových služeb zase používají leckdy nesmyslná a často legislativně závadná pravidla provozu a slabé zabezpečení. Zásada by však měla být, vždy se jako uživatel snažit odpovídat za svou bezpečnost sám. Žádná technologie zabezpečení není neprolomitelná. Nejčastějšími prohřešky nejen českých uživatelů bývá: poskytování (sdílení) osobních informací jako je adresa, nebo telefon. Zaslání (sdílení) fotografií neznámým uživatelům, v nejhorším případě fotografií intimních. Sdělování a v extrémních případech i sdílení hesel. Otevírání neznámých emailových příloh, odkazů a webových aplikací.

Všechny tyto prohřešky plynou z přílišné důvěry uživatelů k jejich (internetovému) sociálnímu okolí. Uživatelé často věří skoro až bezmezně veškerým informacím, které se na ně z internetu a při internetové komunikaci valí. Důvěřivost uživatelů plyne z další sporné schopnosti internetu – internet poskytuje zdánlivou anonymitu. Každý asi zná nějaký z případů, kdy se ze sociálních sítí uvolnily soukromé informace, např. intimní fotky. Nebo, kdy byli lidé v důsledku uvolnění citlivých informací poškozeni, např. vykradením bydliště, zatímco byli na dovolené apod.

Bohužel méně známé jsou jednotlivé případy zneužití komunikace prostřednictvím sociálních sítí a internetových aplikací, které skončily znásilněním nebo zneužitím. Policie ČR v projektu (NE)bezpečný internet uvádí, že nejrizikovější skupinou jsou mladí lidé ve věku 15-20 let, a proto se nejen touto preventivní činností zaměřuje na děti ve věku 14-15 let tzn. na žáky 8. a 9. tříd základních škol.

Dle studie evropské komise zveřejněné v roce 2009 mají uživatelé těchto internetových sociálních služeb s věkem mezi 15 a 24 lety jen z 33% znalost svých práv v oblasti ochrany osobních údajů a jen 18% ví, že existují národní úřady, kam se mohou obrátit pro pomoc.

Výzkumy a statistiky věnující se pouze českým uživatelům prozatím buď nejsou, nejsou přístupné nebo bývají nedůvěryhodné.

2 Praktická část

Výzkum je rozdělen do tří základních oblastí výzkumu:

1. Část: Autentizace
2. Část: Důvěřivost a dostupnost informací
3. Část: Věnující se citlivosti poskytovaných údajů

A dvou sociologicky důležitých oblastí

4. Část: Osobní informace pro statistiku
5. Část: Motivující otázky a snad zajímavé postoje českých uživatelů sociálních sítí.

2.1 Formulace hypotéz

Má-li se tato práce zabývat chováním uživatelů sociálních sítí z hlediska bezpečnosti, pak bychom tzv. základní - teoretickou hypotézu, jakožto cíl této práce mohli formulovat takto:

Chová se opravdu většina (českých) uživatelů sociálních sítí, způsobem, který je ohrožuje z hlediska úniku nežádoucích informací a dat? Do jaké míry? A existují typické ohrožené skupiny?

Tato hypotéza vznikla v důsledku získaných informací o dané problematice v rámci předvýzkumné části. Detailnějším rozbořem teoretické hypotézy byly stanoveny dílčí, tedy pracovní hypotézy:

1. Uvádí čeští uživatelé sociálních sítí v profilu svou adresu? (str. 43)
2. Uvádí čeští uživatelé sociálních sítí v profilu svůj telefon? (str.43)
3. Jsou si vůbec vědomi, jaká data lze o nich na internetu dohledat? (str.41)
4. Skladba uživatelů dle pohlaví je shodná. (str. 45)
5. Je pravdou, že každý pátý uživatel používá slabé heslo? (str. 40)
6. Liší se statistiky českých uživatelů od statistik celosvětových, evropských? (str.48)
7. Jak silná mají hesla? (str. 37)
8. Poskytují uživatelé ve svých profilech takové fotografie, které by je mohly kompromitovat? (str.43)
9. Poskytují taková data (fotografie) jejich známí a přátelé? (str.43)
10. Přijmou tzv. „nabídku na přátelství“ od neznámého člověka? (str. 41)
11. Jak snadno poskytnou svůj email? (str.54)

12. Existuje závislost mezi některými měřenými znaky na věku, pohlaví nebo vzdělání?

13. Jaký je jejich průměrný věk, vzdělání?

2.2 Rozhodnutí o populaci a vzorku

Zkoumaný základní soubor (populace): Všichni čeští (resp. česky mluvící) uživatelé největších zahraničních a českých sociálních sítí. Za reprezentativní vzorek byli zvoleni uživatelé nejznámějších zahraničních sociálních sítí to je Facebook.com, Twitter.com a z největších českých sociálních sítí – tzn. Lide.cz a Libimseti.cz

Výběrový soubor bude vybrán jako náhodný výběrový soubor s opakováním, českých uživatelů sociálních sítí Twitter.com, Facebook.com, Lide.cz. Libimseti.cz.

2.3 Pilotní studie

Nestandardizovaným rozhovorem ve svém okolí jsem si ověřil, že na zamýšlené otázky a okruhy jsou lidé schopni odpovídat způsobem, který jsem schopen analyzovat.

Problematika chování uživatelů sociálních sítí z hlediska bezpečnosti však zanechala stopy na technice sociologického výzkumu a její konstrukce, neboť se zajisté jedná o citlivé téma, které u některých respondentů může vyvolat dojem, že je právě jeho bezpečnost účastí na výzkumu ohrožena.

2.4 Rozhodnutí o technice a sběru informací

Z ohledem na populaci, jejímž specifíkem je výskyt na internetu, resp. na sociálních sítích, zvolil jsem techniku dotazníku, k jehož distribuci byl využit jako zprostředkovatel internet a zejména samotné sociální sítě a jejich uživatelé. U této skupiny je technika dotazníku nejvhodnější s ohledem na velmi nízké náklady a relativně velkou návratnost.

Dotazník byl distribuován do všeobecných skupin a diskuzí o sociálních sítích, tak aby nedošlo ke zkreslení reprezentativnosti výběrového souboru.

2.5 Konstrukce dotazníku

Dotazník spočíval nejprve ve vytvoření samotných otázek a jejich rozčlenění dle sociologických standardů a tématických okruhů (nadpisy částí těchto okruhů - kurzívou - zůstaly uživatelově oku skryty nastavením v souboru kaskádových stylů)

Výsledný dotazník obsahoval 20 otázek, z toho 1 nepovinnou a měl nesledující podobu:

Prosím vyplňte tento dotazník určený k mé bakalářské práci

Je určen pro uživatele internetových sociálních sítí* a veškeré otázky se zaměřují právě na tyto služby*.

Díky. Mara W. Student VUT v Brně

*internetové sociální sítě (dále jen soc. sítě) = facebook, twitter, libimseti apod.

• *Motivující otázka*

1. **Ohodnot'te své znalosti informačních technologií jako ve škole**

 (1-5)

• *1. Část: Autentizace*

2. **Kolik znaků má vaše heslo?**

 (5 a méně, 6, 7... 10 a více)

3. **Používáte funkci "automatického zapamatování hesla"?**

 Ano Ne

4. **Jaké znaky v hesle používáte?**

 Písmena Číslice Speciální znaky

5. **Používáte heslo k vašemu profilu na soc. síti i jinde (např. pro přístup k emailu)?**

 Ano Ne

6. **Používá někdy váš počítač i někdo z vašich přátel nebo lidí mimo rodinu?**

 Ano Ne

• *2. Část: důvěřivost a dostupnost informací?*

7. **Upravujete si nastavení soukromí (nastavení profilu, sdílení informací)?**

 Ano Ne

8. **Přijmuli jste někdy nabídku na přátelství od neznámého člověka?**

 Ano Ne

9. **Víte jaké informace se o Vás dají na internetu dohledat?**

 Ano Ne

10. **Jste si vědomi komerčního a jiného zneužívání tzv. skupin?**

 Ano Ne

• *3. Část věnující se citlivosti poskytovaných údajů*

11. Uvádíte ve svém profilu svou adresu?

Ano Ne

12. Uvádíte ve svém profilu svůj telefon?

Ano Ne

13. Sdílíte fotky, které by vás mohly nějakým způsobem inkriminovat (např. by nebylo vhodné, aby je viděl váš zaměstnavatel, třeba i potenciální.)

Ano Ne

14. Sdílí někdo z Vašich přátel takové fotky, které by vás mohly nějakým způsobem inkriminovat (např. by nebylo vhodné, aby je viděl váš zaměstnavatel, třeba i potenciální.)

Ano Ne

- *Motivující otázka a zároveň snad i zajímavý postoj vhodný k analýze*

15. Se kterým tvrzením se ztotožňujete spíše:

- Mou bezpečnost by měl zajišťovat provozovatel.
 Svou bezpečnost si musím zajistit sám.

- *Osobní informace pro statistiku*

16. Vaše pohlaví:

Muž Žena

17. Váš věk?

(5-90)

18. Vaše nejvyšší dosažené vzdělání (popř. současně studující)?

ZŠ SŠ VŠ

19. Stav:

Student Pracující Ostatní

20. Údaje jsou anonymní a nelze je nijak zneužít. V případě zájmu o výsledky výzkumu mi napište svůj email:

Snahou bylo dodržet většinu otázek uzavřených. Otevřenou zůstala jen jedna, poslední, jenž není otázkou, ale spíše sociotechnickým pokusem, jak snadné bylo z respondentů vylákat jejich email.

Dotazník byl vytvořen jako webová stránka na adrese <http://dotaznik.php5.cz>. Samotný zdrojový kód byl vytvořen jako Html 4.01 Transitional s použitím skriptovacího programovacího jazyku PHP zejména pro ošetření dat a jejich uložení do souboru pojmenovaném *data.dat*. Zdrojový kód navíc obsahoval jednu funkci skriptovacího jazyka JavaScript a to pro návrat k dodatečnému vyplnění otázek případně špatného vyplnění odpovědí.

Po správném vyplnění dotazníku byla data zapsána do souboru *data.dat* a respondent byl automaticky přesměrován na stránku s poděkováním. Přesměrování bylo použito pro znemožnění rychlého vícenásobného zodpovězení dotazníku. Soubor *data.dat* obsahoval odpovědi respondentů většinou v číselné formě (0 pro „ne“, 1 pro „ano“ apod.) oddělené středníkem, každý řádek představoval jednoho respondenta.

Službám a aplikacím ala veřejný dotazník pomocí GoogleDocs jsem se vyhnul záměrně, vlastní vytvoření dotazníku mi umožnilo zachovat data ve formátu, který mi umožňoval plánované zpracování pomocí excelu, nebo podobného tabulkového procesoru.

2.6 Předvýzkum

Dotazník byl ve své formě volně dostupného internetového dotazníku testován na lidech v mém okolí. Nebyly odhaleny žádné významné nedostatky. Doplněna byla pouze informace, že se jedná o dotazník k bakalářské práci.

2.7 Terénní šetření:

Terénní šetření, nebo-li sběr dat proběhlo v době od 1.2.2010 do 8.5.2010. Dotazník byl distribuován do obecných diskuzí a skupin na zmíněných čtyřech největších sociálních sítích Facebook, Twitter, Libimseti a Lide.cz a dále také na diskuzní fóra věnující se těmto sítím. Součástí byla i emailová kampaň s žádostí o další rozeslání tohoto dotazníku.

2.8 Statistické zpracování

2.8.1 Autentizace - základní charakteristiky

Kolik znaků má respondentovo heslo		
Hodnota	Četnost	Relativní četnost
5 a méně	15	0,08
6	25	0,13
7	16	0,09
8	40	0,22
9	16	0,09
10 a víc	74	0,40
Celkem	186	1,00

Tabulka 2; Počet znaků hesla

Odpovědi na 2. otázku „Kolik znaků má vaše heslo“ přinesly zajímavý poznatek v podobě silných hesel délky 10 a více, které užívá 40% respondentů.

Menším překvapením pak byla základní charakteristiky znaku „jaké znaky respondent používá“, kde se nejpopulárnější kombinací ukázala kombinace písmen a čísel, ale ani nejsilnější kombinace čísel, písmen a speciálních znaků nebyla nepopulární, užívá ji 13% respondentů :

Jaké znaky v hesle respondent používá		
Hodnota	Četnost	Relativní četnost
Písmena	36	0,19
Čísla	6	0,03
Speciální znaky	4	0,02
Písmena a čísla	110	0,59
Písmena a speciální znaky	6	0,03
Čísla a speciální znaky	0	0,00
Kombinaci všech tří	24	0,13
Celkem	186	1,00

Tabulka 3; Druhy znaků v hesle

Následující tabulka znázorňuje popularitu tzv. funkce „automatického zapamatování hesla“, která je vhodná pouze u PC, které používá výhradně jeho vlastník, a to zejména v případě, kdy může dojít k „odkroužení“ tisknutých kláves při psaní hesla. Na druhou stranu této funkce využívá často škodlivý software, který má v případě, že se na takovém PC nachází, přístup k zaveslovaným datům.

Používá respondent funkci automatického zapamatování hesla?		
Hodnota	Četnost	Relativní četnost
NE	85	0,46
ANO	101	0,54
Celkem	186	1,00

Tabulka 4; Automatické zapamatování hesla

Odpovědi k užívání této funkce byly velmi vyrovnané, avšak lehce vítězí užívání této funkce, k jejímu užívání se přihlásilo 54%. Zajímavé pak bude zajisté srovnání s odpověďmi na otázku, zda „uživatelské PC“ užívá i někdo jiný mimo rodinu. Základní charakteristiky znaku, zda uživatelské PC užívá i někdo mimo rodinu, znázorňuje následující tabulka:

Používá stejný počítač i někdo jiný mimo respondentovu rodinu?		
Hodnota	Četnost	Relativní četnost
NE	138	0,74
ANO	48	0,26
Celkem	186	1,00

Tabulka 5; Užívání respondentova PC i někým jiným

Při pohledu na data znaku délky hesla respondentů způsobem, kdy ošetříme data tak, že 5 a méně znaků charakterizujeme zkrátka jako 5, neboť menší počet znaků je prakticky u většiny sociálních sítí nemožný, stejně tak odpověď 10 a více charakterizujeme jako prostých 10, neboť větší počet znaků značí již velice silné heslo, pak použitím charakteristik pro kvantitativní znaky jsou charakteristiky znaku délky hesla:

Výběrový průměr	8,28
Výběrový rozptyl	2,96
Směrodatná odchylka	1,72
Minimum	5
Maximum	10
Medián	8
Dolní kvartil	10
Horní kvartil	7

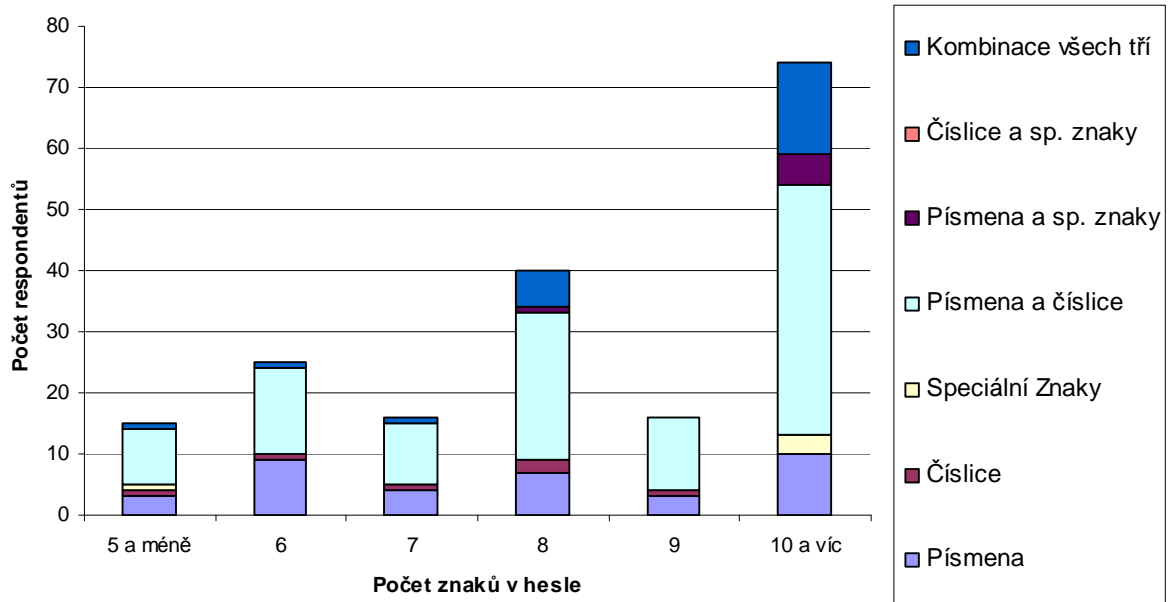
Tabulka 6; Délka hesla - kvant. charakteristiky

Průměrná délka hesla respondentů pak byla 8,28 znaku, přičemž pak odhad směrodatné odchylky je 2,96 znaků.

V základních charakteristikách znaků věnujícím se oblasti autentizace je třeba uvést jejich detailnější poměry a vzájemné vztahy:

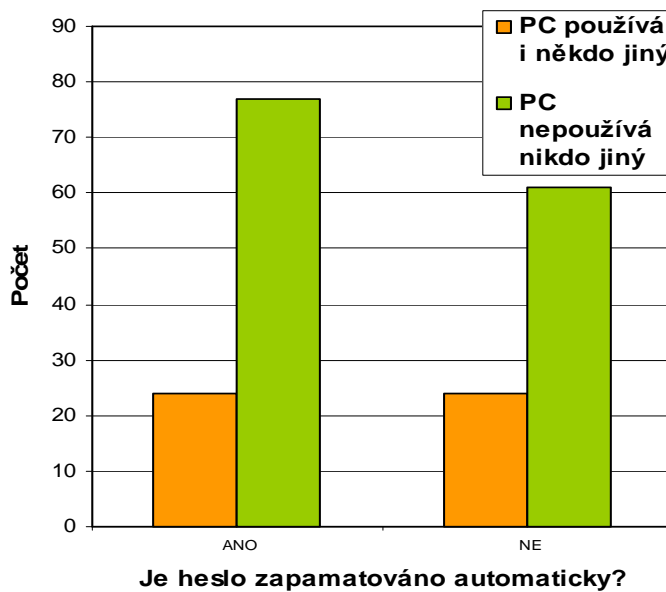
2.8.2 Autentizace - Tabulky a grafy:

Pro znázornění stavby hesel dle jejich délky, znakové skladby a četnosti užití slouží následující skládaný sloupcový diagram:

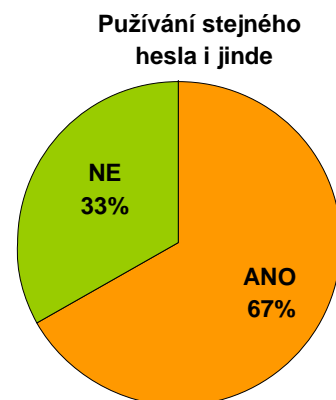


Graf 1; délka, skladba a popularita hesel

K Užívání PC někým jiným (mimo rodinu) se přihlásilo zhruba 26% respondentů. První graf znázorňuje poměry odpovědí, zda používají funkci automatického zapamatování hesla u uživatelů kteří své PC nikomu neposkytují a uživatelů k jejichž PC mají přístup i jiní lidé mimo rodinu.



Graf 2; Sloupcový diagram - automatické zapamatování hesla



Graf 3; Kruhový výšečový diagram

Druhý graf je o něco zajímavější a znázorňuje jak populární je mezi uživateli nešvar používání stejného hesla k účtu na sociální síti u více aplikací.

Skladbu hesel z hlediska délky a kombinace znaků znázorňuje následující kontingenční tabulka četností:

Délka hesla \ Použité znaky	Délka hesla						Celkem
	5 a méně	6	7	8	9	10 a víc	
Písmena	3	9	4	7	3	10	36
Číslice	1	1	1	2	1	0	6
Speciální znaky	1	0	0	0	0	3	4
Písmena a číslice	9	14	10	24	12	41	110
Písmena a sp. znaky	0	0	0	1	0	5	6
Sp. Znaky a číslice	0	0	0	0	0	0	0
Kombinace všech tří	1	1	1	6	0	15	24
Celkem	15	25	16	40	16	74	186

Tabulka 7; Četností délky hesla a použitých znaků

K přehledu síly hesel uvádím i tabulku relativních četností v %, která nám negativně zodpoví pracovní hypotézu, že slabé heslo používá 20% uživatelů.

Délka hesla \ Použité znaky	Délka hesla						Celkem
	5 a méně	6	7	8	9	10 a víc	
Písmena	1,61	4,84	2,15	3,76	1,61	5,38	19,35
Číslice	0,54	0,54	0,54	1,08	0,54	0,00	3,23
Speciální znaky	0,54	0,00	0,00	0,00	0,00	1,61	2,15
Písmena a číslice	4,84	7,53	5,38	12,90	6,45	22,04	59,14
Písmena a sp. znaky	0,00	0,00	0,00	0,54	0,00	2,69	3,23
Sp. Znaky a číslice	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kombinace všech tří	0,54	0,54	0,54	3,23	0,00	8,06	12,90
Celkem	8,06	13,44	8,60	21,51	8,60	39,78	100,00

Tabulka 8; Tabulka relativních četností délky hesla a použitých znaků

Z této tabulky plyne fakt, že silná hesla (délky 7 a delší a zároveň kombinující alespoň dva druhy znaků) užívá 55,91% respondentů. Slabé heslo tedy užívá bezmála 30% respondentů.

Celkové poměry znaků, pro zpřehlednění označených písmeny G a D, tj. odpovědí na otázku

G: zda k respondentově počítači mají přístup i lidé mimo rodinu, a

D: zda užívají funkci automatického zapamatování hesel,

znázorňují následující 2 tabulky četností odpovědí a také relativních četností v %:

G\D	ANO	NE	Celkem
ANO	24	24	48
NE	77	61	138
Celkem	101	85	186

Tabulka 9; Kontingenční tabulka četností znaků G a D

G\D	ANO	NE	Celkem
ANO	12,90	12,90	25,81
NE	41,40	32,80	74,19
Celkem	54,30	45,70	100,00

Tabulka 10; Čtyřpolní tabulka rel. četností G a D v procentech

2.8.3 Dostupnost dat a důvěřivost uživatelů – charakteristiky

Upravujete si nastavení soukromí?		
Hodnota	Četnost	%
NE	32	17,20
ANO	154	82,80
Celkem	186	100

Tabulka 11; Úprava nastavení soukromí

Další znak je odpovědí na otázku, zda si uživatelé na sociální síti upravují nastavení soukromí, tzn. s kým jaká data sdílejí, resp. kdo má k jejich datům přístup a kdo s nimi může ,komunikovat apod. asi 17,2% respondentů toto nastavení nepoužilo.

Přijetí přátelství od neznámého člověka		
Hodnota	Četnost	%
NE	118	63,44
ANO	68	36,56
Celkem	186	100

Tabulka 12; Přijetí přátelství od neznámého člověka

Výše uvedená tabulka základních charakteristik udává kolik respondentů, tzv. přijalo nabídku na přátelství od neznámého člověka, tzn. přijalo ho do bližšího komunitního okruhu s bližším přístupem k jeho datům.

Zajímavým znakem v části dostupnosti dat a důvěřivosti uživatelů byla informace, zda vědí jaké informace se o nich dají na internetu dohledat:

Víte, jaké informace se dají o Vás dohledat na internetu?		
Hodnota	Četnost	%
NE	42	22,58
ANO	144	77,42
Celkem	186	100

Tabulka 13; Ví, co se o něm/ní dá dohledat na int.

Bez mála 23% respondentů tedy neví jaké informace se o nich dají dohledat.

Celkem jasně dopadla odpověď na otázku, zda jsou si respondenti vědomi komerčního a jiného zneužívání skupin. Bez mála 92% respondentů si je této skutečnosti vědoma.

Víte o zneužívání skupin na s.s.?		
Hodnota	Četnost	%
NE	15	8,06
ANO	171	91,94
Celkem	186	100

Tabulka 14; Víte o zneužívání skupin

2.8.4 Dostupnost dat a důvěřivost uživatelů - tabulky a grafy

Tabulka níže uvedená znázorňuje v procentech zastoupení znaků, zda vědí co se o nich dá dohledat na internetu a zda přijali nabídku na přátelství od neznámého člověka tabulka je rozdělena na dvě části, v první části se nacházejí uživatelé kteří si neupravují nastavení svého soukromí:

Vědí co se dá o nich dohledat na internetu	Přijali "přátelství" od neznámého člověka	
	NE	ANO
NE	4,30	2,69
ANO	7,53	2,69
Celkem	11,83	5,38

Tabulka 15; Znaky uživatelů nenastavujících si soukromí - v procentech

A v druhé kteří si nastavení soukromí již upravují:

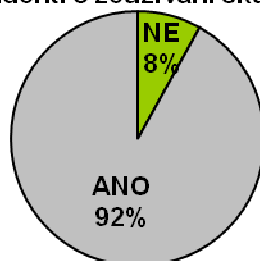
Vědí co se dá o nich dohledat na internetu	Přijali "přátelství" od neznámého člověka	
	NE	ANO
NE	9,14	6,45
ANO	42,47	24,73
Celkem	51,61	31,18

Tabulka 16; Znaky uživatelů nastavujících si soukromí - v procentech

Nejhorší skupinou je 4,3% respondentů, kteří negativně odpověděli současně na všechny tři otázky. Plně uspokojivá nejsou však ani další čísla.

Jasný výsledek znaku K, zda vědí uživatelé o zneužívání skupin znázorňuje následující kruhový výšečový diagram:

Vědí respondenti o zeužívání skupin?



Graf 4; Zneužívání skupin

2.8.5 Citlivost údajů – charakteristiky

Adresa v profilu		
Hodnota	Četnost	%
Ne	159	85,48
Ano	27	14,52
Celkem	186	100,00

Tabulka 17; Adresa v profilu na sociální síti

Pro mne alarmujícím faktem je 14,52% uživatelů udávajících na sociální síti svou adresu. Stejně jako silné zastoupení (16,67%) uživatelů udávajících na sociální síti svůj telefon:

Telefon v profilu		
Hodnota	Četnost	%
Ne	155	83,33
Ano	31	16,67
Celkem	186	100,00

Tabulka 18; Telefon v profilu na sociální síti

Často diskutovanou skutečností je distribuce různých nevhodných fotek, často např. intimních, v profilech uživatelů na sociálních sítích. Následující tabulka znázorňuje kolik uživatelů se přiznalo ke sdílení svých nevhodných fotek, resp. fotek které by je mohly nějakým způsobem inkriminovat, nebo poškodit.

Sdílí uživatel inkriminující fotky?		
Hodnota	Četnost	%
Ne	158	84,95
Ano	28	15,05
Celkem	186	100,00

Tabulka 19; Sdílí ve svém profilu inkriminující fotky

K tomuto nešvaru se přiznalo 28 respondentů ze 186, tzn 15,5%. Často se ale stává, že samotný uživatel takové fotky nesdílí, ale jeho přátelé ano. Úplnou katastrofou jsou pak případy, kdy je na de facto cizích fotkách respondent identifikován, např. v popisu fotky.

Informace zda uživatelé ví o existenci fotek, které by je mohli nějakým způsobem inkriminovat, v profilech svých přátel přináší tabulka základních charakteristik tohoto znaku:

Sdílí fotky inkriminující respondenta jeho přátelé?		
Hodnota	Četnost	%
Ne	129	69,35
Ano	57	30,65
Celkem	186	100,00

Tabulka 20; Sdílí vaši přátele inkriminující fotky

2.8.6 Citlivost údajů - Tabulky a grafy:

Následující kontingenční tabulky znázorňují vztahy mezi znaky týkající se citlivosti údajů (v procentech). První z nich se týká uživatelů nesdílejících nevhodné fotky:

Uvádí telefon	Uvádí Adresu		Celkový součet
	NE	ANO	
NE	63,44	6,99	70,43
ANO	11,29	3,23	14,52

Tabulka 21; Údaje uživatelů nesdílejících nevhodné fotky v procentech

A druhá naopak uživatelů, kteří potenciálně škodlivé fotky sdílejí:

Uvádí telefon	Uvádí Adresu		Celkový součet
	NE	ANO	
NE	10,22	2,69	12,90
ANO	0,54	1,61	2,15

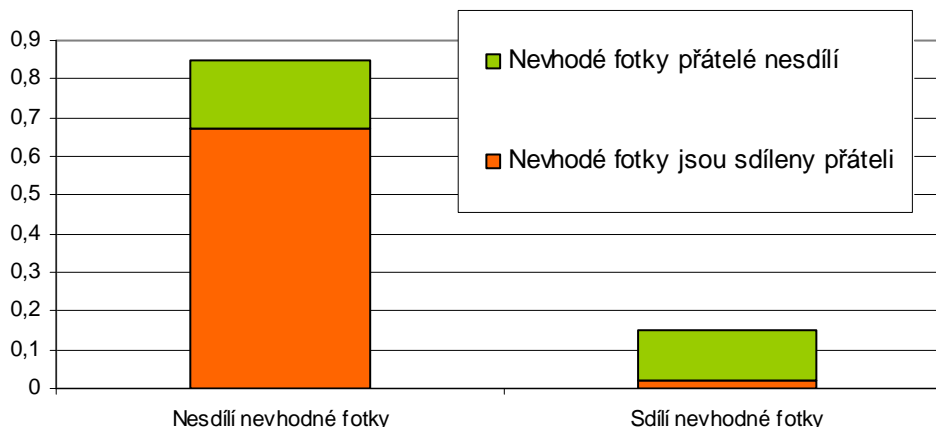
Tabulka 22; Údaje uživatelů sdílejících nevhodné fotky v procentech

Sdílí nevhodné fotky	Kompromitující fotky jsou sdíleny někým jiným		Celkem
	NE	ANO	
NE	67,20	17,74	84,95
ANO	2,15	12,90	15,05
Celkem	69,35	30,65	100,00

Tabulka 23; Čtyřpolní tabulka o sdílení fotek v procentech

Dále uvedený graf ukazuje silný podíl sdílení nevhodných fotek přáteli u uživatelů co sami takové fotky sdílejí. Znamená to tedy, že pokud jsou uživatelé v tomto ohledu,

řekněme, shovívaví, pak stejně shovívaví jsou pak často i jejich přátelé. Naštěstí takovou skupinu tvoří asi 15% respondentů



Graf 5; Nevhodné fotky sdílejí vs. jsou o nich sdíleny – v procentech

2.8.7 Osobní údaje - charakteristiky

Výzkumu se zúčastnilo 105 mužů a 81 žen, procentuálně byla tedy pohlaví zastoupena podobně, což je očekávatelné.

Pohlaví		
Hodnota	Četnost	%
Muž	105	56,45
Žena	81	43,55
Celkem	186	100,00

Tabulka 24; Pohlaví

Zajímavější skladbu však u českých uživatelů sociálních sítí nalezneme z pohledu jejich vzdělání (i současně studujících), kde nejvyšší zastoupení mají středoškoláci – 63%.

Nejvyšší vzdělání (+ právě studující)		
Hodnota	Četnost	%
ZŠ	27	14,52
SŠ	118	63,44
VŠ	41	22,04
Celkem	186	100,00

Tabulka 25; Vzdělání

Výzkumu se zúčastnilo 117 studentů, což tvořilo 62% z celkového počtu respondentů.

Stav		
Hodnota	Četnost	%
Student	117	62,90
Pracující	50	26,88
Ostatní	19	10,22
Celkem	186	100,00

Tabulka 26; Stav

Detailnější charakteristiky věku uživatelů nám poskytují zajímavé informace zejména odhadu průměrného věku 22,96 ve srovnání s mediánem 22. Což značí celkem rovnoměrnou distribuci četností jednotlivých věků.

Věk	
Počet dat	186
Suma	4270
Výběrový průměr	22,96
Výběrový rozptyl	43,75
V. směrodatná	6,61
Minimum	11
Maximum	65
Medián	22
Dolní kvartil	25
Horní kvartil	20

Tabulka 27; Věk respondentů

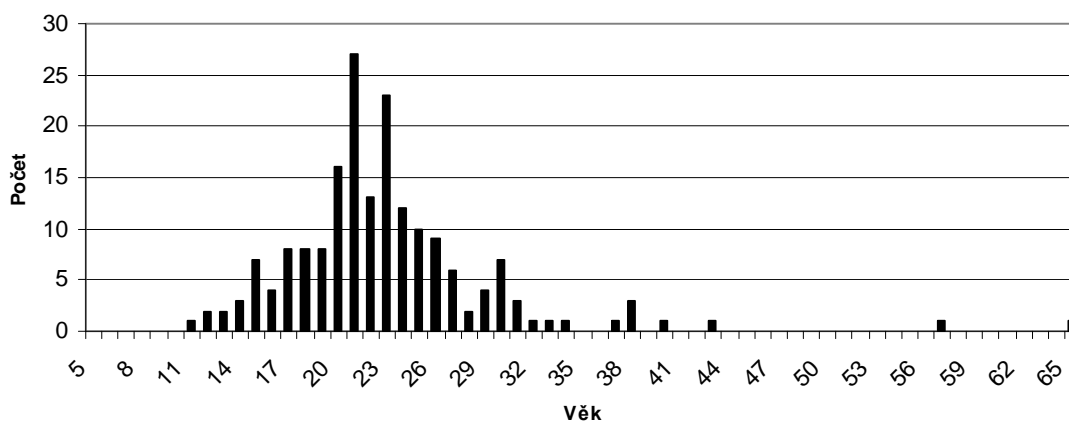
2.8.8 Osobní-údaje - tabulky a grafy

Konkrétní skutečnosti, zda se jednalo o studenty, či dostudované a jakým vzděláním znázorňuje následující kontingenční tabulka:

Vzdělání	Stav			Celkem
	Student	Pracující	Ostatní	
ZŠ	27			27
SŠ	68	39	11	118
VŠ	22	11	8	41
Celkem	117	50	19	186

Tabulka 28; Znaky stav a vzdělání

Histogram četností věků ukazuje i extrémní hodnoty, které byly započítány, ale vzhledem k jejich četnosti neovlivnily výrazně žádnou ze statistických charakteristik:



Graf 6; Věkové rozdělení

2.8.9 Intervalové odhady věkových charakteristik

Dvoustranný 95% interval spolehlivosti střední hodnoty μ pro věk uživatelů sociálních sítí s rozdělením $N(\mu, \sigma^2)$ je interval od 22,00 do 24 let (zaokrouhлено na celá místa).

Dvoustranný 95% interval spolehlivosti rozptylu σ^2 pro věk uživatelů sociálních sítí s rozdělením $N(\mu, \sigma^2)$ je interval od 35,94 do 54,08 let (zaokrouhлено na dvě desetinná místa). Směrodatnou odchylku pokrývá s 5% možností omylu interval od 6 do 7,35 let.

2.8.10 Testy hypotéz:

Hladina významnosti všech testů byla zvolena $\alpha = 0,05$.

Průměrný věk českých uživatelů soc. sítí je nižší 26,9 let

Pro relevantnost výzkumu je důležité, aby výběrový soubor odpovídal svými vlastnostmi co nejvěrněji základnímu souboru. Pro otestování, nakolik je výběrový soubor tohoto výzkumu shodný jsem si vybral test střední hodnoty věku českých uživatelů sociálních sítí, která se podle agentury PR.Konektor rovná 26,9 let

Výzkum agentury PR.Konektor je však nejspíš zkreslen směrem k vyšším výsledkům, neboť dotazník byl v angličtině a statistiky serveru ChekFacebook.com a zprávy ČTK uvádějí, že věk českých uživatelů sociálních sítí se pohybuje mezi 18 až 24 let.

1. Testujeme hypotézu: průměrný věk českých uživatelů int. soc. sítí je nižší 26,9 let.

Proto formuluji nulovou hypotézu o střední hodnotě věku českých uživatelů sociálních sítí

$$H_0: \mu \leq \mu_0, \text{ kde } \mu_0 = 26,9$$

a alternativní hypotéza $H_1: \mu > \mu_0$

2. Realizovaná hodnota testového kritéria $t = \frac{22,96 - 26,9}{6,61} \sqrt{186} = -8,13$

3. Kritický obor $W_{0,05} = \{t: t > 1,1653\}$

4. Protože se hodnota testového kritéria v kritickém oboru nerealizovala, přijímáme nulovou hypotézu H_0 . To znamená, že střední hodnota věku českých uživatelů sociálních sítí je menší nebo rovna 26,9 let. Což v důsledku potvrzuje kvalitu našeho výběrového souboru.

Celosvětový věkový průměr je vyšší než u českých uživatelů

1. Testujeme hypotézu: průměrný věk českých uživatelů int. soc. sítí je nižší než celosvětový věkový průměr 37 let (průměrný věk uživatelů soc. sítí, dle celosvětové studie švédské společnosti Pingdom AB).

$H_0: \mu \geq \mu_0$, kde $\mu_0 = 37$

$H_1: \mu < \mu_0$

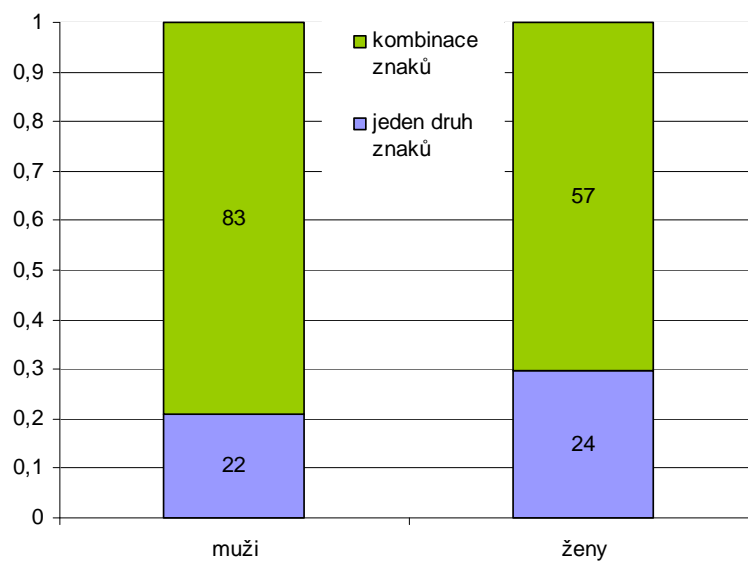
2. Realizovaná hodnota testového kritéria $t = -28,97$

3. Kritický obor $W_{0,05} = \{t: t \leq 1,1653\}$.

4. Protože se hodnota testového kritéria realizovala v kritickém oboru, nulovou hypotézu H_0 zamítneme s 5% možností omylu a přijímáme alternativní hypotézu H_1 , tedy; Střední hodnota věku českých uživatelů je nižší než celosvětová.

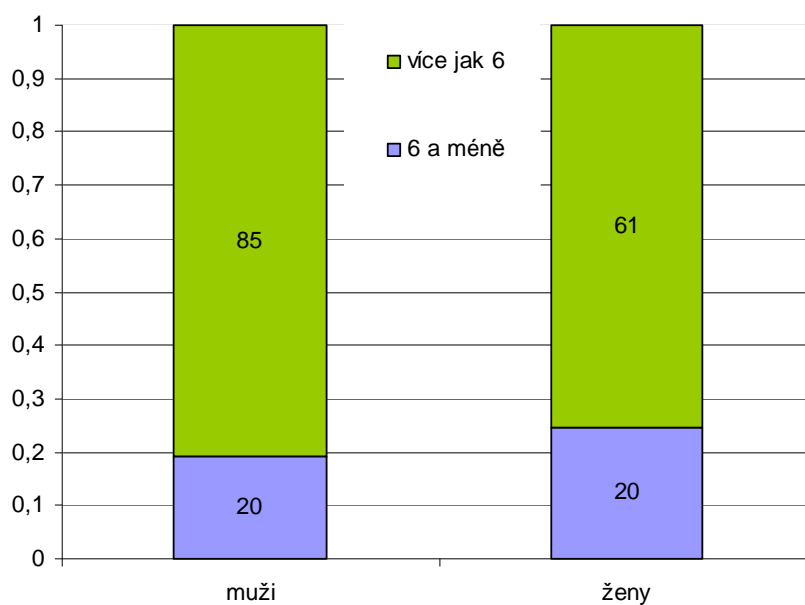
Ze získaných dat nebyly patrné žádné závislosti na pohlaví ani věku, pro názornost jsem však zvolil dva příklady; první dvojce grafů pro znázornění v jakých poměrech byla data, která nenaznačovala žádnou závislost a druhý, který sice závislost naznačoval, ale testováním se ukázala jako velmi slabá.

Jak znázorňuje následující dvojce grafů, na pohlaví zajisté kvalita hesla nezávisí:



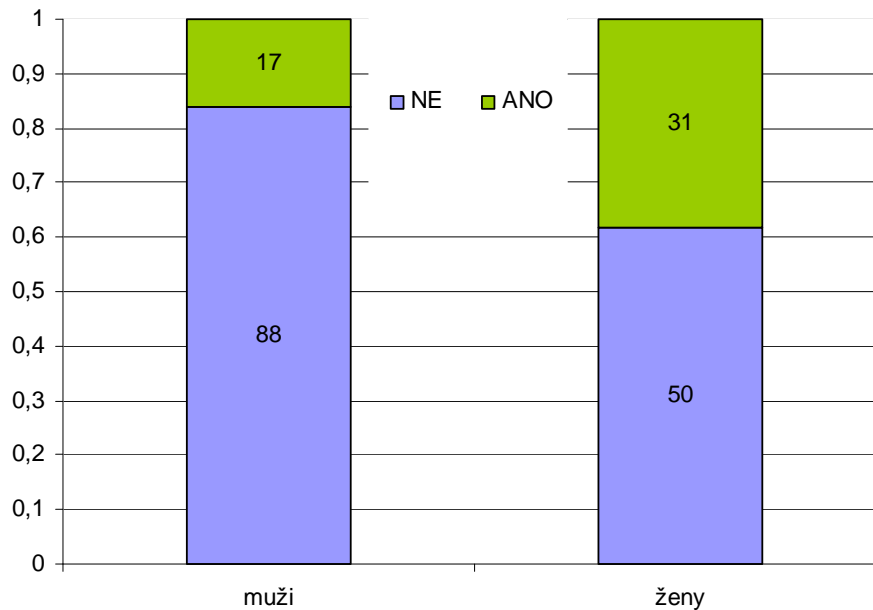
Graf 7; kombinace znaků a pohlaví

Stejně tak i níže uvedený graf závislost nenaznačoval, a test ji samozřejmě ani nepotvrdil.



Graf 8; počet znaků a pohlaví

Na otázku, zda ke svému počítači respondenti pouštějí i lidi mimo rodinu však již ženy značně vedou a graf naznačuje silnou závislost:



Graf 9; Má přístup k počítači i někdo jiný?

Existuje závislost mezi pohlavím a poskytováním přístupu ke svému PC lidem mimo rodinu?

1. Testujeme hypotézu: poskytování přístupu ke svému PC je u českých uživatelů soc. sítí závislé na pohlaví.

H_0 : Neexistuje závislost mezi pohlavím a poskytováním přístupu ke svému PC lidem mimo rodinu.

H_1 : Závislost mezi pohlavím a poskytováním přístupu ke svému PC lidem mimo rodinu existuje.

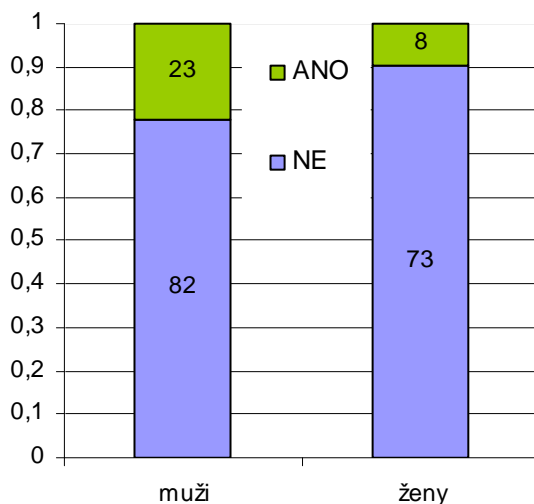
2. Hodnota testového kritéria $\chi^2 = 11,64$

3. Kritickým obor $W_\alpha = \{\chi^2 : \chi^2 \geq 3,841\}$,

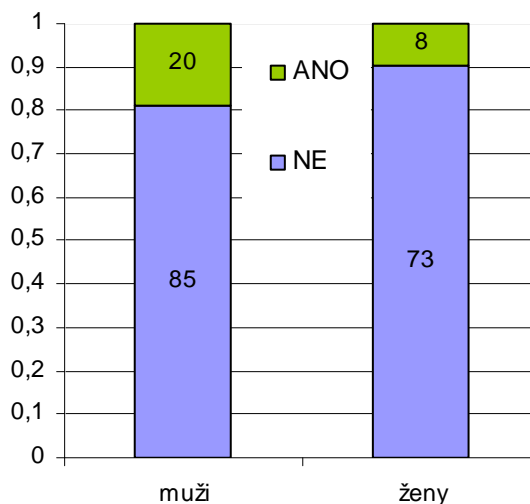
4. Protože se hodnota testového kritéria v kritickém oboru realizovala, přijímáme alternativní hypotézu H_1 tedy, že pohlaví ovlivňuje poskytování přístupu k PC i lidem mimo rodinu.

Pro zjištění těsnosti této závislosti jsme vypočítali Cramérův koeficient kontingence $V=0,25$, což značí, že závislost je slabší.

Soukromí si uživatelé nastavují rovněž bez rozdílu pohlaví. Telefon uvádějí častěji muži než ženy, stejně tak své inkriminující fotografie viz. graf:



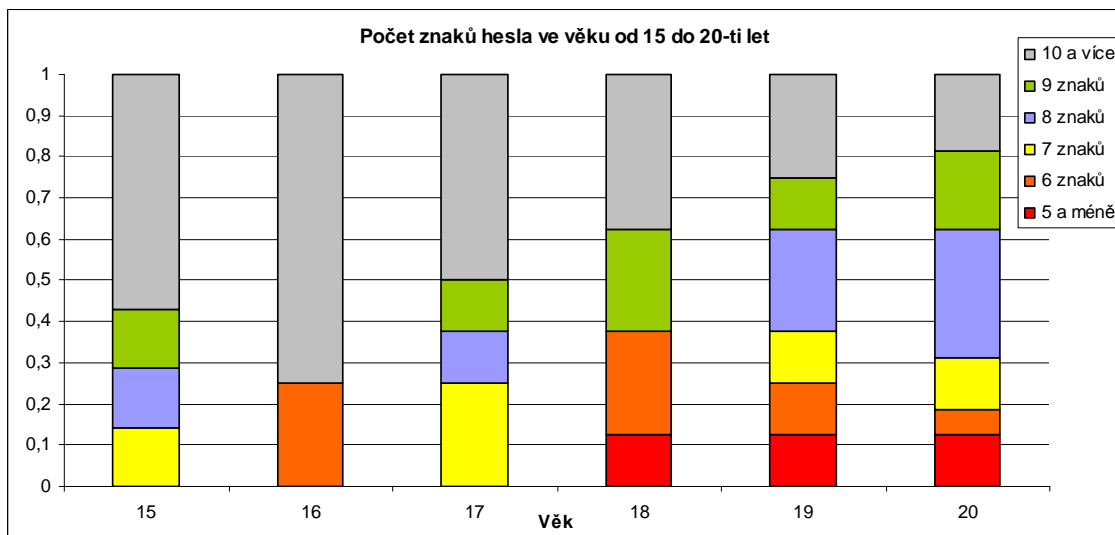
Graf 10; Uvádění telefonu - muži vs. ženy



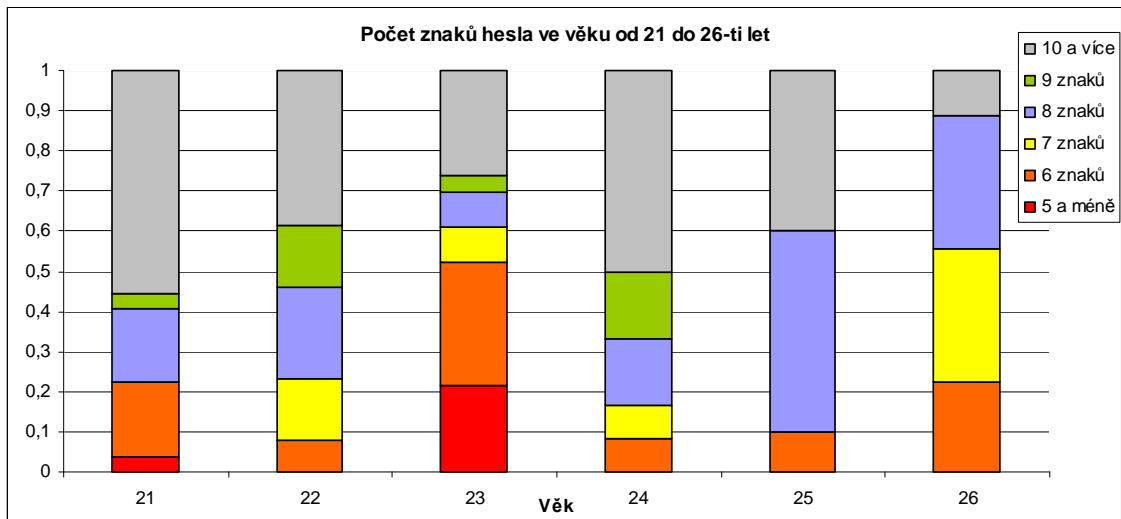
Graf 51; Uvádění ink. fotek - muži vs. ženy

Skupina 15-20 let a její srovnání

Dle policie ČR je nejrizikovější skupinou uživatelů internetu skupina ve věku 15-20 let. Tuto skupinu jsem se rozhodl srovnávat s druhou nejpočetnější skupinou v mém dotazníku, tj skupinou ve věku 21-26 let. Popularita délky hesel žádnou závislost, nebo snad bezpečnější chování (týkající se autentizace) u starších ani mladších uživatelů neodhalila, viz. tabulky dále:

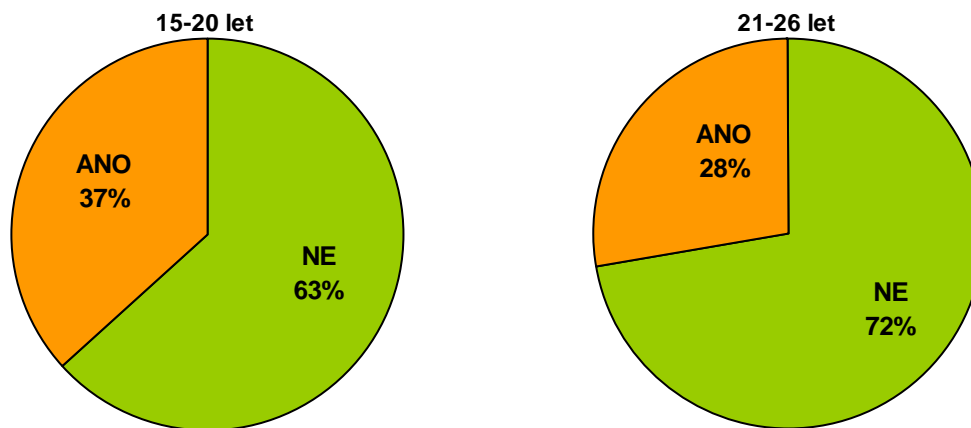


Graf 12; síla hesla 15-20 let



Graf 13; síla hesla 21-26 let

Avšak graf níže, ukazující skutečnost, zda uživatelé někdy přijali přátelství od neznámého člověka naznačuje jistou závislost:



Graf 64; přijali jste někdy přátelství od neznámého člověka?

Existuje závislost na věku a přijetí tzv. nabídky na přátelství od neznámého člověka?

1. Testujeme hypotézu: existuje závislost mezi věkovou skupinou a zda někdy uživatelé přijali tzv. nabídku na přátelství od neznámého člověka

H_0 : Neexistuje závislost mezi věkovou kategorií a přijmutí přátelství od neznámého člověka.

H_1 : Závislost mezi věkovou kategorií a přijmutí přátelství od neznámého člověka existuje.

2. Hodnota testového kritéria $\chi^2 = 2,27$

3. Kritický obor $W_\alpha = \{\chi^2 : \chi^2 \geq 3,841\}$,

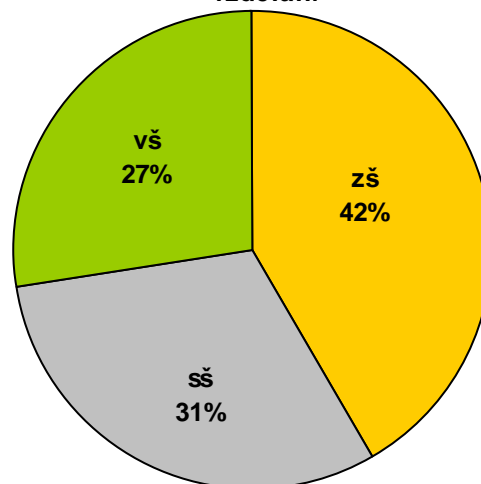
4. Protože se hodnota testového kritéria v kritickém oboru nerealizovala, přijímáme nulovou hypotézu H_0 tedy, že věková kategorie neovlivňuje skutečnost, zda uživatelé někdy přijali nabídku na přátelství od neznámého člověka.

Tato závislost nebyla potvrzena, proto jsem se rozhodl testovat závislost znaku zda přijal respondent někdy nabídku na přátelství od neznámého uživatele na vzdělání i současně studujícím.

Test závislosti vzdělání na přijetí přátelství od neznámého člověka

Graf rozdělení uživatelů, kteří někdy přijali nabídku na přátelství od neznámého člověka a vzdělání zde:

Poměry uživatelů, kteří přijali nabídku na přátelství od neznámého člověka, dle vzdělání



Graf 15; Vzdělání vs. přijetí přátelství neznámého člověka

Samotný graf naznačuje, že s rostoucím vzděláním klesá tendence přijímat nabídky na přátelství od neznámých lidí. Test této závislosti:

1. Testujeme hypotézu: existuje závislost mezi vzděláním a skutečností zda někdy uživatelé přijali tzv. nabídku na přátelství od neznámého člověka

H_0 : Neexistuje závislost mezi vzděláním a přijmutí přátelství od neznámého člověka

H_1 : Závislost mezi věkovou kategorií a přijmutí přátelství od neznámého člověka existuje.

2. Hodnota testového kritéria $\chi^2 = 30,10$

3. Kritický obor $W_\alpha = \{\chi^2 : \chi^2 \geq 5,991\}$

4. Protože se hodnota testového kritéria v kritickém oboru realizovala, přijímáme alternativní hypotézu H_0 tedy, že vzdělání s 5% rizikem omylu ovlivňuje přijetí přátelství od neznámého člověka.

Pro úplnost uvádím tabulku, ze které byly výpočty provedeny:

Cramérův koeficient kontingence $V=0,44$ nám říká, že tato závislost je středně slabá.

Vzdělání	Přijetí přátelství od neznámého	
	ne	ano
zš	14	13
sš	76	42
vš	28	13
celkem	118	68

Tabulka 29; Vzdělání vs. přijetí přátelství neznámého

Test závislosti vzdělání na užívání funkce automatického zapamatování hesla

Při testování úrovně vzdělání používání funkce automatického zapamatování hesla vypadá graf v procentech naprosto shodně, otestováním této zjevné závislosti se nám hodnota testového kritéria s hodnotou 30,1 v kritickém oboru opět realizovala, přičemž nulová hypotéza zněla, že tato závislost neexistuje. Přijali jsme tedy hypotézu alternativní, ze které plyne, že s rostoucím vzděláním klesá užívanost funkce automatického zapamatování hesla.

Testy závislosti vzdělání na udávání telefonu a adresy v profilu na sociální síti.

Oba testy potvrdily hypotézu, že existuje závislost mezi vzděláním a udáváním těchto citlivých informací v profilech uživatelů.

Uživatelé uvádějící telefon byly z 48% uživatelé se základním vzděláním, 30% se středoškolským vzděláním a 22% s vysokoškolským vzděláním. Hodnota testového kritéria byla 20,11 a Cramérův koeficient kontingence 0,33.

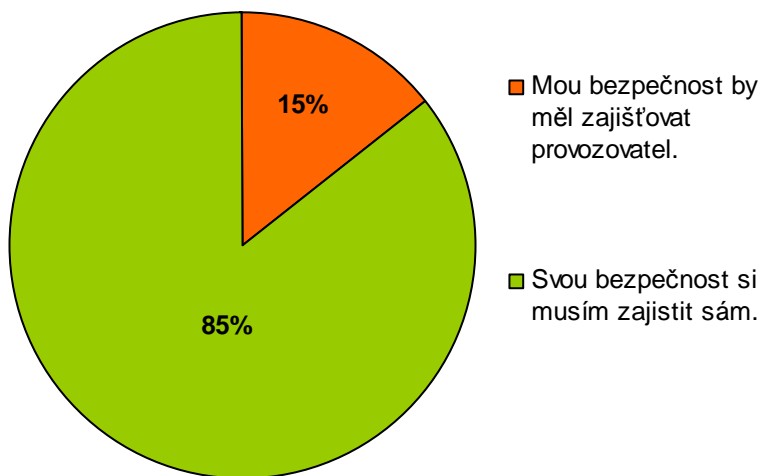
Uživatelé uvádějící svou adresu byly z 29% uživatelé se základním vzděláním, 45% se středoškolským vzděláním a 26% s vysokoškolským vzděláním. Hodnota testového kritéria byla 25,80 a Cramérův koeficient kontingence 0,37.

2.8.11 Motivační otázky a znaky mimo kategorie:

Pro mne zajímavými ukazateli byly i tzv. motivační otázky, které měly respondenty navadit k zodpovězení celého dotazníku. Osobně považuji za nejzajímavější otázku

s kterým ze dvou tvrzení se uživatelé ztotožňují nejvíce, zda by jejich bezpečnost měl zajišťovat provozovatel, nebo zda si svou bezpečnost musí zajistit sami.

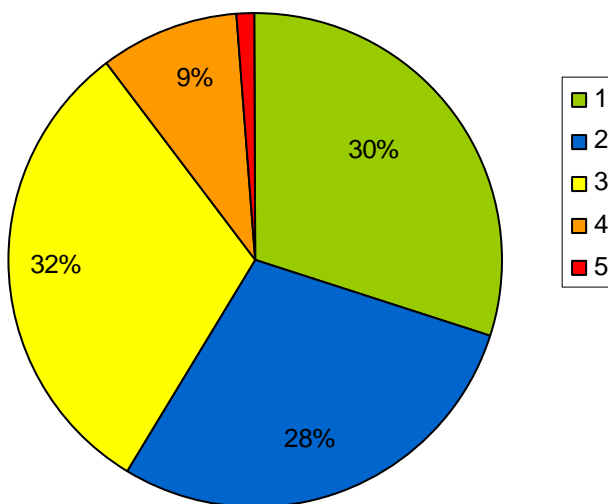
S jakým názorem se uživatelé stotožňují spíše:



Graf 76; s jakým tvrzením se ztotožňujete spíše

Vůbec první otázkou, která měla respondenty navnadit, byl požadavek, aby své znalosti informačních technologií ohodnotili známkou jako ve škole:

Oznámkujte své znalosti IT jako ve škole



Graf 87; oznámkujte své znalosti IT

Poslední (a jedinou otevřenou) otázkou dotazníku nebyla ve skutečnosti otázka, ale pokus o znázornění klasické sociotechnicky jak získat spoustu emailů, které lze potenciálně zneužít např. k spamu v podobě nevyžádané reklamy (v lepším případě). Ze 186 respondentů svůj email do kolonky s popisem „V případě zájmu o výsledky výzkumu mi napište svůj email:“ vložilo 69, tzn. 37% respondentů. Jelikož se jedná o

seriózní výzkum, emaily nebudou zneužity a výsledky snad respondentům i distribuovány.

3 Zhodnocení

Dotazníku se zúčastnilo 186 respondentů, z toho 105 mužů a 81 žen. Průměrný věk respondenta byl 23 (22,96 zaokrouhlo na dvě desetinná místa) let. Nejmladším respondentem byla 11-ti letá dívka a nejstarším 65 letá žena.

Výzkum potvrdil, že věk českých uživatelů je nižší, než u celosvětových statistik.

Údajně nejohroženější skupina ve věku 15-20 let se v části věnující se autentizaci - tedy síle hesel a možnosti přihlášení se pod uživatelským účtem někým jiným – neukázala nijak odlišná od jiných věkových kategorií – avšak skutečnost, že v části věnující se důvěřivosti uživatelů 37% respondentů ve věku od 15-20 let přijalo tzv. „nabídku na přátelství“ od neznámého člověka, je hrozivá. Ve skupině 21-26 let takovou nabídku přijalo od 9% méně respondentů.

Co se tzv. nabídky na přátelství od neznámého člověka a vzdělání týče, výzkum prokázal závislost na vzdělání, přičemž nejhůře si vedli studenti základních škol a lidé se základním vzděláním. Stejně tak u používání funkce pro automatické zapamatování hesla.

37% respondentů v závěru dotazníku poskytlo svůj email a 15% se ztotožňuje s tvrzením, že jejich bezpečnost by měl zajišťovat provozovatel.

3.1 Shrnutí nejvýraznějších charakteristik

3.1.1 Autentizace

- Nejužívanější je heslo délky 10 a více, užívá jej 39,78% respondentů.
- Nejužívanějšími znaky je kombinace písmen a číslic, tuto kombinaci používá 59,14% respondentů.
- Nevhodná slabá hesla, kdy se uživatelské heslo skládá pouze z číslic, nebo pouze s písmen, 22,58% respondentů.
- Slabá hesla délky 6 a kratší užívá 22,04% respondentů.
- Hesla, která můžeme považovat za silná používá 77,96% respondentů.
- 12% uživatelů používá funkci automatického zapamatování hesel na svém počítači, ačkoliv k němu dávají přístup i osobám mimo rodinu.

3.1.2 Dostupnost dat a důvěřivost uživatelů

- 36,55% respondentů někdy přijalo nabídku na přátelství od neznámého člověka
- Avšak 51,61% respondentů nabídku na přátelství od neznámého člověka nepřijala a současně si je vědoma informací, jaké je možno o nich naleznout na internetu a současně si upravují nastavení soukromí na sociální síti.
- Jasný výsledek měla otázka, zda jsou si respondenti vědomi zneužívání sociálních skupin (ať už samotných, nebo dílčích skupin), 92% respondentů si této skutečnosti je vědoma.
- 22% uživatelů vůbec netuší jaké informace se o nich dají na internetu dohledat.

3.1.3 Citlivost údajů

- Na 32% respondentů existují na sociálních sítích fotky, které by je mohly nějakým způsobem inkriminovat např. v zaměstnání. Pouhých 15,05% je sdílí samo a 30,65% respondentů přiznává, že takové fotky o nich sdílí přátelé.
- Svůj telefon v na sociální síti uvádí 16,67% respondentů, svou adresu 14,52%.
- Adresu včetně telefonu sdílí 4,84% respondentů.

3.1.4 Osobní údaje

- Dotazník zodpovědělo zhruba 62% studentů, z toho bylo 58,12% středoškoláků, 18,8% vysokoškoláků a zbylých 23% studovalo základní školu.
- Průměrný věk respondenta byl 22,96 let
- Střední hodnota věku uživatelů sociálních sítí je s 95% spolehlivostí pokryta intervalem od 22 do 23,91 let
- Směrodatná odchylka věku uživatelů sociálních sítí je s 95% spolehlivostí pokryta intervalem od 6 do 7,35 let.

3.1.5 Ostatní

- Známkami 4 a 5 se ohodnotilo celkem 10% respondentů.
- Svůj email poskytlo 37% respondentů
- Čeští uživatelé jsou mladší než celosvětový průměr
- Závislosti na věku se nepotvrdily
- Potvrdily se některé závislosti na vzdělání

Závěr a návrhy

Můj sociologický výzkum znázornil mnoho zajímavých poměrů mezi českými uživateli sociálních sítí, nejenže potvrdil odlišnost českých uživatelů od celosvětových statistik – zejména průměrným věkem, ale také zjistil závislost mezi vzděláním a poskytováním adresy a telefonu v profilu na sociální síti. Stejně tak potvrdil závislost přijetí přátelství od neznámého člověka na vzdělání. Naopak závislost přijetí takové nabídky na věku byla vyvrácena. Prokázané závislosti dávají za pravdu všem skeptikům, kteří hlásají po změně výuky informatiky už na základních školách. Ani v osnovách pro rozšířenou výuku informatiky na základních školách nenaleznete něco jako „bezpečnost chování na internetu“, jediný termín týkající se bezpečnosti z hlediska poskytovaných informací bylo konstatování „*co by měl žák umět: je seznámen s faktem, že informační média mohou obsahovat i eticky závadné informace.*“ (MŠMT ČR, 2001, s. 4)

Evropská komise sice volá po zpřísnění bezpečnostních standardů týkajících se internetu a sociálních sítí, nicméně všechny tyto návrhy směřují spíše k „tlačítkům nahlášení závadného obsahu na webových stránkách“ nikoliv k samotnému chápání možností, jak lze získat zdánlivě soukromé a osobní informace a už vůbec ne k chápání možných důsledků špatného sebe zabezpečení se.

Sociologický výzkum by měl spíš poskytovat informace o skutečnostech a nevynášet soudy, ty necht' vynášejí jiní a k tomu jim právě sociologický výzkum má sloužit; jako nástroj pro podporu rozhodování. Význam statistiky jednotlivých měřených znaků má zajisté pro různé obory různý (resp. jinak silný) význam.

Osobně hodnotím výsledky výzkumu spíš kladně, má očekávání byla horší, zejména co se autentizace týče. Výsledky v dostupnosti a důvěřivosti uživatelů jsou však horší. Stejně jako skutečnost, že stále značná část uživatelů (15%) chápe bezpečnost jako něco, co by měl zajistit provozovatel poskytnuté služby.

Přál bych si změnit uvažování lidí, kteří žijí v domněnce, že internet je stále něco abstraktního, co nezasahuje do jejich životů. Pravda je opačná a krutá. Personalisté už nechtou životopisy, ale přinejmenším „googlí“ potenciální zaměstnance, kteří se pak třeba ani nedoví, proč ve skutečnosti nebyli přizváni k pracovnímu pohovoru. Ano, data na internetu se někdy těžko udržují skrytě a na jednom místě, obzvláště při takovém nástupu velice efektivních nástrojů nejen pro komunikaci a práci, zejména teamovou

spolupráci. Ale námaha, vynaložená na důkladné rozmyšlení si, jakou informaci a kam vložím, je tisíckrát menší, než námaha s jakou se pak snaží uživatelé svá data získat zpět, nebo je znovu ukryt.

Osobně bych si přál, aby můj výzkum (a jemu podobné) inspirovaly nejen společnosti zabývající se bezpečností, ale zejména pedagogy a uživatele samotné.

Použitá literatura

- [1] ČIHOVSKÝ, Jaroslav. *Sociologický výzkum : Studijní text pro posluchače FTK UP Olomouc*. Olomouc : [s.n.], 2006. 41 s. Dostupný z WWW: <www.upol.cz/fileadmin/user_upload/FTK-dokumenty/Katedra_rekreologie/Sociologicky_vyzkum_def_1_.doc>.
- [2] JEŘÁBEK, Hynek. *Úvod do sociologického výzkumu*. 1. vyd. Praha : Karolinum, 1993. 162 s. ISBN 80-7066-662-5. Dostupný z WWW: <<http://www.ftvs.cuni.cz/hendl/jerabek3/jerabek.htm>>.
- [3] CHROBOČEK, M. *Případové studie pro statistickou analýzu dat*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 84 s. Vedoucí diplomové práce Ing. Radovan Novotný, Ph.D.
- [4] KLIMEŠ, Lumír. *Slovník cizích slov*. Praha : SPN, 1981. 791 s.
- [5] KROPÁČ, Jiří. *Statistika B : Jednorozměrné a dvourozměrné datové soubory, Regresní analýza, Časové řady*. 2. dopl. vyd. Fakulta podnikatelská, VUT v Brně. 151 s. ISBN 978-80-214-3295-6.
- [6] NOVOTNÁ, Eliška. *Základy sociologie*. 1. vyd. Havlíčkův Brod : Grada Publishing, 2008. 192 s. ISBN 978-80-247-2396-9.
- [7] NOVOVIČOVÁ, Jana. *Pravděpodobnost a matematická statistika : Skriptum Fakulty dopravní CVUT*. Praha : [s.n.], 1999. 157 s. Dostupný z WWW: <<http://euler.fd.cvut.cz/publikace/files/skripta3.pdf>>.
- [8] PECÁKOVÁ, Iva. *Testy nezávislosti v řídkých kontingenčních tabulkách* [online]. [2007] [cit. 2010-01-05]. Dostupný z WWW: <<http://panda.hyperlink.cz/cestapdf/pdf07c1/pecakova.pdf>>.
- [9] REICHEL, Jiří. *Kapitoly metodologie sociálních výzkumů*. Praha : Grada Publishing a.s., 2009. 192 s. ISBN 978-80-247-30066.
- [10] ZICH, František. *Sociologický výzkum*. 1. vyd. Praha : Svoboda, 1976. 266 s.
- [11] DISMAN, Miroslav. *Jak se vyrábí sociologická znalost ; Příručka pro uživatele*. 3. vydání. Praha : Karolinum, 2009. 372 s. ISBN 978-80-246-0139-7.
- [12] press.linkedin.com [online]. 2009 [cit. 2010-05-07]. Company History. Dostupné z WWW: <<http://press.linkedin.com/history>>.

- [13] *Ct24.cz* [online]. 17. 3. 2010 [cit. 2010-05-07]. Italský mafiánský boss se prozradil díky Facebooku. Dostupné z WWW: <<http://www.ct24.cz/svet/84148-italsky-mafiansky-boss-se-prozradil-diky-facebooku/>>.
- [14] KRATOCHVÍL, Luboš. *Technet.idnes.cz* [online]. 19. 2. 2010 [cit. 2010-05-07]. Facebook a Twitter mohou prozradit zlodějům, kdy nejste doma. Dostupné z WWW: <http://technet.idnes.cz/facebook-a-twitter-mohou-prozradit-zlodejum-kdy-nejste-doma-pl5-/sw_internet.asp?c=A100219_111325_sw_internet_pka>.
- [15] *Facebook.com* [online]. 2010 [cit. 2010-05-07]. Tiskové středisko. Dostupné z WWW: <<http://www.facebook.com/press.php>>.
- [16] *Twitter.com* [online]. 2010 [cit. 2010-05-07]. Twitter blog. Dostupné z WWW: <<http://blog.twitter.com/>>.
- [17] *Policie.cz* [online]. 2010 [cit. 2010-05-07]. Aktuality: (NE)bezpečný Internet. Dostupné z WWW: <<http://www.policie.cz/clanek/ne-bezpecny-internet.aspx>>.
- [18] *Evropská komise* [online]. 2010 [cit. 2010-05-07]. Tiskové centrum. Dostupné z WWW: <http://ec.europa.eu/ceskarepublika/press/index_cs.htm>.
- [19] ULLMAN, Larry; POŠMURA, Vlastimil. *PHP a MySQL : Náзорný průvodce tvorbou dynamických WWW stránek*. Brno : Computer Press, 2004. 534 s. ISBN 80-251-0063-4.
- [20] MITNICK, Kevin. *Umění klamu*. Gliwice : Helion, 2003. 348 s. ISBN 83-7361-210-6.
- [21] *Inflow.cz* [online]. 2010 [cit. 2010-05-01]. Studie: Věk uživatelů sociálních sítí. Dostupné z WWW: <<http://www.inflow.cz/studie-vek-uzivatelu-socialnich-siti>>.
- [22] *MŠMT ČR* [online]. 1. září 2005 [cit. 2010-05-10]. Učební osnova předmětu INFORMAČNÍ A KOMUNIKAČNÍ TECHNOLOGIE. Dostupné z WWW: <<http://aplikace.msmt.cz/PDF/AKICTproSOUrevUOUupr.pdf>>.
- [23] *MŠMT ČR* [online]. 2001 [cit. 2010-05-10]. Doplnění základní pedagogické dokumentace vzdělávacího programu “Základní škola”. Dostupné z WWW: <aplikace.msmt.cz/DOC/Upravy_VP_ZS.doc>.

Seznam obrázků a grafů

GRAF 1; DÉLKA, SKLADBA A POPULARITA HESEL

39

GRAF 2; SLOUPCOVÝ DIAGRAM - AUTOMATICKÉ ZAPAMATOVÁNÍ HESLA	39
GRAF 3; KRUHOVÝ VÝSEČOVÝ DIAGRAM	39
GRAF 4; ZNEUŽÍVÁNÍ SKUPIN	43
GRAF 5; NEVHODNÉ FOTKY SDÍLEJÍ VS. JSOU O NICH SDÍLENY – V PROCENTECH	45
GRAF 6; VĚKOVÉ ROZDĚLENÍ	47
GRAF 6; KOMBINACE ZNAKŮ A POHLAVÍ	49
GRAF 7; POČET ZNAKŮ A POHLAVÍ	49
GRAF 8; MÁ PŘÍSTUP K POČÍTAČI I NĚKDO JINÝ?	50
GRAF 9; UVÁDĚNÍ TELEFONU - MUŽI VS. ŽENY	51
GRAF 10; UVÁDĚNÍ INKR. FOTEK - MUŽI VS. ŽENY	51
GRAF 11; SÍLA HESLA 15-20 LET	51
GRAF 12; SÍLA HESLA 21-26 LET	52
GRAF 13; PŘIJALI JSTE NĚKDY PŘÁTELSTVÍ OD NEZNÁMÉHO ČLOVĚKA?	52
GRAF 14; VZDĚLÁNÍ VS. PŘIJETÍ PŘÁTELSTVÍ NEZNÁMÉHO ČLOVĚKA	53
GRAF 15; S JAKÝM TVRZENÍM SE ZTOTOŽŇUJETE SPÍŠE	55
GRAF 16; OZNÁMKUJTE SVÉ ZNALOSTI IT	55

Seznam tabulek

TABULKA 1; KONTINGENČNÍ TABULKA	17
TABULKA 2; POČET ZNAKŮ HESLA	37
TABULKA 3; DRUHY ZNAKŮ V HESLE	37
TABULKA 4; AUTOMATICKÉ ZAPAMATOVÁNÍ HESLA	38
TABULKA 5; UŽÍVÁNÍ RESPONDENTOVA PC I NĚKÝM JINÝM	38
TABULKA 6; DÉLKA HESLA - KVANT. CHARAKTERISTIKY	38
TABULKA 7; ČETNOSTÍ DÉLKY HESLA A POUŽITÝCH ZNAKŮ	40
TABULKA 8; TABULKA RELATIVNÍCH ČETNOSTÍ DÉLKY HESLA A POUŽITÝCH ZNAKŮ	40
TABULKA 9; KONTINGENČNÍ TABULKA ČETNOSTÍ ZNAKŮ G A D	41
TABULKA 10; ČTYŘPOLNÍ TABULKA REL. ČETNOSTÍ G A D V PROCENTECH	41
TABULKA 11; ÚPRAVA NASTAVENÍ SOUKROMÍ	41
TABULKA 12; PŘIJETÍ PŘÁTELSTVÍ OD NEZNÁMÉHO ČLOVĚKA	41
TABULKA 13; VÍ, CO SE O NĚM/NÍ DÁ DOHLEDAT NA INT.	41
TABULKA 14; VÍTE O ZNEUŽÍVÁNÍ SKUPIN	42
TABULKA 15; ZNAKY UŽIVATELŮ NENASTAVUJÍCÍCH SI SOUKROMÍ - V PROCENTECH	42
TABULKA 16; ZNAKY UŽIVATELŮ NASTAVUJÍCÍCH SI SOUKROMÍ - V PROCENTECH	42
TABULKA 17; ADRESA V PROFILU NA SOCIÁLNÍ SÍTI	43
TABULKA 18; TELEFON V PROFILU NA SOCIÁLNÍ SÍTI	43
TABULKA 19; SDÍLÍ VE SVÉM PROFILU INKRIMINUJÍCÍ FOTKY	43
TABULKA 20; SDÍLÍ VAŠI PŘÁTELE INKRIMINUJÍCÍ FOTKY	44
TABULKA 21; ÚDAJE UŽIVATELŮ NESDÍLEJÍCÍCH NEVHODNÉ FOTKY V PROCENTECH	44

TABULKA 22; ÚDAJE UŽIVATELŮ SDÍLEJÍCÍCH NEVHODNÉ FOTKY V PROCENTECH	44
TABULKA 23; ČTYŘPOLNÍ TABULKA O SDÍLENÍ FOTEK V PROCENTECH	44
TABULKA 24; POHLAVÍ	45
TABULKA 25; VZDĚLÁNÍ	45
TABULKA 26; STAV	46
TABULKA 27; VĚK RESPONDENTŮ	46
TABULKA 28; ZNAKY STAV A VZDĚLÁNÍ	46
TABULKA 29; VZDĚLÁNÍ VS. PŘIJETÍ PŘÁTELSTVÍ NEZNÁMÉHO	54

Přílohy

Přílohy byly umístěny na webových adresách z důvodu jejich přílišné obsáhlosti.

Zdrojový kód dotazníku dostupný na adrese: <http://dotaznik.php5.cz/kod.txt>

Obsah ošetřeného souboru data.dat, použitého v této práci (odstraněny emailové adresy respondentů) dostupný na adrese: <http://dotaznik.php5.cz/data.txt>