

UNIVERZITA PALACKÉHO V OLMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

## DIPLOMOVÁ PRÁCE

Zátěžový model dotazníkového šetření na  
respondenty



Vedoucí bakalářské práce: **Doc. RNDr. Karel Hron, PhD.**

Vypracovala: **Bc. Edita Mojzíková**

Studijní program: B1103 Aplikovaná matematika

Studijní obor: Aplikace matematiky v ekonomii

Forma studia: prezenční

Rok odevzdání: 2015

## BIBLIOGRAFICKÁ IDENTIFIKACE

**Autor:** Bc. Edita Mojžíková

**Název práce:** Zátěžový model dotazníkové šetření na respondenty

**Typ práce:** Diplomová práce

**Pracoviště:** Katedra matematické analýzy a aplikací matematiky

**Vedoucí práce:** Doc. RNDr. Karel Hron, Ph.D.

**Rok obhajoby práce:** 2015

**Abstrakt:** Diplomová práce svým obsahem vychází z dotazníkového šetření Životní podmínky, které na našem území každoročně provádí Český statistický úřad. Je studií vybraných typů otázek, které jsou v dotaznících vůči respondentovi obsaženy, s důrazem na časovou zátěž otázek. Cílem práce je tak najít, jakému rozdělení odpovídá skutečné rozdělení pravděpodobnosti časové zátěže. Vycházet přitom budeme z reálných časových schémat otázek, která byla Českým statistickým úřadem poskytnuta. Pro tyto účely budou otázky popsány a zároveň vysvětleny chí-kvadrát testy dobré shody, které budou při testování použity. Výsledkem by měl být obecný model použitelný při tvorbě dotazníkové šetření s možností odhadnout časovou zátěž přidáním dodatečné proměnné.

**Klíčová slova:** Dotazníkové šetření, časová zátěž, rozdělení pravděpodobnosti, testy dobré shody.

**Počet stran:** 78

**Počet příloh:** 4

**Jazyk:** český

## BIBLIOGRAPHICAL IDENTIFICATION

**Author:** Bc. Edita Mojžíková

**Title:** Exercise test of questionnaire survey on respondents

**Type of thesis:** Diploma

**Department:**

Department of Mathematical Analysis and Application of Mathematics

**Supervisor:** Doc. RNDr. Karel Hron, Ph.D.

**The year of presentation:** 2015

**Abstract:** The thesis is based on questionnaire survey called Living Conditions in the Czech Republic. It is a study of selected questions from the questionnaire with focus on time burden of questions. The aim of the thesis is to find out the real probability distribution of the time burden based on data provided by the Czech Statistical Office. For that reason we will present the questions and also chi square goodness of fit tests that are applied for the purpose of testing. The results could be used for a general model to estimate time burden of the questionnaire survey by providing an additional variable.

**Key words:** Questionnaire survey, time burden, probability distribution, chi square goodness of fit test.

**Number of pages:** 78

**Number of appendices:** 4

**Language:** Czech

### **Prohlášení**

Prohlašuji, že jsem diplomovou práci zpracovala samostatně pod vedením pana Doc. RNDr. Karla Hrona, Ph.D. a že jsem v seznamu literatury uvedla všechny použité zdroje.

V Olomouci dne 3. dubna 2015.

## Poděkování

Na tomto místě bych ráda poděkovala svému vedoucímu diplomové práce panu Doc. RNDr. Karlu Hronovi, Ph.D. a konzultantům z Českého statistického úřadu, zejména Ing. Táně Dvornákové, za cenné rady a podněty, které vedly k úspěšnému vypracování práce. Děkuji typografickému programu  $\text{\TeX}$ , který dal za vznik vizuální stránce práce, a statistickému softwaru R, který umožnil zpracování rozsáhlého souboru dat a veškeré matematické výpočty. Velký dík ale patří zejména mé rodině a mým blízkým, protože mi při studiu a vypracování práce byli oporou. Děkuji jim zejména za vytrvalou psychickou podporu a materiální zajištění.

# Obsah

Úvod	7
<b>1 Šetření Životní podmínky</b>	<b>9</b>
<b>2 Analýza proměnných</b>	<b>13</b>
2.1 Typy proměnných . . . . .	13
2.1.1 Objektivní vs. subjektivní . . . . .	13
2.1.2 Výběrové otázky . . . . .	14
2.1.3 Vypisovací otázky . . . . .	14
2.1.4 Škálové otázky . . . . .	15
2.2 Analytická část . . . . .	16
2.2.1 Alternativní otázky . . . . .	17
2.2.2 Výběrové otázky . . . . .	19
2.2.3 Vypsání textového údaje . . . . .	20
2.2.4 Vypsání číselného údaje . . . . .	21
2.2.5 Otázky na částku . . . . .	24
2.2.6 Škálové otázky . . . . .	25
<b>3 Teoretická rozdělení pravděpodobnosti</b>	<b>27</b>
3.1 Normální rozdělení . . . . .	27
3.2 Logaritmicko-normální rozdělení . . . . .	31
3.3 Exponenciální rozdělení . . . . .	34
3.4 Gama rozdělení . . . . .	36
3.5 Weibullovo rozdělení . . . . .	39
<b>4 Testování shody rozdělení</b>	<b>42</b>
4.1 Metoda maximální věrohodnosti . . . . .	42
4.2 Multinomické rozdělení . . . . .	45
4.3 Testy dobré shody při neznámých parametrech . . . . .	47
4.4 Ověřování normálního rozdělení . . . . .	50
<b>5 Praktická část</b>	<b>54</b>
5.1 Testování otázky A6_UK_VZD . . . . .	55
5.2 Testy ostatních proměnných . . . . .	61
5.3 Statistika $\phi$ . . . . .	67
5.4 Odhady parametrů lognormálního rozdělení . . . . .	68
<b>Závěr</b>	<b>74</b>
<b>Literatura</b>	<b>76</b>
<b>Přílohy</b>	<b>78</b>

# Úvod

Pro Českou republiku se v souvislosti se vstupem do Evropské Unie stalo povinné pravidelné zjišťování úrovně životních podmínek českých domácností. Od roku 2005 proto Český statistický úřad provádí toto šetření pod názvem Životní podmínky.

Šetření u předem vybrané domácnosti probíhá formou pohovoru tazatele s respondentem. Náplň rozhovoru je dána čtyřmi typy dotazníků, v nichž je obsaženo přes 100 nejrůznějších otázek. V kompetenci tazatele je otázky respondentovi co nejvíce osvětlit tak, aby odpověď na ně měla maximální výpovědní hodnotu a byla co nejpřesnější, a tím i celé šetření bylo kvalitnější.

Pohovor mezi tazatelem a respondentem má předem vymezený časový interval, který by neměl překročit 60 minut. Problémem však je, že tazatelé potřebují k vyšetření jedné domácnosti okolo 90 minut, zvláště je-li tazatel v domácnosti poprvé a musí si rodinu takzvaně „získat“. Náročnější na čas je šetření i ve chvíli, kdy se do dotazníku přidá další otázka. A právě časová zátěž vybraného vzorku proměnných bude předmětem našeho zkoumání.

Nejprve si ale ve čtyřech teoretických kapitolách připravíme půdu pro praktickou část práce. První kapitola bude věnována stručnému popisu šetření Životní podmínky. Půjde nám o to nastínit princip šetření, abychom si udělali obrázek o tom, jak vlastně sběr dat v praxi probíhá. Ve druhé části práce se podrobněji podíváme na nástroj šetření, tím jsou dotazníky, respektive otázky, které jsou v nich obsaženy. Uvedeme zde kritéria, na základě nichž jsme vybrali proměnné pro testování v praktické části práce, a tyto proměnné představíme. Budeme se přitom snažit o komplexní popis proměnné se zaměřením na možné faktory, které mohou mít vliv na časovou zátěž.

Třetí a čtvrtá kapitola se už nese ve statistickém duchu. Ve třetí kapitole provedeme stručný vhled do pravděpodobnostních rozdělení s tím, že si pomocí základních charakteristik představíme pět spojitých rozdělení pravděpodobnosti. A konečně čtvrtá kapitola už bude přechodem k praktické části práce. Představíme zde chí-kvadrát testy dobré shody, kterými budeme zvolená teoretická

rozdělení ověřovat.

Cílem práce tak bude pomocí uvedené statistické inference a na konkrétních datech testovat, kterému z teoretických rozdělení nejvíce odpovídá skutečné rozdělení časové zátěže vybraného modulu otázek. Hlavním cílem pak bude umožnit vytvoření obecného modelu, který bude vyjadřovat závislost mezi typem otázky a časem a bude užitečný jako podklad pro dotazníkové šetření při jeho tvorbě. Tento výstup bude v praxi sloužit zejména metodikům pro odhad časové zátěže celého dotazníku. Důležitá bude také předpovědní funkce modelu pro možnost posouzení zátěže na respondenta přidáním dodatečné proměnné.



# 1 Šetření Životní podmínky

Jak již bylo řečeno v úvodu, Český statistický úřad (dále jen ČSÚ) každoročně organizuje výběrové šetření Životní podmínky. Toto šetření je českou modifikací evropského šetření EU-SILC (European Union-Statistics on Income and Living Conditions), které na svém území aplikují všechny členské země Evropské unie (EU). K šetření se každoročně připojuje také Island, Norsko a Turecko, které členy EU nejsou. V České republice šetření probíhá na základě Nařízení (EC) 1177/2003, v souladu se zákonem č. 89/1995 Sb., o státní statistické službě, ve znění pozdějších předpisů, a při respektování zákona č. 101/2000 Sb., o ochraně osobních údajů. Celá tato kapitola byla vypracována s použitím zdroje [7], [10] a [20].

Podstatou šetření je zjišťování příjmů a životních podmínek domácností. Výstupem jsou informace o sociální situaci v zemi, o životní úrovni obyvatelstva, o jejich majetkové vybavenosti a finanční náročnosti bydlení, o zdravotních podmínkách osob apod. Zjištěná data slouží při zkoumání úrovně peněžní a materiální chudoby nebo při posuzování intenzity ekonomické aktivity. Díky jednotné metodice, kterou státy v šetření používají, můžeme získané údaje mezi sebou porovnávat a hodnotit na celoevropské úrovni.

Šetření probíhá prostřednictvím čtyř typů dotazníků, které uvádíme v příloze. Ucelenou složku šetření tvoří tři dotazníky A, B a C. Dotazník A šetří informace za byt. Do něj se zaznamenávají demografické údaje o osobách žijících v šetřené domácnosti, o jejich společném hospodaření a jejich vzájemné vztahy. Dotazník za hospodařící domácnost je označován jako dotazník B. Do něj se zaznamenávají informace ohledně velikosti bytu, jeho vybavení, pobíraných sociálních dávek nebo např. informace o tom, jak domácnost vychází se svými příjmy. Dotazník C obsahuje informace za osobu týkající se zaměstnání, ekonomické aktivity, příjmů či zdraví. Dotazník C vyplňují osoby starší 16 let.

Čtvrtou, každoročně se měnící složkou šetření, je tzv. modul. Většinou tématicky rozšiřuje některou z oblastí, na niž je šetření zaměřeno. Například materiální deprivace, mezigenerační přenos chudoby, podmínky bydlení, sociální participace,

správa financí domácností, zadluženost a finanční vyloučení. Například pro rok 2013 byl modul zaměřen na téma well-being, kterým se označuje pocit životní pohody a spokojenosti v různých oblastech života.

Jednotkou zjišťování je byt. Jejich výběr probíhá dvoustupňovým náhodným výběrem. V první fázi jsou vybrány sčítací obvody, jakožto nejmenší územní celky v ČR, a ve fázi druhé se z každého obvodu vybírá 10 bytů. Tyto byty musí být jednoznačně určené adresou a číslem domu, případně číslem bytu v domě. Aby bylo zaručeno, že bude tímto výběrem postihnuta celá Česká republika, jsou do výběru zařazeny všechny kraje tak, že rozsah výběru odpovídá velikosti kraje. Do šetření jsou pak zahrnuty všechny osoby, které mají na dané adrese své obvyklé bydliště.

Šetření se skládá z průřezové a panelové části, přičemž panelová složka kombinuje údaje za čtyři roky po sobě prostřednictvím opakovaných každoročních návštěv. Tomuto systému se říká *čtyřletý rotační panel*. Osoby, které byly členy domácnosti vybrané v tzv. první vlně, se nazývají panelové osoby. Ty jsou opakovaně šetřeny, aby mohl být sledován vývoj jejich životních podmínek v čase. Může se stát, že se některá osoba nebo celá hospodařící domácnost během šetření přestěhuje. V takové situaci je dohledána na nové adrese (je-li známá), aby se soubor nezmenšoval. Osoby přistěhované, popř. narozené ve vybrané domácnosti, které žijí s panelovou osobou, jsou považovány za osoby mimo panel. To znamená, jsou šetřeny, ale v případě stěhování nejsou dohledány.

Základní složkou celého šetření jsou tazatelé. Právě oni totiž zajišťují sběr dat v terénu. To znamená, že navštěvují vybrané domácnosti a prostřednictvím dotazníků příslušné informace šetří. Tazatelem mohou být zaměstnanci ČSÚ i externisté. Při své práci jsou vázáni povinností zachovávat mlčenlivost o všech šetřených skutečnostech, což zaručuje přísnou anonymitu zjištěných údajů.

Šetření může probíhat buď formou PAPI nebo CAPI. Zkratka PAPI (Paper Assisted Personal Interviewing) se používá pro šetření za pomoci „tužky a papíru“, kdy s sebou tazatelé do domácností nosí papírové dotazníky. V CAPI šetření (Computer Assisted Personal Interviewing) jsou tazatelé vybaveni table-

tem, notebookem či netbookem, v němž mají nainstalovaný elektronický dotazník. S PAPI dotazníky převážně šetří tazatelé externisté, CAPI dotazník mají k dispozici jen zaměstnanci úřadu. Oba dotazníky jsou svým obsahem téměř totožné. Oba způsoby šetření mají své pro a proti, vlivem technologického pokroku je ale přeci jen v případě CAPI výhod trochu více.

Víceméně jediným problémem elektronického šetření je paradoxně technika sama. S počítači jsou totiž často spojeny technické problémy a komplikace může působit i šetření v hůře dostupných lokalitách. Bezesporu výhodou oproti PAPI je ale rychlé a pohodlné zpracování zjištěných údajů. U papírových dotazníků je totiž právě zpracování výstupu velmi problematické. Archy musejí být opticky snímány, proto je třeba s nimi zacházet opatrně a věnovat velkou pozornost uvedeným údajům, aby byly výrazné a čitelné. Proto také tazatel pro jejich vyplnění používá výhradně psací potřeby k tomuto účelu určené - tenkou černou fixu, resp. černou propisovací tužku. Žádné slabě písíci psací potřeby, ve stylu čínského pera, nejsou povoleny.

Výhodou CAPI dotazníku je i automatické „trasování“, díky němuž je tazatel směřován k dalším otázkám. Tazatel proto v dotazníku nahodile „neskáče“ z jedné otázky ke druhé, ale plynule jím prochází tak, jak mu určí počítač, je tedy automaticky „trasován“ na navazující otázky. Nedochozí tak často k tomu, že by nějaké odpovědi chyběly v důsledku toho, že otázky byly přeskočeny. Stejně tak otázky, které se na základě vyhodnocení předchozích odpovědí netýkají konkrétní domácnosti nebo daného respondenta, jsou automaticky překročeny.

V souvislosti s výhodami CAPI nesmíme opomenout zmínit možnost automatické kontroly vyplněných údajů. *Informativní* kontrola slouží k upozornění tazatele, že jde o neobvyklou situaci. Například respondent uvedl údaj, který se liší v porovnání s předchozím rokem, i přesto, že by měl zůstat neměný. *Závazná* kontrola už poukazuje na závažnou chybu. Upozorňuje tazatele, že zadal hodnotu, která se běžné situaci vymyká. Tuto kontrolu už nelze potvrdit. Jediným řešením je opravit chybný kontrolovaný údaj. Příkladem chybného údaje může být situace, kdy respondent udá u příjmu částku 50 Kč. Upozorněn bude

tazatel i v případě zadání chybného kódu nebo numerické hodnoty v tom smyslu, že neodpovídá povolenému rozmezí. Lze ji opravit jen smazáním chybně zapsané hodnoty a následným zadáním některého kódu z nabízených kategorií.

Nutno podotknout, že role tazatele v šetření je spíše psychologická. Nejen, že musí respondentovi otázky co nejlépe osvětlit a následně do dotazníku zaznamenat všechny zjištěné odpovědi, metodologické průzkumy ukazují, jak moc je důležitá také neverbální komunikace, a podtrhují navození harmonie mezi tazatelem a respondentem. Vedou-li totiž šetření nezkušení tazatelé, může se stát, že věnují kupříkladu přílišnou pozornost počítači a soustředí se na to, aby získané odpovědi správně do počítače zadali. Takové jednání vztahy mezi tazatelem a respondentem oslabuje a může způsobit, že respondent poskytne nežádoucí odpověď nebo celý pohovor odmítne. Tazatelé si proto musejí v domácnosti počínat profesionálně a navodit patřičnou atmosféru, aby se respondent cítil příjemně a měl k tazateli důvěru. Jedině pak bude otevřenější a ochotný poskytnout požadované informace, zejména pak v případě citlivějších otázek.

## 2 Analýza proměnných

V této kapitole zaměříme svou pozornost na otázky, které jsou obsahem dotazníků v šetření Životní podmínky 2013. Zejména se budeme snažit otázky otypovat, abychom mohli proměnné seskupit na základě určitých společných znaků, a následně tyto celé skupiny testovat v praktické části diplomové práce vždy na jednom vybraném zástupci. Uvedli jsme totiž v úvodu, že otázek je v dotaznících obsaženo více jak sto, bylo by proto velmi časově náročné testovat každou zvlášť. Tato kapitola bude tedy sloužit k popisu typů proměnných, jež se v šetření vyskytují, a k charakteristice otázek vybraných pro praktické testování. Kapitola byla zpracována za podpory interního zdroje ČSÚ [7], konzultací na ČSÚ a vzorových dotazníků, které uvádíme v příloze.

### 2.1 Typy proměnných

Ať už si z dotazníku vybereme kteroukoliv z otázek, budeme vždy váhat, je-li objektivní nebo subjektivní, a současně na ni budeme nuceni odpovědět jedním ze čtyř způsobů: výběrem z daného výčtu variant, vypsáním číselného či textového údaje, nebo samotným ohodnocením variant. Jiné typy otázek v dotaznících nenajdeme.

#### 2.1.1 Objektivní vs. subjektivní

*Objektivní* otázka je taková, která vyžaduje objektivní odpověď. Je založena na faktech a zjištěných skutečnostech. Odpověď na ni je věcná, předmětná, shodující se se skutečností. Oproti tomu *subjektivní* otázka vyžaduje odpověď osobní, nevěcnou, předpojatou, založenou na subjektivním mínění a chtění. Odpověď na otázky zpravidla vychází z respondentových osobních názorů, výkladů, úsudků, úhlů pohledů.

Jako příklad objektivní otázky můžeme uvést: „Vlastní vaše rodina osobní automobil?“ nebo „Kolik obytných místností vaše rodina využívá?“. Obě dvě otázky vyžadují věcnou odpověď. Rodina buď automobil má nebo nemá, stejně tak počet obytných místností je daný. Otázka je objektivní právě v tom, že

nezávisí na postoji a názoru dotazovaného, ale shoduje se se skutečností. Zatímco u subjektivní otázky neexistuje všeobecně správná odpověď, ale je závislá na úsudku konkrétního respondenta. Taková je třeba otázka „Ohodnoťte, prosím, svůj celkový zdravotní stav.“

### 2.1.2 Výběrové otázky

Výraz *výběrová* otázka budeme používat právě pro ty proměnné, na něž je odpověď v podobě výběru z několika variant.<sup>1</sup> V marketingových průzkumech jde o nejpoužívanější typ proměnné a je obecně známý pod názvem uzavřená otázka. Je charakteristická výčtem alternativ možných odpovědí, ze kterých respondent vybírá jednu nebo více preferovaných variant. Odpovídající odpověď zvolí buď označením křížkem, nebo vypsáním kódu příslušné odpovědi.

Výhodou tohoto typu otázek je bezpochyby jejich jednoduchost, a to jak pro respondenta, tak pro zadavatele. Respondentovi se takové otázky lehce zodpovídají, zadavateli snadno zpracovávají. Jsou však poněkud náročnější na přípravu. Navržené alternativy zde vymezují určitý rámec odpovědí, o kterých chce tazatel získat představu. Jednoho respondenta mohou inspirovat, rozšířit mu jeho obzor zkušeností, a poskytnout tak odpovědi, na které by třeba sám nepřišel. Jiného respondenta však mohou svazovat a bránit mu v jeho iniciativě, nemusí zkrátka nutně vystihnout jeho názor.

Nejjednodušším typem výběrové otázky je tzv. *alternativní otázka* nebo někdy používáme pojem *dichotomická otázka*. Respondent má na výběr pouze ze dvou odpovědí. Nejčastěji volba je mezi ano - ne, muž - žena, má - nemá.

### 2.1.3 Vypisovací otázky

Protipólem k výběru z navržených variant je odpověď formou vypsání textového nebo číselného údaje. Otázka je odlišná v tom, že vybízí respondenta

---

<sup>1</sup>Kdybychom chtěli být přesní, výběrová otázka je výraz používaný, vybíráme-li jen jednu ze zvolených variant. Jestliže označíme odpovědi více, jde o otázku výčtovou. V textu však budeme pro jednoduchost uvažovat pojem *výběrová otázka* všude tam, kde vybíráme z variant, ať už jednu nebo více.

k vlastní samostatné odpovědi, ten tedy nemá k dispozici žádné alternativy odpovědí.

V případě *vypsání textového údaje* se v průzkumech často hovoří jako o tzv. otevřené otázce. Obecně jsou vyhledávány pro svou originalitu výstupu a volnost při odpovídání. Dotazovaný má prostor k vyjádření svého vlastního názoru a může se hlouběji nad tématem zamyslet. Je tak zamezeno frustraci, která může být vyvolána omezených výčtem alternativ. Výhodou jsou také podrobnější a rozsáhlejší odpovědi, což vede k předpokládanému kvalitnímu výstupu. Problémem však mohou být nestandardizované a nepředpokládané odpovědi. Řada respondentů může mít totiž s vlastní formulací odpovědi problém, může pro ně být problematické utříbit si myšlenky a výstižně a zřetelně zformulovat svůj názor. I zpracování dat je více problematické, protože je navíc třeba provést kategorizaci, která převede otázku na uzavřenou. V šetření Životní podmínky se takový typ otázky vyskytuje jen ve dvou případech, navíc nejsou zcela typické. Po respondentovi se nechtějí žádné delší výpovědi, je vybízen ke krátké a víceméně heslovité odpovědi, proto by formulace odpověď zas takové problémy činit neměla.

Otázka vyžadující odpověď formou *vypsání číselného údaje* vyžaduje odpověď číslicí. Může se jednat o otázku, kolik hodin týdně respondent odpracuje ve svém zaměstnání, nebo jaký je rok dokončení nejvyššího vzdělání. Otázka tedy nenabízí výběr z variant odpovědí, neřadí se ani mezi ryze otevřené otázky.

Pro pozdější účely budeme v otázce vyžadující odpověď číslicí rozlišovat *otázku na částku*. Ukázalo se totiž, že dotaz na finance je pro respondenty poměrně citlivá záležitost, a z časového hlediska bývá rozdíl zodpovědět otázku: „Jaký byl v loňském roce Váš zisk či ztráta z podnikání?“, než uvést třeba rok, kdy se respondent přistěhoval do svého bytu.

#### **2.1.4 Škálové otázky**

V modulové části šetření se nejčastěji vyskytuje ještě *škálová otázka*. Tady má dotazovaný k dispozici navržené varianty, ze kterých ale nevybírám,

nýbrž je hodnotí výběrem ze stupnice od 0 do 10 dle svého uvážení. Přiřadí-li respondent hodnocení 0 znamená to, že chce vyjádřit postoj *zcela nespokojen, nemá rozhodně smysl* nebo *zcela nedůvěřuji*, naopak 10 bodů přidělí v opačném případě, je-li *zcela spokojen, má-li rozhodně smysl* nebo tehdy, kdy *zcela důvěřuje* třeba v právní systém v ČR. Často se s tímto typem proměnné setkáváme ve škole při hodnocení žáků formou známkování. Přidělujeme body/známky na škále od 1 do 5, jednička vyjadřuje nejlepší možné hodnocení, pětka to nejhorší. Obecně můžeme mít ale i jinou hodnotící stupnici, třeba od jedné do sta, od jedné do padesáti, atd. Tento typ proměnných je nejvhodnějším nástrojem pro měření postojů a názorů.

## 2.2 Analytická část

Dáme-li dohromady výše uvedené typy proměnných, dostaneme 12 skupin otázek, které jsme pro ilustraci zpracovali do následující tabulky. V šetření není žádná otázka, která by byla subjektivní a vypisovací, též škálová a objektivní, proto jsou tyto skupiny prázdné. Takto dostaneme právě 10 otázek, se kterými budeme dále pracovat. Označení otázek vzniklo pro interní potřebu a je dáno číslem otázky, pod jakým se vyskytuje v šetření, a zkratkou, která naznačuje, čeho se otázka týká.

Typ proměnné	Objektivní	Subjektivní
Výběr z variant	C17_POST_ZAM	CM4.POCITY
Alternativní otázka	B27_PRIJ_ZPRON	B18_SLUZ_VYR
Vypsání textového údaje	C15_HL_ZAM	-
Vypsání číselného údaje	A6_UK_VZD	C21_ODPRAC_HOD
Otázka na částku	B9_VYSE_HYPO	B13_ODH_CEN
Škálová otázka	-	CM8_DUVERA

Tabulka 1: Přehled vybraných otázek a používaných značení



### 2.2.1 Alternativní otázky

Analýzu začneme alternativními otázkami. Ty jsou zastoupeny proměnnými B27\_PRIJ\_ZPRON a B18\_SLUZ\_VYR. Obě otázky jsou součástí dotazníku B, v němž jsou obsaženy otázky, které se týkají bydlení, vlastnictví, vybavení bydlení, nákladů na bydlení atp. Alternativní otázky patří mezi nejjednodušší typ otázek, jaký nás může v šetření potkat. Respondent odpovídá výběrem z variant, které jsou však jen dvě. V našem případě volí mezi možnostmi *ano* a *ne*. Z hlediska časového by se mělo jednat o nejméně náročnou otázku.

Na následujícím obrázku 1 je znázorněna otázka B27\_PRIJ\_ZPRON.

B27 Měl(a) jste Vy nebo někdo jiný z Vaší domácnosti v minulém kalendářním roce příjem z pronájmu nemovitosti (bytu nebo jeho části, domu, nebytových prostor, chaty, pozemku), popř. movitých věcí (auta, strojů atd.)?		
1. ano	<input type="checkbox"/>	2. ne <input type="checkbox"/> → B30

Obrázek 1: Otázka B27.

[Zdroj: Dotazník za hospodařící domácnost, ČSÚ, Životní podmínky 2013.]

Zařadila jsem otázku mezi objektivní, protože se netýká pocitů nebo názorů respondenta, ale skutečnosti, která se buď dotazovaného týká nebo ne. Otázka se dotýká finanční situace respondenta. Konkrétně chce vědět, měl-li respondent nebo jiná osoba v domácnosti v loňském roce příjem z různých druhů pronájmů. Dotýká se tak tématu, o kterém se všeobecně moc nemluví a respondenti se s těmito informacemi neradi svěřují. Potíž může působit i fakt, že se otázka neptá jen na příjmy respondenta, ale i ostatních spolubydlících. Z tohoto hlediska bude čas zodpovězení této otázky v porovnání s jinou alternativní otázkou teoreticky větší. Jinou rozumíme třeba otázku A6\_12 z dotazníku A, která šetří, je-li dotazovaný muž či žena, proto je zde předpoklad, že s odpovědí respondent nebude váhat.

Stále se však jedná jen o odpověď formou *ano* či *ne*, otázka nejde nikterak do hloubky, respondent odpověď na ni zná, a tak by mu neměla činit větší problém. A tak fakt, že otázka vlastně šetří finanční situaci respondenta, by se v časové zátěži neměl nijak zvlášť projevit.

Nutno podotknout, že tato otázka je *filtrační*, což znamená, že z odpovědi

na další otázku takzvaně „vyfiltruje“ určité skupiny respondentů, kterých se následující otázka netýká, a odkáže je na otázku další. Odpoví-li respondent kladně, posouvá se na otázku další, kde už se po něm jistá specifikace příjmu chce. Pokud však zvolí odpověď ne, je odkázán až na otázku B30.

Otázka B18\_SLUZ\_VYR je subjektivní (viz obrázek 2).

B18 Může si Vaše domácnost dovolit uvedené služby nebo výrobky?		
	ano	ne
1. zaplatit ročně všem členům HD alespoň týdenní dovolenou mimo domov (včetně pobytu na chatě/chalupě, u přátel/příbuzných apod.)	<input type="checkbox"/>	<input type="checkbox"/>
2. jíst maso, drůbež nebo ryby každý druhý den (nebo jejich vegetariánské náhražky)	<input type="checkbox"/>	<input type="checkbox"/>
3. dostatečně vytápět byt	<input type="checkbox"/>	<input type="checkbox"/>

Obrázek 2: Otázka B18.

[Zdroj: Dotazník za hospodařící domácnost, ČSÚ, Životní podmínky 2013.]

Ve třech samostatných podotázkách se otázka ptá, zda-li si dotazovaná domácnost může dovolit uvedené výrobky a služby. V prvním bodě se zjišťuje možnost zaplatit jednou za rok celé hospodařící domácnosti týdenní pobyt v rámci dovolené mimo domov, v bodě druhém vyšetřujeme možnost koupit maso či ryby k jídlu každý druhý den a v posledním bodě možnost dostatečně vytápět byt.

Otázku jsem záměrně zařadila mezi subjektivní, protože se bude odvíjet od životní situace konkrétního respondenta. Záležet bude na jeho úsudku, postoji a hodnotách. Třeba pod týdenní dovolenou si zkrátka každý představíme asi jiný standard a pro každého je pojem *normální* a *přirozené* něco jiného.

Otázka B18 bude co obsahu zřejmě problematičtější než otázka předchozí, nicméně opět jde jen odpověď, která tvrzení buď vyvrátí nebo potvrdí, čímž není tak složitá, jak se na první pohled může zdát. Každá rodina ví, jestli ročně na nějakou dovolenou vyjede, ať už v létě nebo v zimě. Pokud ne, tak asi ví, z jakého důvodu, je-li tím důvodem tíživá finanční situace, nebo prostě jen fakt, že rodinná dovolená není předmětem jejich zájmu. Stejně tak u dalších dvou bodů.

### 2.2.2 Výběrové otázky

Jak již bylo řečeno, výběrové otázky se od těch alternativních odlišují tím, že při odpovědi na kteroukoliv z nich volíme z množiny nabízených variant. Už se nejedná o odpověď *ano* či *ne*. Máme v nabídce konkrétní odpovědi, čímž jsme při zodpovídání omezeni. Musíme věnovat svůj čas tomu, abychom si navržené možnosti alespoň pročetli, a měli tak jasnou představu o tom, z čeho budeme odpověď vybírat. Za tento typ proměnné uvažujeme otázky C17\_POST\_ZAM a CM4\_POLOCITY.

Otázka C17\_POST\_ZAM je typickým zástupcem skupiny. Je jednoduchá, objektivní, věcná a krátká a při odpovědi vybíráme jednu z pěti jednoduchých variant. Je součástí dotazníku za osobu, který obsahuje otázky z oblasti pracovní aktivity, současného i bývalého hlavního zaměstnání, nezaměstnanosti a všech druhů příjmů či zdraví. Otázka je uvedena na obrázku 3.

C17 Jaké postavení máte ve svém současném hlavním zaměstnání?	
1. zaměstnanec	<input type="checkbox"/>
2. společník, jednatel s.r.o.	<input type="checkbox"/>
3. osoba samostatně výdělečně činná se zaměstnanci	<input type="checkbox"/>
4. osoba samostatně výdělečně činná bez zaměstnanců	<input type="checkbox"/>
5. pomáhající rodinný příslušník	<input type="checkbox"/>

→ C20

Obrázek 3: Otázka C17.  
[Zdroj: Dotazník za osobu, ČSÚ, Životní podmínky 2013.]

Týká se současného hlavního zaměstnání a ptá se na postavení, jaké v něm respondent má. Z variant zaměstnanec, společník nebo jednatel s.r.o., OSVČ se zaměstnanci nebo bez nich, a pomáhající rodinný příslušník respondent vybere jednu, kterou odpoví. Tím, že otázka šetří klasické informace spojené s osobou respondenta, neměla by být problematická, a respondent by tak měl být schopen odpovědět, aniž by nad tím musel nějak zdlouhavě přemítat.

O něco komplikovanější už je otázka CM4\_POLOCITY viz obrázek 4.

CM4 Jak často (z hlediska doby) za poslední 4 týdny:						
	stále	často	občas	málokdy	nikdy	nedokážu posoudit
1. jste se cítil(a) velmi nervózní	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. jste pociťoval(a) tak špatnou náladu, že Vás nic nemohlo rozveselit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. jste pociťoval(a) klid a pohodu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. jste cítil(a) skepsi a smutek	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. jste byl(a) šťastný(-á)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Obrázek 4: Otázka CM4.  
[Zdroj: Modul 2013, ČSÚ.]

Otázka je subjektivní a velice komplexní. V jedné otázce je zahrnuta široká škála pocitů, které respondent hodnotí výběrem z pěti druhů příslovce, „stále“, „často“, „občas“, „málokdy“, „nikdy“. Jako únikovou cestu při odpovědi může využít varianty „nedokážu posoudit“. Subjektivita zde vychází z tématu, na něž je modul pro tento rok zaměřen. Pro rok 2013 byl tímto tématem *well-being*, tedy pocit životní pohody a spokojenosti v různých oblastech lidského života. Otázka se proto zaměřuje na respondentovy pocity a názory. Doslova otázka chce vědět, jak často respondent v uplynulých čtyřech týdnech pociťoval nervozitu, špatnou náladu až skepsi a smutek, nebo naopak klid, pohodu a štěstí. Je jasné, že se otázka dotýká vždy konkrétního respondenta a je tak zcela individuální. Respondenti jsou v případě takové otázky často zdrženliví a zdráhají se odpovědět. Neradi se svěřují s tím, jak se cítí, a to tím spíš, že tazatel je pro ně vlastně cizí osoba. To se jistě v časové zátěži může projevit.

### 2.2.3 Vypsání textového údaje

Typ otázky týkající se vypsání textu jsme v úvodu charakterizovali jako otevřenou otázku. Ty jsou v celém šetření dvě a obě jsou objektivní. Subjektivní otevřená otázka se v šetření nevyskytuje, takže tuhle kategorii otázek bude zastupovat jen jedna proměnná. Pro popis byla vybrána otázka C15\_HL\_ZAM. Na obrázku 5 je znění otázky znázorněno.

Že je otázka otevřená, vidíme na první pohled, protože dává prostor k vy-

C15 Jaké je Vaše současné hlavní zaměstnání (profese)? Popište prosím co nejpodrobněji práci, kterou vykonáváte.	
→ dvoumístný kód CZ-ISCO:	<input type="text"/> <input type="text"/>

Obrázek 5: Otázka C15.  
[Zdroj: Dotazník za osobu, ČSÚ, Životní podmínky 2013.]

psání textu. Ptá se na náplň hlavního zaměstnání. Tím, že otázka neuvádí varianty odpovědí, se může zdát složitější a může se předpokládat větší časová u této otázky. Nicméně k odpovědi nejsou třeba žádné sáhodlouhé výpovědi, naopak otázka vybízí k odpovědi jednou větou, případně slovním spojením. Například jako svou odpověď respondent uvede *vedoucí odboru makroekonomických analýz ČNB*. Otázka navíc nevyšetřuje žádné závažné či osobní téma, je objektivní a věcná, sjata s profesí, kterou respondent vykonává. To vše působí na snižování časové zátěže. Nutno podotknout, že respondent na otázku vůbec neodpovídá v případě, že je aktuálně nezaměstnaný nebo třeba důchodce.

#### 2.2.4 Vypsání číselného údaje

Skupinu proměnných vyžadujících k odpovědi vypsání číslice reprezentují proměnné A6\_UK\_VZD a C21\_ODPRAC\_HOD. První otázka A6\_UK\_VZD, jak už z označení plyne, je součástí dotazníku A za byt a zároveň je součástí otázky A6. V dotazníku A se celkově vyskytuje jen šest otázek, otázka A6 je v něm a v celém šetření nejrozsáhlejší. Obsahuje podotázky různého typu: alternativní, výběrové i otázky vyžadující vypsání číselného údaje. Otázka A6 je pro ilustraci celá znázorněna na obrázku 6, přičemž podotázky jsou zde barevně odlišeny podle typu, modře jsou vyznačeny alternativní podotázky, výběrové žlutě a vypisovací červeně.

Celá otázka A6 je objektivní a týká se věcných informací. Její úlohou je vy-

Pořadové číslo osoby	Identifikační číslo		Křestní jméno	Vztah k uživateli bytu, osobě v čele HD				Identifikační číslo (ze sl. 3)					
	HD	osoby						otce	matky	partnera			
			Slouží ke snadnější identifikaci osob při rozhovoru. Jako první se zapisuje uživatel/ka bytu, pak jeho/její partner/ka, děti a ostatní osoby.			BD	HD						
1	2	3	4			5	6	7	8	9			
1.	<input type="text"/>	<input type="text"/>				<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>			
Měsíc a rok narození		Pohlaví	Rodinný stav	Rok sňatku	Přítomnost/nepřítomnost v domácnosti (demografický pohyb, stěhování)						Kontrolní součet		
měsíc	rok (poslední dvojčíslí)	1 muž 2 žena	Zapište se dvojčíslí roku posledního sňatku*	1 panělová osoba 2 osoba mimo panel	Druh přítomnosti (2. až 4. vlna)	1 přítomný 2 částečně nepřítomný	Zapište se měsíc a dvojčíslí roku, kdy se osoba do domácnosti přistěhovala	Zapište se měsíc a dvojčíslí roku, kdy se osoba z domácnosti odstěhovala, popř. zemřela	Místo odstěhování	Součet číselných hodnot ve sloupcích 2, 3, 5 až 22			
10	11	12	13	14	15	16	17	18	19	20	21	22	23
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Pořadové číslo osoby	Nejvyšší dokončené vzdělání				Dvojčíslí roku dokončení nejvyššího vzdělání	Současné studium							
	24					25	26						
1.	<input type="text"/>				<input type="text"/>	<input type="text"/>							

Obrázek 6: Otázka A6.  
[Zdroj: Dotazník za byt, ČSÚ, Životní podmínky 2013.]

šetřit údaje za osobu, konkrétně pak, jaký vztah má šetřená osoba k uživateli bytu, jaký je měsíc a rok jejího narození, je-li muž či žena, jaký je rok případného sňatku, je-li šetřená osoba momentálně přítomna, měsíc a rok, kdy se osoba do bytu nastěhovala, stejně tak kdy se z bytu odstěhovala, případně kdy zemřela, místo odstěhování, nejvyšší dokončené vzdělání a rok jeho dosažení, a současné studium. Z toho je zřejmé, že otázka se vskutku týká základních informací k identifikaci respondenta a osob žijících s respondentem v jedné domácnosti, časová zátěž těchto podotázek by proto měla být minimální.

Předmětem našeho zájmu je pouze podotázka A6\_UK\_VZD. Ta je vypisovací a odpovědí je číslice. Týká se měsíce a dvojčíslí roku dosažení nejvyššího vzdělání. Otázka je objektivní, odpověď je reálná a respondent ji zná, a týká se skuteč-

nosti. Výběrové otázky jsou obecně možná lehčí v tom, že respondenta navržené alternativy směřují, jak má odpovídat, a v případě, že by si třeba na odpověď nemohl vzpomenout, mohou ho inspirovat, aby si vzpomněl. Zde varianty na výběr nejsou, problém s odpovědí by tak možná mohly mít starší generace. Kdy ukončily nejvyšší vzdělání by si hned nemusely vzpomenout, protože vzdělání ukončily podstatně dříve než mladší respondenti. Navíc je to datum, které si lidé tolik nepřipomínají a neoslavují jako třeba datum narození.

Stejným typem, nicméně subjektivní, je otázka C21\_ODPRAC\_HOD, která je uvedena na obrázku 7.

<b>C21</b> Kolik hodin odpracujete ve svém zaměstnání/podnikání nebo při práci pro rodinnou firmu? Uvedte prosím obvyklý počet hodin týdně. (U zaměstnanců se zahrne i neplacená práce přesčas.)	
hlavní zaměstnání	<input type="text"/> <input type="text"/>
další zaměstnání (přichází-li v úvahu)	<input type="text"/> <input type="text"/>

Obrázek 7: Otázka C21.  
[Zdroj: Dotazník za osobu, ČSÚ, Životní podmínky 2013.]

Otázka se ptá na obvyklý počet odpracovaných hodin týdně v zaměstnání nebo podnikání včetně přesčasů u zaměstnaneckého poměru. Přitom se dotazuje na hlavní zaměstnání a uvažuje i situaci, že má respondent mimo to i jiné zaměstnání.

U zaměstnanců, kteří mají většinou vytyčenou fixní pracovní dobu bez přesčasů, (řekněme třeba klasicky osmihodinovou pracovní dobu včetně půlhodinové neplacené přestávky), bude odpověď jednodušší a určitě bude přesnější. Stejně tak u zaměstnání, kde se zaznamenává příchod a odchod do práce a rovnou se zaměstnanci znázorní celkový odpracovaný čas. Problém ale mohou mít ti respondenti, kteří vykonávají nepravidelné nárazové přesčasy. Také podnikatelé, kteří mohou mít pracovní dobu zcela nepravidelnou, a příchody a od-

chody do práce si nikam nezaznamenávají, bude odpovídat pouhým odhadem odpracovaných hodin. Z těchto důvodů jsem otázku zařadila mezi subjektivní, protože závisí na úsudku respondenta a je pouze na něm a na jeho odhadu, co odpoví.

### 2.2.5 Otázky na částku

Otázky na částku jsou také vypisovací a ptají se na číslíci. Tady však působí problém fakt, že se tazatel ptá na důvěrné informace, jako jsou finance a peněžní závazky respondenta. Jsou to citlivější informace než v případě „obyčejné“ otázky na číslíci, proto je uvádíme ve zvláštní skupině. Z této kategorie byly vybrány otázky B9\_VYSE\_HYPO a B13\_ODH\_CEN.

B9 Kolik činila počáteční výše Vaší hypotéky/úvěru, tj. výše jistiny?	
částka v Kč	<input type="text"/>

Obrázek 8: Otázka B9.  
[Zdroj: Dotazník za hospodařící domácnost, ČSÚ, Životní podmínky 2013.]

Otázka B9\_VYSE\_HYPO, znázorněná na obrázku 8, patří mezi objektivní otázky. Ptá se na počáteční výši hypotéky nebo úvěru, který respondent splácí. V případě, že žádný úvěr na bydlení nemá nebo má, ale splácí až od roku 2013, je odkázán až na otázku B13, a na tuto otázku nebere zřetel. Jinak uvede částku ve výši jistiny, tedy počáteční zapůjčené sumy v korunách, neuvádí se zde výše úroků, o kterou respondent na konci zaplatí více. Výše úvěru je daná, sjednaná ve smlouvě a respondent do ní může kdykoliv nahlédnout. Odhadovat by proto neměl, odpověď by v tomhle případě měla být přesná. Už proto, že v dnešní době téměř každý financuje svoje bydlení nějakým úvěrem a v rodinném rozpočtu tvoří tak zásadní položku, že každý zřejmě i po pěti letech bude vědět, kolik si půjčil. Na časovou zátěž otázky tak bude nejspíš působit jen fakt, že se otázka ptá na citlivější informace. Bude proto záviset na konkrétním respondentovi, jak moc bude sdílný.



Otázka B13\_ODH\_CEN se ptá na odhad tržní ceny bytu či domu, který respondent obývá, a je uvedena na obrázku 9.

<b>B13</b> Následující otázka slouží k odhadu současných cen bydlení v ČR. Pokuste se prosím odhadnout tržní cenu bytu/domu, ve kterém bydlíte.	
cena v Kč	<input type="text"/>

Obrázek 9: Otázka B13.  
[Zdroj: Dotazník za hospodářící domácnost, ČSÚ, Životní podmínky 2013.]

Tržní cena hraje roli při koupi či prodeji nemovitosti a závisí na mnoha faktorech, rozloze, umístění nemovitosti a její dostupnosti. Tržní cena je vlastně prodejní hodnotou nemovitosti. Odhad této ceny zpravidla zpracovávají odborné firmy, které tuto službu nabízejí. Základem pro stanovení této ceny je totiž srovnávací metoda, která porovnává danou nemovitost s jinou jí podobnou. Navíc nemovitost, podle jejíž ceny se chceme řídit, by měla být úspěšně prodána, poněvadž to vypovídá o dobře určené tržní ceně.

Otázka je tedy svým obsahem zaměřena na současné ceny bydlení v ČR. K tomu, aby byl respondent vůbec schopen na tuto otázku kvalitně odpovědět, měl by se v této problematice alespoň trochu orientovat. Předpokládám ale, že řada respondentů o tom příliš povědomí nemá, proto bude potřeba, aby tazatel otázku vysvětlil.

V porovnání s otázkou předchozí se sice netýká osobních věcí respondenta, nicméně je velmi spjata s jeho úsudkem a názorem na tuto nelehkou problematiku. Odpověď na ni bude opět předpojatá a závislá na respondentovi. Otázka tak zastupuje v této skupině jednoznačně subjektivní proměnnou, navíc bych řekla, že je v šetření jednou z nejtěžších otázek vůbec.

### 2.2.6 Škálové otázky

Nakonec popíšeme vybranou škálovou proměnnou. Je subjektivní, objektivní škálové otázky se v šetření nevyskytují. Navíc stojí za zamyšlení, zda vůbec škálová otázka může být objektivní. Tím, že respondent navržené alternativy hod-

notí, je tím odpověď automaticky předpojatá. I kdybychom tak měli věcné téma otázky, vždy bude na respondentovi, jak varianty ohodnotí.

Škálové otázky jsou celkově v šetření obsaženy pouze v modulu a je jich jen pět. Pro rozbor jsme vybrali otázku CM8\_DUVERA, ty ostatní jsou této otázce hodně podobné. Otázka je znázorněna na obrázku 10.

CM8 Do jaké míry důvěřujete: Při odpovědi použijte, prosím, stupnici od 0 ( <i>vůbec nedůvěřuji</i> ) do 10 ( <i>zcela důvěřuji</i> ).		
	důvěra	nedokážu posoudit
1. státnímu systému ČR	<input type="checkbox"/>	<input type="checkbox"/>
2. právnímu systému ČR	<input type="checkbox"/>	<input type="checkbox"/>
3. Policii ČR	<input type="checkbox"/>	<input type="checkbox"/>

Obrázek 10: Škálová otázka.  
[Zdroj: Modul 2013, ČSÚ.]

Každá škálová otázka je uvozena stejným výrazem: „Do jaké míry..?“ Pak následuje celá škála dotazů týkající se spokojenosti se svým životem, bydlením, zaměstnáním, dojížděním, osobními vztahy atp. Otázka CM8\_DUVERA je věnována míře důvěry ve státní a právní systém v České republice a v Policii ČR. Dotazovaný hodnotí varianty vždy na stupnici 0 (toto bodové hodnocení přísluší jazykovým termům *vůbec nedůvěřuji*, *vůbec nemá smysl*, *zcela nespokojen*) až 10 (nahrazuje hodnocení pomocí termů *zcela důvěřuji*, *má rozhodně smysl*, *zcela spokojen*). Hodnotící jazykové termy jsou zde uvedeny tedy pouze pro představu vždy ke krajnímu ohodnocení body 0 a 10. Co si respondent představí mezi těmito krajními body, je na něm. Takže třeba hodnocení číslem 3 pak každý respondent může vnímat různě. Proto je hodnocení v tomto případě dosti subjektivní.

## 3 Teoretická rozdělení pravděpodobnosti

V úvodu jsme uvedli, že cílem diplomové práce je určit rozdělení pravděpodobnosti času, který je zapotřebí k zodpovězení jednotlivých typů otázek. V této sekci si představíme některá teoretická spojitá rozdělení pravděpodobnosti, která by mohla odpovídat skutečnému rozdělení pravděpodobnosti časové zátěže a již budeme v praktické části testovat. Jelikož podrobná analýza pravděpodobnostních rozdělení jako takových není hlavním cílem diplomové práce, omezíme se při jejich popisu zejména na hustotu a distribuční funkci a jejich grafickou interpretaci s důrazem na vliv změny parametrů na tvar hustoty a distribuční funkce. Uvažovat budeme exponenciální rozdělení, gama rozdělení, normální a log-normální rozdělení, a Weibullovo rozdělení. Kapitola vznikla za použití [1], [14], [18].

### 3.1 Normální rozdělení

Normální rozdělení, označováno také jako Gaussovo rozdělení, patří k základním stavebním kamenům teorie pravděpodobnosti a statistiky. Je jedním z nejznámějších a snad nejdůležitějších rozdělení pravděpodobnosti, a to zejména díky své roli, jakou hraje v centrální limitní větě. Řídí se jím mnoho náhodných veličin. Máme-li být ale přesní, je třeba říct, že řada náhodných veličin se tímto rozdělením neřídí přesně, ale za určitých podmínek normální rozdělení dobře aproximuje jiná, ať už spojitá či diskrétní rozdělení. Řada veličin se jím řídí alespoň přibližně, nejběžnějším takovým typem jsou náhodné chyby nebo veličiny, jejichž hodnoty lze pokládat za výsledek velkého množství nepatrných vzájemně nezávislých vlivů. Normalita je také důležitou podmínkou při statistické analýze k tomu, aby mohla být aplikována statistická inference (testování hypotéz, intervaly spolehlivosti) v populárních statistických metodách.

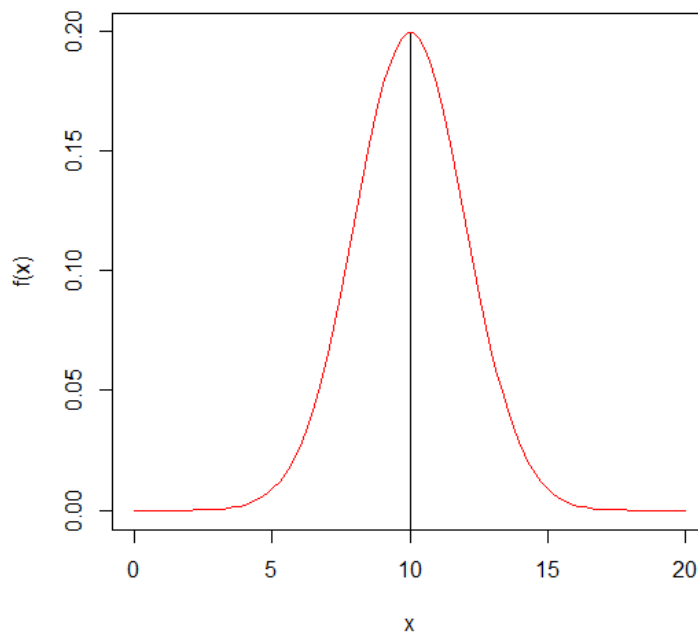
Normální rozdělení je označováno symbolem  $N(\mu, \sigma^2)$  a je definováno pro spojitou náhodnou veličinu  $X$ , jejímž oborem hodnot je množina reálných čísel.

Rozdělení je charakterizováno hustotou ve tvaru

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right],$$

kde  $\mu$  a  $\sigma$  jsou parametry rozdělení a platí, že  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$ .

Graf hustoty normálního rozdělení pro hodnoty parametrů  $\mu = 10$ ,  $\sigma = 2$  uvádíme na následujícím obrázku 11.



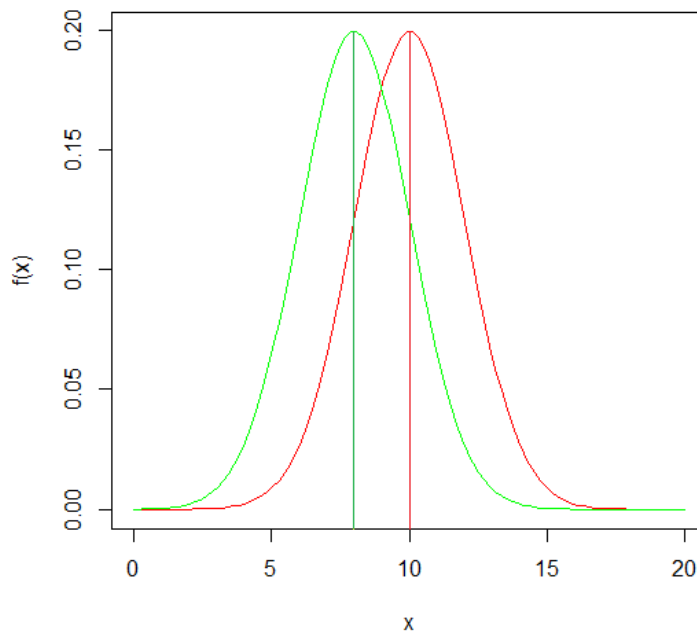
Obrázek 11: Funkce hustoty normálního rozdělení.

Z grafu je patrné, že normální rozdělení je unimodální rozdělení. Hustota rozdělení je zvonovitého tvaru, proto se normálnímu rozdělení často přezdívá „rozdělení ve tvaru zvonu“. Hustota je symetrická a veličiny s tímto rozdělením mají nulovou šikmost. Je koncentrována kolem střední hodnoty  $\mu$ , která je na obrázku znázorněna vertikální úsečkou. Je tedy rovna hodnotě deset. V důsledku symetričnosti jsou odchylky od střední hodnoty mající stejnou velikost, ale opačné znaménko, stejně pravděpodobné. V bodě střední hodnoty dosahuje funkce

hustoty svého maxima a posuneme-li se od tohoto bodu vlevo či vpravo, hustota rozdělení klesá.

Jak již bylo zmíněno, parametr  $\mu$  je střední hodnotou rozdělení,  $\sigma^2$  potom představuje rozptyl veličiny  $X$ . V intervalu mezi  $\mu - \sigma$  a  $\mu + \sigma$  je 68% hodnot, mezi  $\mu - 2\sigma$  a  $\mu + 2\sigma$  je 95% hodnot a konečně mezi  $\mu - 3\sigma$  a  $\mu + 3\sigma$  je 99,7% hodnot, takže skoro všechny.

Nyní se budeme věnovat vlivu parametrů na tvar hustoty. Nejprve zafixujeme hodnotu parametru  $\sigma$  na hodnotě 2 a budeme měnit hodnotu  $\mu$ . Už z grafu 11 je patrné, že se zvyšující se střední hodnotou se hustota rozdělení bude posunovat po  $x$ -ové ose doprava, jinak doleva. Názorně si to ukážeme na následujícím obrázku 12.

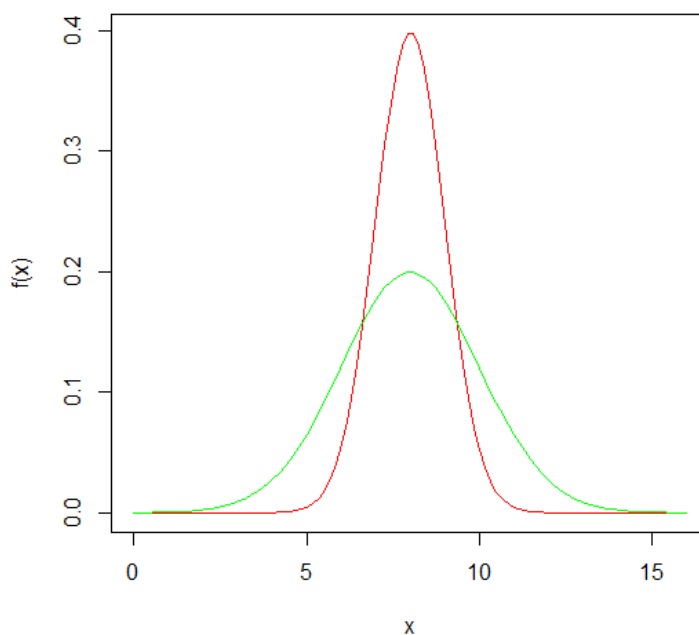


Obrázek 12: Vliv změny střední hodnoty na tvar hustoty.

Zelená křivka přísluší hustotě rozdělení pro  $\mu_1 = 8$ , červená pro  $\mu_2 = 10$ . Parametr rozptylu zůstává zachován na hodnotě  $\sigma^2 = 4$ . Vidíme, že opravdu s klesající střední hodnotou rozdělení se hustota posunuje na ose  $x$  vlevo, tvar

i velikost hustoty přitom zůstávají zachovány.

Nyní zafixujeme střední hodnotu na hodnotě  $\mu = 8$  a budeme sledovat vliv změny rozptylu na hustotu rozdělení viz obrázek 13. Uvažovat budeme hodnoty rozptylu  $\sigma_1^2 = 1$  a  $\sigma_2^2 = 4$ . Zelená křivka hustoty přísluší hodnotě rozptylu čtyři, červená jedné.



Obrázek 13: Vliv změny rozptylu na tvar hustoty.

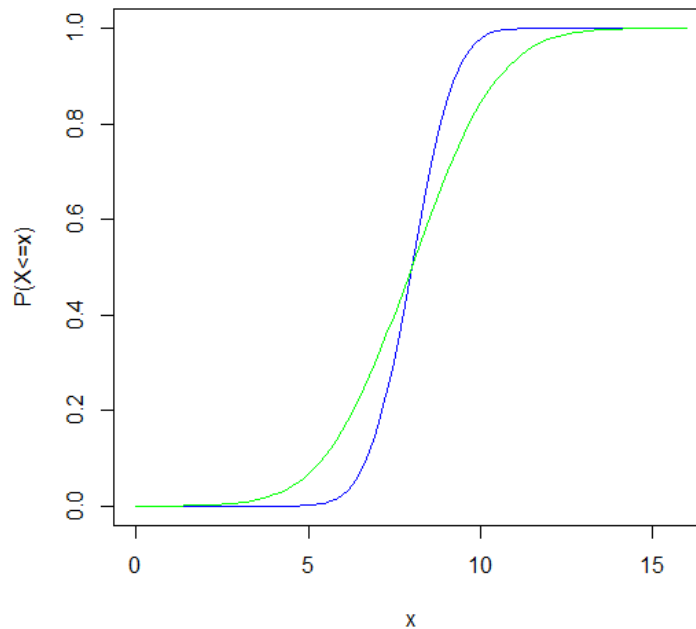
Na první pohled vidíme, že vlivem změny rozptylu dochází ke změně tvaru hustoty rozdělení. Střední hodnota je konstatní, křivka se tím pádem neposouvá po horizontální ose, nemění polohu. Rostoucí rozptyl, znázorněný zelenou křivkou, má za důsledek, že se hustota rozdělení zplošťuje, roste její rozpětí a stává plošší. Menší rozptyl zase způsobí, že je hustota rozdělení špičatější a má větší výšku.

Závěrem kapitoly o normálním rozdělení přidáme ještě pár slov o distribuční funkci. Z teoretického hlediska je totiž pro popis pravděpodobnostního chování náhodné proměnné nejvhodnější právě distribuční funkce, která, podobně jako hustota, podává úplnou informaci o rozdělení pravděpodobnosti náhodné veličiny.

Distribuční funkci normálního rozdělení  $F(x)$  nelze vyjádřit pomocí elementárních funkcí, proto ji vyjádříme jednoduše jako integrál hustoty

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] dt.$$

Pro ilustraci uvádíme graf distribuční funkce normálně rozdělené náhodné proměnné a pro úplnost uvažujeme dvě situace pro dva různé rozptyly,  $\sigma_1^2 = 4$  (příslušná této hodnotě rozptylu je hustota daná zelenou křivkou) a  $\sigma_2^2 = 1$  (znázorněno modrou křivkou), střední hodnota je společná  $\mu = 8$ .



Obrázek 14: Distribuční funkce normálního rozdělení.

S rostoucím rozptylem je distribuční funkce méně strmá a konce jsou „zaoblenější“. Střední hodnota je v grafu inflexním bodem, kde se obě křivky protínají a kde se funkce z konvexní mění na konkávní.

### 3.2 Logaritmicko-normální rozdělení

Na normální rozdělení navážeme logaritmicko-normálním rozdělením. Někdy se můžeme setkat s názvem lognormální rozdělení nebo log-normální rozdělení.

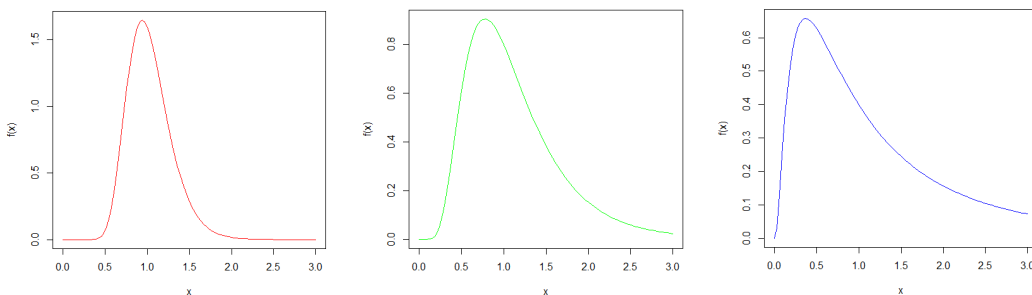
Již z názvu plyne, že mezi těmito dvěma rozděleními je úzká souvislost.

Řekneme, že náhodná veličina  $X$  má logaritnicko-normální rozdělení s parametry  $\mu$  a  $\sigma$ , kde  $\mu \in \mathbb{R}$ ,  $\sigma \in (0, \infty)$ , jestliže proměnná  $\ln(X)$  má normální rozdělení pravděpodobnosti se střední hodnotou  $\mu$  a směrodatnou odchylkou  $\sigma$ . Jinak řečeno, jestliže  $Y$  je náhodná veličina s normálním rozdělením, potom  $X = e^Y$  má lognormální rozdělení, a naopak, když  $X$  má lognormální rozdělení, pak  $Y = \ln(X)$ .

Lognormální rozdělení, označováno symbolem  $LN(\mu, \sigma)$ , je definováno hustotou ve tvaru

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{[\ln x - \mu]^2}{2\sigma^2}\right), \quad x > 0.$$

Graf hustoty lognormálního rozdělení si ukážeme na obrázku 15 viz níže.



Obrázek 15: Hustota lognormálního rozdělení pro různé hodnoty parametru  $\sigma$ .

Vidíme na něm tři grafy vedle sebe, které znázorňují hustotu pro různé parametry  $\sigma$  s fixní nulovou střední hodnotou  $\mu = 0$ . Hodnoty směrodatné odchylky uvažujeme různé. Zleva doprava, červená křivka je pro hodnotu parametru  $\sigma = 0.25$ , zelená pro  $\sigma = 0.5$  a modrá křivka přísluší parametru  $\sigma = 1$ .

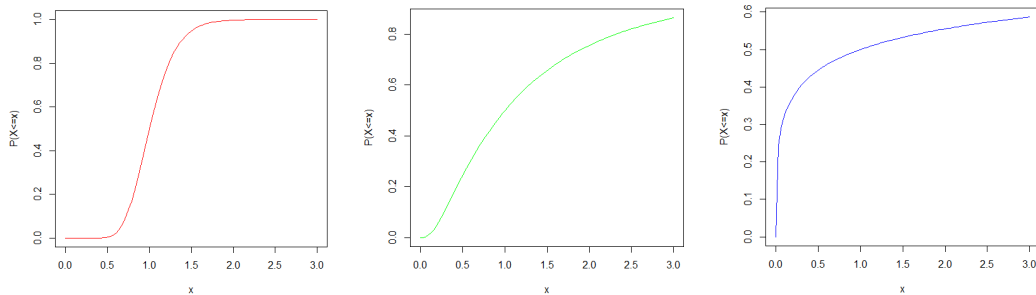
Je patrné, že hustota rozdělení již není symetrická, jako tomu bylo v případě normálního rozdělení, a je nyní zešikmená doprava. Čím je hodnota parametru  $\sigma$  menší, tím je hustota špičatější a o to déle její levý konec kopíruje osu  $x$  než začne růst. Vidíme, že červená křivka kopíruje osu  $x$  zhruba až do bodu  $x = 0.4$ , zatímco modrá křivka od nuly víceméně okamžitě roste. Jakmile vzroste do bodu svého maxima, následně o to prudčeji klesá, o co je hodnota  $\sigma$  nižší.



To je patrné na dalších dvou grafech. Pravé rameno hustoty klesá s rostoucím parametrem pomaleji, stejně tak klesá i výška hustoty. Hustota je v takovém případě plošší, aby zůstala splněna podmínka jednotkového obsahu plochy pod hustotou.

Dalo by se říct, že s rostoucím parametrem  $\sigma$  se hustota rozdělení jakoby posouvá na ose  $x$  doleva. Pro větší hodnoty parametru už bychom viděli jen klesající exponenciálu, která je pravým ramenem hustoty.

Distribuční funkci můžeme, podobně jako u normálního rozdělení, zapsat pouze jako integrál z hustoty. Pro ilustraci uvedeme obrázek 16, v němž znázorníme opět tři grafy distribuční funkce, jako tomu bylo u hustoty rozdělení. Budeme zkoumat vliv změny  $\sigma$  při konstantním  $\mu = 0$  na tvar distribuční funkce. Uvažujeme větší hodnoty  $\sigma$  než u hustoty, konkrétně zleva doprava  $\sigma_1 = 0.25$ ,  $\sigma_2 = 1$  a  $\sigma_3 = 5$ . Interpretace je podobná. S rostoucím  $\sigma$  posouváme křivku doleva, od počátku roste čím dál prudčeji a výška se snižuje.



Obrázek 16: Distribuční funkce lognormálního rozdělení pro různé hodnoty parametru  $\sigma$ .

Ještě než přejdeme k popisu exponenciálního rozdělení, uvedeme střední hodnotu a rozptyl lognormálního rozdělení v analytickém tvaru. Zde je vyjádření podstatně složitější než u normálního, kde platilo, že  $E(X) = \mu$  a  $\text{var}(X) = \sigma^2$ . Příslušná střední hodnota a rozptyl jsou v následujícím tvaru

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}, \quad \text{var}(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}.$$

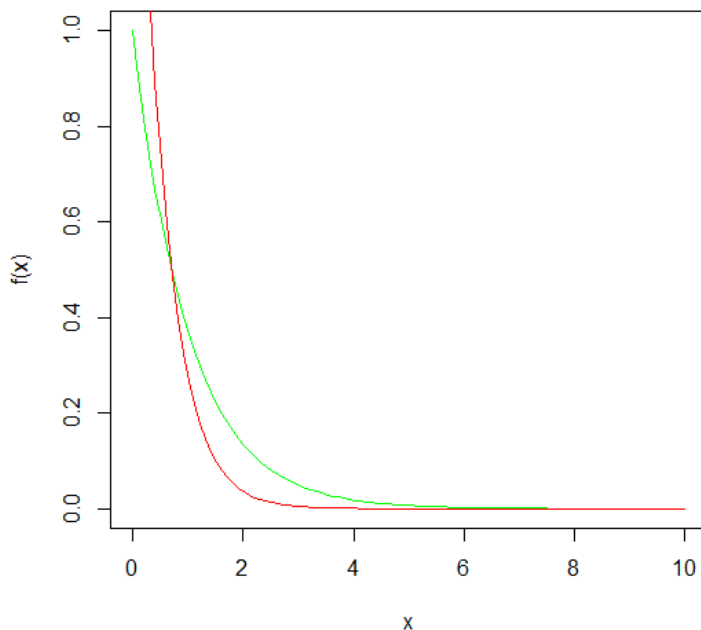
### 3.3 Exponenciální rozdělení

Exponenciální rozdělení se v praxi nejčastěji používá k modelování doby čekání na výskyt nějakého jevu. Řekněme třeba, že chceme dostat odpověď na otázku, kdy do našeho obchodu vstoupí zákazník, nebo jak dlouho vydrží nějaké zařízení bez poruchy. Obě doby, ať už do příchodu zákazníka nebo do poruchy zařízení, považujeme za náhodné proměnné a jejich rozdělení pravděpodobnosti lze často aproximovat právě exponenciálním rozdělením.

Exponenciální rozdělení zapisujeme symbolem  $E(\lambda)$  nebo  $Exp(\lambda)$  a je určeno jediným parametrem  $\lambda > 0$ . Hustota exponenciálního rozdělení pro náhodnou veličinu  $X$ , která je spojitá a nabývá hodnot z intervalu  $(0, \infty)$ , je ve tvaru

$$f(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right).$$

Graf hustoty exponenciálního rozdělení uvádíme na obrázku viz [17](#).



Obrázek 17: Funkce hustoty exponenciálního rozdělení.

Uvažujeme přitom dvě situace pro dvě různé hodnoty parametru  $\lambda$ ,  $\lambda_1 = 1$  a  $\lambda_2 = 2$ . První hodnota přísluší zelené křivce, druhá červené. Obě křivky mají tvar exponenciály. Výšku hustoty udává právě parametr  $\lambda$ , jehož hodnotu vynásíme na osu  $y$ . Od tohoto bodu funkce hustoty exponenciálně klesá a je o to více prohnutá, čím je hodnota parametru vyšší. Lze názorně vidět i na obrázku 17, že červená hustota je prohnutější, a klesá tak strměji.

Pro shrnutí vlastností rozdělení náhodné veličiny lze kromě hustoty a distribuční funkce použít také základní charakteristiky jako je střední hodnota a rozptyl. Tyto se v případě exponenciálního rozdělení vyjádří jednoduše, střední hodnota je inverzí parametru  $\lambda$  a rozptyl je její druhou mocninou, tedy

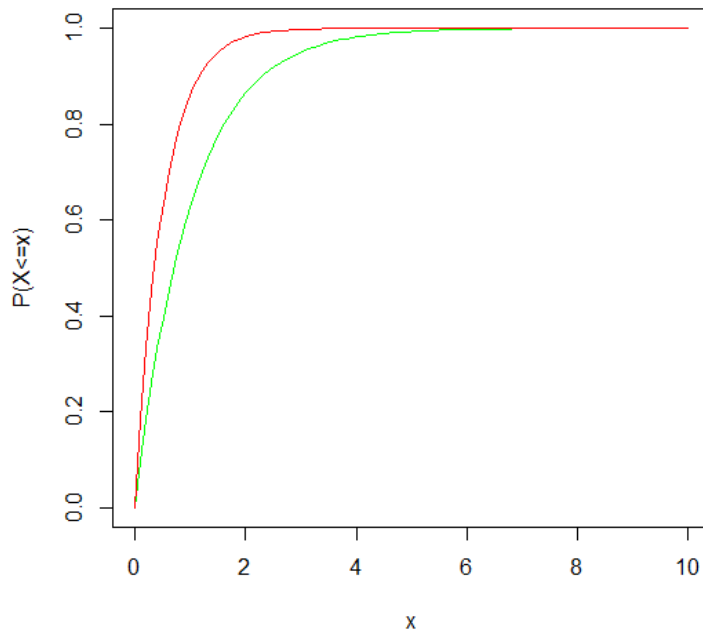
$$E(X) = \frac{1}{\lambda}, \quad \text{var}(X) = \frac{1}{\lambda^2}.$$

Pro získání distribuční funkce vyjdeme opět z integrace hustoty, tentokrát jsme schopni získat její explicitní tvar,

$$F(x) = \int_0^x \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) dt = 1 - \exp\left(-\frac{x}{\lambda}\right).$$

Na obrázku 18 vidíme graf distribuční funkce exponenciálního rozdělení. Podobně jako u hustoty exponenciálního rozdělení prezentujeme dvě varianty pro hodnoty parametru  $\lambda_1$ ,  $\lambda_2$ , stejné jako v případě hustoty. Zachováno zůstává i barevné rozlišení.

Svým tvarem připomíná distribuční funkce logaritmickou funkci. Je rostoucí, shora ohraničena přímkou  $y = 1$ . Přímkou se funkce nikdy nedotkne, proto platí, roste-li si  $x$  nad všechny meze, funkce se limitně blíží jedné. Opět platí, že čím vyšší hodnota parametru, tím je funkce prohnutější a tím déle kopíruje osu  $y$  a přímkou  $y = 1$ .



Obrázek 18: Distribuční funkce exponenciálního rozdělení.

### 3.4 Gama rozdělení

Gama rozdělení je v teorii pravděpodobnosti a matematické statistice spojitým rozdělením pravděpodobnosti se dvěma parametry. Jeho speciálními případy jsou  $\chi^2$ -rozdělení a právě exponenciální rozdělení. Gama rozdělení se podobně jako exponenciální rozdělení používá k modelování doby čekání na výskyt nějakého jevu, nebo v pojistné matematice k modelování výše pojistných plnění. Protože je ale gama rozdělení závislé na dvou parametrech, je pro svůj účel vhodnější a flexibilnější.

K definici rozdělení si opět dopomůžeme hustotou. Hustotu gama rozdělení lze vyjádřit pro spojitou náhodou proměnnou  $X$ , jejíž obor hodnot je  $(0, \infty)$  ve tvaru

$$f(x) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1}, \quad x > 0, \quad a > 0, \quad p > 0,$$

kde  $\Gamma()$  je gama funkce. Značíme  $X \sim \text{Ga}(a, p)$  nebo  $X \sim \Gamma(a, p)$ .

Dalo by se říct, že gama funkce je zobecněním funkce faktoriál. Podobně jako pro  $n!$  platí

$$n! = n \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1$$

respektive

$$n! = n \cdot (n - 1)!,$$

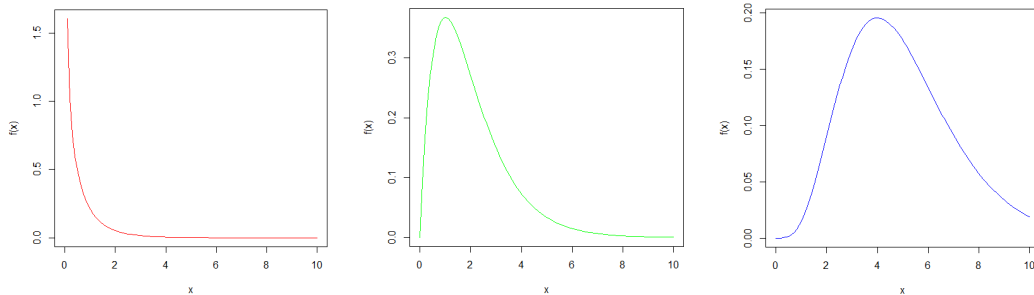
kde  $n \in \mathbb{N}$ , dá se i gama funkce  $\Gamma(x)$ , pro  $x > 0$ , zapsat ve tvaru, který vyhovuje podobné struktuře

$$\Gamma(x) = \Gamma(x - 1) \cdot (x - 1).$$

Formálně se potom gama funkce definuje jako

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

Na následujícím obrázku 19 uvádíme graf hustoty gama rozdělení. Jako u lognormálního rozdělení uvádíme v obrázku tři grafy hustoty pro různé hodnoty parametru  $a$ . Zleva doprava křivky přísluší hodnotám parametrů  $a = 0.5$ ,  $a = 2$ ,  $a = 5$ . Hodnotu parametru  $p$  nyní zafixujeme na hodnotě jedna.

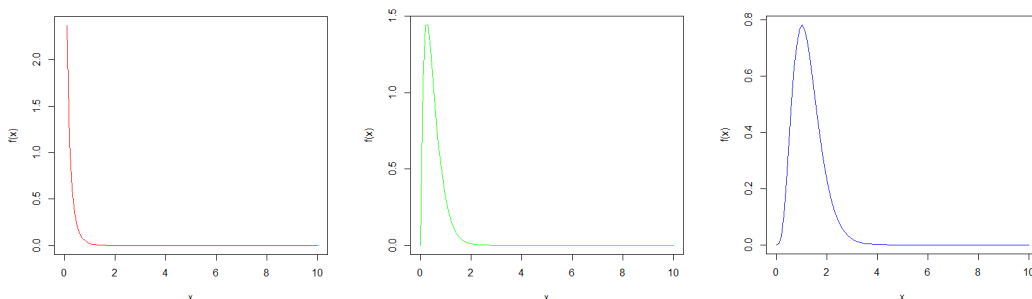


Obrázek 19: Hustota gama rozdělení v závislosti na měnícím se  $a$  pro  $p = 1$ .

Zde se zdá, jako by se hustota posouvala s rostoucí hodnotou parametru doprava. Zatímco pro  $a = 0.5$ , má hustota jen jedno rameno, je exponenciální a klesající, zelená i modrá křivka pro větší hodnoty  $a$  už mají dvě ramena a hustoty do určitého bodu rostou a pak klesají. S rostoucí hodnotou parametru

se hustota rozdělení stává více symetrická, modrá hustota svým tvarem nejvíce připomíná hustotu normálního rozdělení.

Zvýšíme-li hodnotu parametru  $p$  na čtyři, opět pro hodnotu parametru  $a$  rovno 0.5, 2 a 5, můžeme posoudit vliv této změny na hustoty na obrázku 20.



Obrázek 20: Hustota gama rozdělení v závislosti na měnícím se  $a$  pro  $p = 4$ .

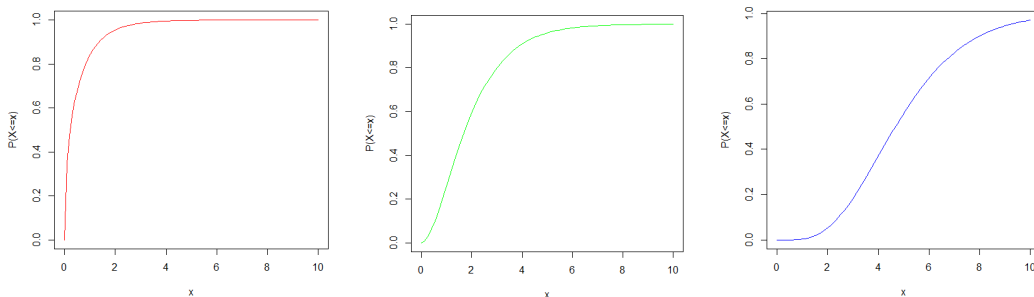
V první situaci je funkce více prohnutá, stáhla se k počátku soustavy souřadnic. V dalších dvou situacích se zmenšilo rozpětí hustoty, hustota je špičatější a též došlo k posunu celé hustoty doleva. Podobná bude situace u distribuční funkce.

Distribuční funkci gama rozdělení prezentujeme ve tvaru integrálu

$$F(x) = \int_0^x \frac{a^p}{\Gamma(p)} \exp(-at) t^{p-1} dt.$$

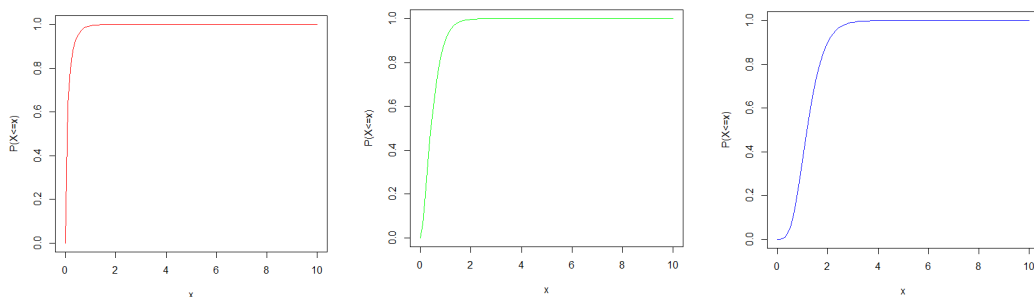
Na následujících obrázcích 21 a 22 vidíme její grafické znázornění. Opět uvažujeme stejné hodnoty parametru  $a$  jako u hustoty, parametr  $p$  jsme nyní zafixovali na hodnotě jedna.

Pro hodnotu  $a = 0.5$  je hustota logaritmickou funkcí a připomíná distribuční funkci exponenciálního rozdělení. Je konkávní, prohnutá k levému hornímu rohu obrázku. Zelená křivka už je méně prohnutá, zpočátku je lehce konvexní, poté následně přechází do konkávního tvaru. Pro  $a = 5$  se mírně zakulacuje a zdá se být konkávní. Pro  $a = 5$  je inflexní bod přesně ve středu definičního oboru distribuční funkce, funkce je proto do půlky definičního oboru konvexní, od půlky zase konkávní.



Obrázek 21: Distribuční funkce gama rozdělení v závislosti na měnícím se  $a$  pro  $p = 1$ .

Nakonec ještě uvažujme variantu, kdy  $p$  se zvedlo na hodnotu čtyři (viz obrázek 22). Hodnoty  $a$  se nezměnily.



Obrázek 22: Distribuční funkce gama rozdělení v závislosti na měnícím se  $a$  pro  $p = 4$ .

Odtud vidíme, že vlivem rostoucího parametru  $p$  distribuční funkce roste stále strměji.

### 3.5 Weibullovo rozdělení

Jako poslední ze spojitých pravděpodobnostních rozdělení uvedeme Weibullovo rozdělení. Pro svou všestrannost je hojně používáno např. v analýze přežití, v předpovědních modelech a také v teorii spolehlivosti. Jeho speciálním případem je také exponenciální rozdělení.

Weibullovo rozdělení pravděpodobnosti je označováno symbolem  $W(c, p)$

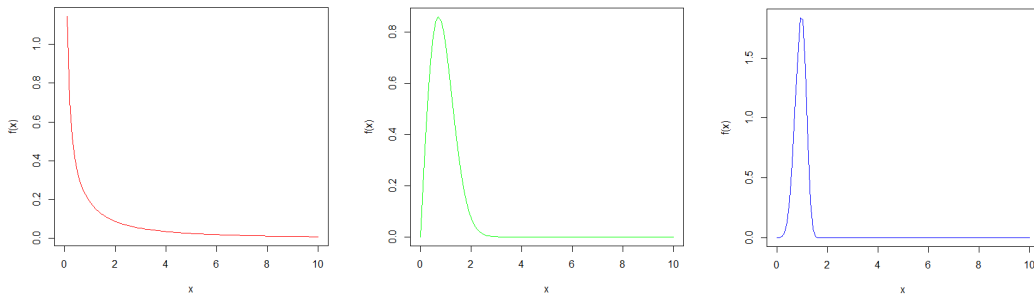
s hustotou

$$f(x) = cpx^{p-1} \exp(-cx^p), \quad x > 0,$$

kde parametry  $c$  a  $p$  jsou kladné.

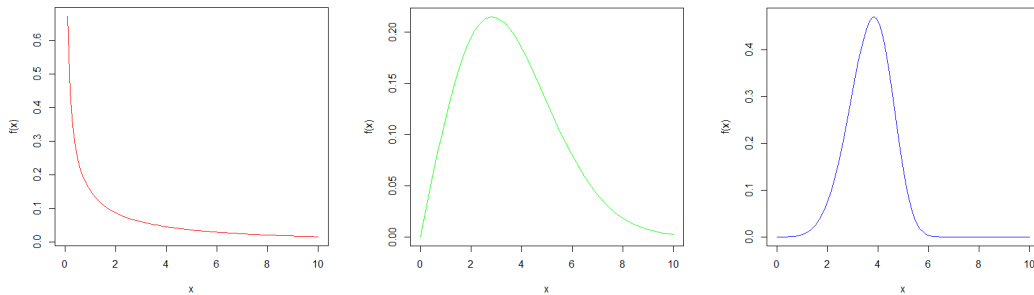
Grafickou studii hustoty a distribuční funkce Weibullova rozdělení provedeme stejně jako u předchozího gama rozdělení. Nejprve budeme uvažovat fixní hodnotu parametru  $p = 1$  a budeme měnit hodnotu  $a$ . Uvažovat budeme  $a = 0.5, 2, 5$ . Poté hodnotu  $p$  zvýšíme na 4 a opět porovnáme pro již uvedené volby parametru  $a$ .

Když se podíváme na obrázek 25, je zřejmé, že jeho interpretace bude analogická, jako v případě hustoty gama rozdělení pro  $p = 4$ , (viz obrázek 20).



Obrázek 23: Hustota Weibullova rozdělení v závislosti na měnícím se  $c$  pro  $p = 1$ .

Zvedneme-li  $p$  na hodnotu čtyři, grafická interpretace hustot je naopak dost podobná hustotám gama rozdělení pro  $p = 1$ , (viz obrázek 19).



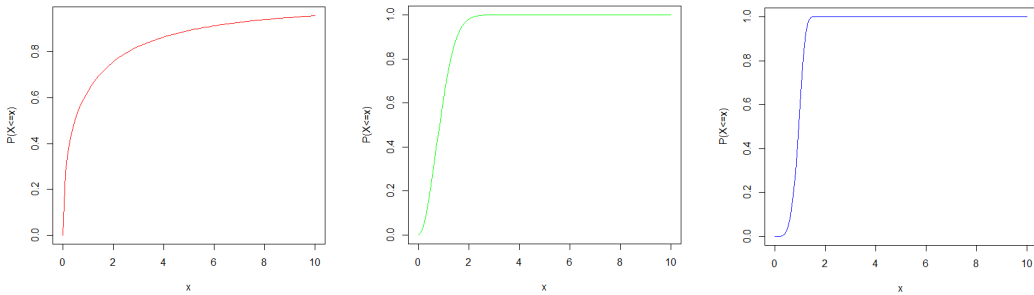
Obrázek 24: Hustota Weibullova rozdělení v závislosti na měnícím se  $c$  pro  $p = 4$ .



Distribuční funkce Weibullova rozdělení  $F(x)$ , kde

$$F(x) = \int_0^t cpx^{p-1} \exp(-cx^p) dt$$

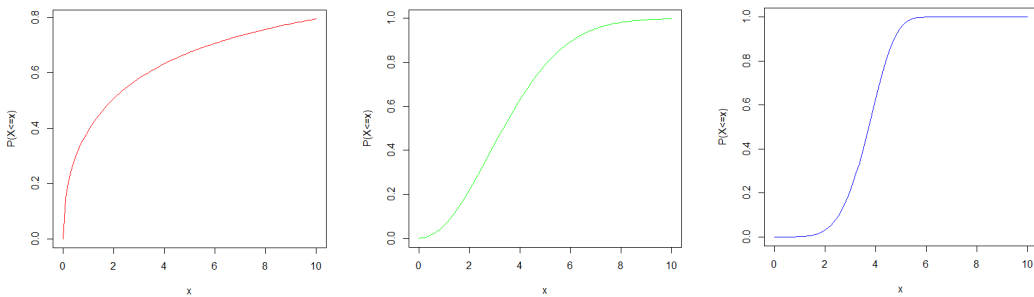
je znázorněna na další šestici grafů. Opět uvažujeme nejprve  $p = 1$ .



Obrázek 25: Distribuční funkce Weibullova rozdělení v závislosti na měnícím se  $c$  pro  $p = 1$

Distribuční funkce má tvar podobný funkci logaritmu. První červená funkce, pro nejmenší hodnotu parametru  $a = 0.5$ , je vizuálně čistě konkávní. Další distribuční funkce už jsou ostřeji zaoblené a jakoby se „zaseknou“ v bodě maxima, od něž je již funkce prakticky konstatní.

V druhém případě můžeme sledovat podobný vývoj, jen v méně extrémní podobě.



Obrázek 26: Distribuční funkce Weibullova rozdělení v závislosti na měnícím se  $c$  pro  $p = 4$ .

## 4 Testování shody rozdělení

Následující kapitola bude věnována procesu, o kterém se v českém překladu hovoří jako o *fitování rozdělení*. Jedná se o proces hledání vhodného teoretického pravděpodobnostního rozdělení na námi zkoumaná data. Jinak řečeno vybíráme takové statistické rozdělení, které nejlépe vyhovuje našemu datovému souboru a nejlépe jej popisuje.

V práci se v souvislosti s touto problematikou nejprve budeme zabývat metodou maximální věrohodnosti, což je jedna z parametrických metod, (jako další můžeme zmínit známou metodu momentů), a následně přejdeme k testu dobré shody při neznámých parametrech, který slouží k testování konkrétních rozdělení. Vzhledem k tomu, že pro test dobré shody je základní stavební jednotkou multinomické rozdělení, protože v principu test používá právě aproximaci multinomického rozdělení rozdělením  $\chi^2$ , charakterizujeme předtím multinomické rozdělení v samostatné kapitole. Celý oddíl byl zpracována primárně za použití zdrojů [1], [11], [16].

### 4.1 Metoda maximální věrohodnosti

Metoda maximální věrohodnosti je univerzální metodou odhadu neznámých parametrů ze statistického modelu. Je založena na tzv. věrohodnostní funkci a jejím výstupem jsou maximálně věrohodné odhady, tzv. MLE odhady (zkratka MLE pochází z anglického *Maximum Likelihood Estimation*).

Metoda spočívá v tom, že námi naměřené hodnoty tvoří vektor  $\mathbf{X}$  nezávislých stejně rozdělených náhodných veličin  $X_1, X_2, \dots, X_n$ . Jedná se tedy o náhodný výběr. Jde nám o to stanovit odhad  $\hat{\theta}$  hodnoty neznámého parametru  $\theta$ , skalárního či vektorového, v rozdělení, které je charakterizováno známou třídou distribučních funkcí  $\{F(x, \theta), \theta \in \Theta\}$ .

Pro odhad  $\theta$  je zapotřebí sestavit *věrohodnostní funkci*  $L(\theta|\cdot)$ , která v případě spojitého rozdělení odpovídá sdružené hustotě  $f(\mathbf{x}, \theta)$  náhodného vektoru  $\mathbf{X}$  pro pevné  $\mathbf{x}$ . Vzhledem k tomu, že máme předpoklad nezávislých stejně rozdělených náhodných veličin, sdruženou hustotu lze vyjádřit jako součin hustot

veličin  $X_1, \dots, X_n$ ,

$$L(\theta|x_1, \dots, x_n) = f(x_1, \dots, x_n|\theta)$$

$$f(x_1, \dots, x_n|\theta) = f(x_1|\theta)f(x_2|\theta) \dots f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta).$$

Jestliže existuje hodnota parametru  $\hat{\theta}$ , která maximalizuje přes  $\theta$  věrohodnostní funkci tak, že platí

$$L(\hat{\theta}|x_1, \dots, x_n) = \arg \max_{\theta \in \Theta} L(\theta|x_1, \dots, x_n),$$

pak parametr  $\hat{\theta}$  nazveme *maximálně věrohodným odhadem* (dále jen ML odhadem).

Pro názornost můžeme uvažovat situaci, kdy máme náhodný výběr z normálního rozdělení. Předpokládáme tedy normálně rozdělený náhodný vektor  $\mathbf{X} = (X_1, \dots, X_n) \sim N(\mu, \sigma^2)$ . ML odhady jsou řešením rovnice

$$L(\mu, \sigma^2) \rightarrow \max_{\mu, \sigma^2}.$$

Častěji než s věrohodnostní funkcí samotnou pracujeme s jejím přirozeným logaritmem, pro další výpočty je to pohodlnější, poněvadž lze následně nahradit součin součtem. Úloha je potom ve tvaru

$$\ln L(\mu, \sigma^2) \rightarrow \max_{\mu, \sigma^2}.$$

Věrohodnostní funkce je v našem případě sdružená hustota normálního rozdělení, tedy za předpokladu náhodného výběru, součin jednorozměrných hustot  $f(x)$ ,

$$L(\mu, \sigma^2) = \prod_{i=1}^n f(x_i),$$

kde

$$f(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x_i - \mu)^2}{2\sigma^2} \right].$$

Použitím vztahu

$$\ln \left[ \prod_{i=1}^n f(x_i) \right] = \sum_{i=1}^n \ln f(x_i)$$

dostáváme vyjádření ve tvaru

$$L(\mu, \sigma^2) = n \ln \left( \frac{1}{\sigma \sqrt{2\pi}} \right) - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \quad (1)$$

nebo též

$$L(\mu, \sigma^2) = -n \ln(\sigma) - n \ln(\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (2)$$

Nyní je třeba vyřešit úlohu maximalizace věrohodnostní funkce, což znamená, že funkci zderivujeme podle jednoho z parametrů a položíme rovno nule.

1. Prvně provedeme derivaci podle parametru  $\mu$

$$\frac{\partial L}{\partial \mu} \stackrel{!}{=} 0.$$

Použijeme k tomu věrohodnost ve tvaru (2). První dva výrazy z věrohodnostní rovnice (1) vypadnou, vzhledem k parametru  $\mu$  se jedná o konstanty. Derivace vypadá takto

$$-\frac{1}{2\sigma^2}(-2) \sum_i (x_i - \mu) = 0 \iff \sum_i (x_i - \mu) = 0.$$

Jdeme-li se sumou dovnitř závorky, dostáváme vztah, odkud už jednoduše vyjádříme odhad  $\mu$ ,

$$\begin{aligned} \sum_i x_i &= n\mu, \\ \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}. \end{aligned} \quad (3)$$

Maximálně věrohodným odhadem parametru střední hodnoty normálního rozdělení je tedy *výběrový průměr*, němž víme, že je nestranný, dokonce nejlepší nestranný, a konzistentní.

2. Nyní se pokusíme o vyjádření MLE odhadu parametru  $\sigma^2$ .

Vyjdeme zase z věrohodnostní funkce (2), kterou zderivujeme, tentokrát podle parametru  $\sigma^2$ , a položíme rovno nule. Dostáváme následující

$$\begin{aligned}\frac{\partial L}{\partial \sigma^2} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_i (x_i - \mu)^2 \stackrel{!}{=} 0 \\ \sum_i (x_i - \mu)^2 &= n\sigma^2 \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} S^2,\end{aligned}\quad (4)$$

kde  $S^2$  je výběrový rozptyl.

Maximálně věrohodným odhadem pro parametr  $\sigma^2$  je tedy statistika, která má menší rozptyl než  $S^2$ , ale zato je pouze asymptoticky nestranná.

Protože je většina vlastností ML odhadu asymptotická, (za poměrně obecných podmínek mají dokonce ML odhady asymptoticky normální rozdělení), je dobré se metodě maximální věrohodnosti vyhnout za předpokladu, že je počet pozorování  $n$  malý. Problematická může být metoda i v tom, že pro své účely využívá přesného pravděpodobnostního modelu. S tím souvisí fakt, že věrohodnostní funkce mohou být různě komplikované, vyřešit úlohu nalezení jejího maxima pak může vyžadovat použití speciálních iterativních metod algoritmů.

Nyní už přejdeme k testům dobré shody, ve kterých ML odhady parametrů rozdělení používáme. Začneme přitom definicí multinomického rozdělení.

## 4.2 Multinomické rozdělení

Opakujeme-li  $n$ -krát nějaký experiment, přičemž výsledkem jednotlivého pokusu může být buďto úspěch nebo neúspěch, potom počet úspěchů v daném experimentu popisuje náhodná veličina, která má binomické rozdělení. Jestliže opakujeme  $n$ -krát pokus, který má  $k$  různých výstupů, kde  $k$  je libovolné přirozené číslo, pak náhodný vektor  $\mathbf{X} = (X_1, \dots, X_k)$ , kde  $X_i$  je náhodná veličina, která značí počet výsledků  $i$ -tého pokusu, má multinomické rozdělení.

Ve většině zdrojů (např. [1]) je multinomické rozdělení interpretováno na konkrétním příkladu s osudím a kuličkami. I my se v práci budeme držet zavedené koncepce. Na příkladu si také demonstrujeme fakt, že multinomické rozdělení je zobecnění toho binomického.

Máme urnu s kuličkami  $k$  barev. Pravděpodobnost vytažení kuličky  $i$ -té barvy je  $p_i$  pro  $i = 1, \dots, k$ . Současně platí, že  $p_i \in (0, 1)$  a  $\sum_{i=1}^k p_i = 1$ . Nezávisle na sobě vybereme  $n$ -krát jednu kuličku a zase ji do urny vrátíme. Nechť veličina  $X_i$  značí počet kuliček  $i$ -té barvy, které byly takto vybrány,  $\mathbf{X} = (X_1, \dots, X_k)$  je vektor veličin  $X_i$ ,  $\mathbf{p} = (p_1, \dots, p_k)$  je vektor pravděpodobností  $p_i$ ,  $i = 1, \dots, k$ . Sdružené rozdělení náhodného vektoru  $\mathbf{X}$  se nazývá *multinomické* s parametry  $n; p_1, \dots, p_k$ . Značíme  $\mathbf{M}(n; p_1, \dots, p_k)$  nebo  $\mathbf{M}(n; \mathbf{p})$ .

Z výše popsaného modelu následně plyne

$$P(X_1 = x_1 \dots X_k = x_k) = \begin{cases} \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} & \text{pro } x_i = 0, \dots, n, \\ & \text{kde } i = 1, \dots, k, \\ & x_1 + \dots + x_k = n \\ 0 & \text{jinak} \end{cases}$$

Z příkladu tedy plyne, že náhodné veličiny  $X_i$ ,  $i = 1, \dots, k$ , které tvoří složky multinomicky rozděleného náhodného vektoru  $\mathbf{X}$ , mají binomické rozdělení s parametry  $n$  a  $p_i$ . Náhodná veličina  $X_i$  například představuje počet bílých kuliček tažených v  $n$  pokusech, tzn.  $X_i \sim Bi(n, p_i)$ .

Nechť  $\mathbf{X} \sim \mathbf{M}(n, \mathbf{p})$ , čili máme náhodný vektor s multinomickým rozdělením, pak střední hodnoty a rozptyly složek  $\mathbf{X}$  jsou charakterizovány pomocí jednorozměrné binomicky rozdělené náhodné veličiny  $X_i$ , a tedy

$$EX_i = np_i, \quad \text{var} X_i = np_i(1 - p_i), \quad 1 \leq i \leq k.$$

Jelikož se jedná o náhodný vektor musíme uvažovat ještě kovarianci mezi jednotlivými náhodnými veličinami  $X_i$  a  $X_j$ ,  $i, j = 1, \dots, k$ , tedy

$$\text{cov}(X_i, X_j) = -np_i np_j, \quad i \neq j.$$

Důkaz lze najít v [1], str. 270.

Multinomické rozdělení je pro své výhodné vlastnosti jedním z předních diskrétních rozdělení pravděpodobnosti náhodného vektoru. Svým významem by se dalo přirovnat k mnohorozměrnému normálnímu rozdělení. Tomu se podobá zejména v tom, že podmíněná i marginální rozdělení jsou opět multinomická. Stěžejní na multinomickém rozdělení je jeho souvislost s  $\chi^2$ -rozdělením. Vhodnou statistikou je totiž možné aproximovat jeho diskrétní rozdělení rozdělením spojitým. Vlastnost nám shrne následující věta. Text jsme záměrně naformulovali do věty, abychom se na ni mohli v průběhu textu dále odkazovat.

**Věta 4.1.** *Mějme náhodný vektor  $\mathbf{X}$  s multinomickým rozdělením s parametry  $n$  a  $p_1, \dots, p_k$ . Potom náhodná veličina*

$$Z = \sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i} \quad (5)$$

*má při  $n \rightarrow \infty$  asymptoticky rozdělení  $\chi_{k-1}^2$ .*

Statistika (5) má své využití právě v testu dobré shody při neznámých parametrech, se kterým se záhy seznámíme. Předtím je třeba ještě zmínit, že tato aproximace funguje pouze za předpokladu dostatečně velkého rozsahu výběru  $n$ , potažmo při splnění podmínky  $np_i \geq 5$ . Splnění této podmínky zřejmě při aplikaci testu pro rozsáhlé soubory z ČSÚ nebude problém.

### 4.3 Testy dobré shody při neznámých parametrech

Tato kapitola bude věnována testu dobré shody. Podstatou testu je, že se snaží zjistit, zda má náhodná veličina předem dané rozdělení pravděpodobnosti s parametry, které v praxi většinou neznáme. Pro naše účely tak potřebujeme právě testy dobré shody s neznámými parametry, proto se na ně v práci zaměříme.

Pro úplnost se ovšem na úvod zmiňme i o situaci, kdy by parametry byly neznáme. Základem testu je i v takovém případě tzv. Pearsonova statistika (5), v níž konfrontujeme námi naměřené hodnoty, tzv. empirické (skutečné) četnosti

$X_1, \dots, X_k$ , s jejich středními hodnotami, očekávanými (teoretickými) četnostmi  $np_1^0, \dots, np_k^0$ . Na hladině  $\alpha$  chceme testovat nulovou hypotézu  $H_0 : p_1 = p_1^0, \dots, p_k = p_k^0$ , která tvrdí, že skutečné hodnoty pravděpodobností multinomického rozdělení jsou rovny hypotetickým pravděpodobnostem  $p_1^0, \dots, p_k^0$ . Alternativou je, že alespoň jedna z rovností neplatí.

Pro zjištění četností vyjdeme z toho, že nejprve rozdělíme obor hodnot náhodné veličiny na  $k$  částí (tříd), které se vzájemně nepřekrývají. Pro každou třídu se pak (např. na základě známého pravděpodobnostního modelu) přidělí pravděpodobnost  $p_i^0$ , že náhodná veličina nabude hodnoty z  $i$ -té části. Následně se provede  $n$  pokusů, z čehož zjistíme četnosti  $X_1, \dots, X_k$ , tedy kolikrát realizace padnou do jednotlivých tříd. Jako testovací kritérium použijeme statistiku  $Z$ , která je nyní ve tvaru

$$Z = \sum_{i=1}^k \frac{(X_i - np_i^0)^2}{np_i^0}$$

a má za platnosti  $H_0$  asymptoticky  $\chi^2$ -rozdělení o  $k - 1$  stupních volnosti. Pokud hodnota testovací statistiky  $z$  překročí hodnotu  $\chi_{k-1}^2(1 - \alpha)$ , tedy hodnotu  $(1 - \alpha)$ -kvantilu příslušného rozdělení, nulovou hypotézu na hladině testu  $\alpha$  zamítáme. Stále je třeba mít na paměti, že test je asymptotický, tedy lze ho doporučit pro velký rozsah výběru, např. při splnění podmínky  $np_i^0 \geq 5$ , pro  $i = 1, \dots, k$ .

V praxi jsou většinou situace, kdy pravděpodobnosti  $p_1, \dots, p_k$  uvažovaného multinomického rozdělení závisí na nějakém neznámém vektorovém parametru  $\mathbf{a} = (a_1, \dots, a_m)$ . Řešíme pak situaci, jak pozměnit větu 4.1, aby vyhovovala i tomuto případu. Můžeme parametry odhadnout, následně odhadnout i pravděpodobnosti  $p_1, \dots, p_k$  a do vztahu (5) dosadit tyto odhady.

Předpokládáme tedy, že  $p_1 = p_1(\mathbf{a}), \dots, p_k = p_k(\mathbf{a})$ . Protože platí

$$p_1(\mathbf{a}) + \dots + p_k(\mathbf{a}) = 1,$$

dostáváme

$$\frac{\partial p_1(\mathbf{a})}{\partial a_j} + \dots + \frac{\partial p_k(\mathbf{a})}{\partial a_j} = 0, \quad j = 1, 2, \dots, m.$$



Vzniká otázka, jakým způsobem získat odhady parametru  $\mathbf{a}$ . Častým řešením je *metoda minimálního  $\chi^2$* , která je jistým způsobem analogií metody nejmenších čtverců a spočívá v tom, že jako odhad vezmeme tu hodnotu  $\mathbf{a}$ , která při daných  $X_1, \dots, X_k$  minimalizuje hodnotu  $Z$ .

Aplikujeme-li postup na statistiku  $Z$ , kterou ještě dále upravíme

$$Z(\mathbf{a}) = \sum_{i=1}^k \frac{(X_i - np_i(\mathbf{a}))^2}{np_i(\mathbf{a})} = \sum_{i=1}^k \frac{X_i^2}{np_i(\mathbf{a})} - n, \quad (6)$$

dostáváme

$$\frac{\partial Z(\mathbf{a})}{\partial a_j} = -\frac{1}{k} \sum_{i=1}^k \frac{X_i^2}{p_i^2(\mathbf{a})} \frac{\partial p_i(\mathbf{a})}{\partial a_j} = 0 \quad j = 1, \dots, m.$$

I přesto, že takto obdržená hodnota parametru  $\mathbf{a}$  opravdu minimalizuje  $Z(\mathbf{a})$ , právě díky tomu, že všechny druhé parciální derivace jsou kladné, je řešení této soustavy poměrně problematické. Používáme proto častěji *modifikovanou metodu minimálního  $\chi^2$* , která je chápána jako jisté zjednodušení předchozí metody. Proto pro další úvahy uvažujeme vztah (6) upravený do tvaru

$$\sum_{i=1}^k \frac{X_i}{p_i(\mathbf{a})} \frac{\partial p_i(\mathbf{a})}{\partial a_j} = 0, \quad j = 1, \dots, m. \quad (7)$$

Uvedená soustava má řešení v podobě odhadu parametru  $\mathbf{a}$  modifikovanou metodou minimálního  $\chi^2$ .

Mějme pro  $m < k - 1$  splněny podmínky:

1.  $p_1(\mathbf{a}) + \dots + p_k(\mathbf{a}) = 1$ ,
2.  $\exists c > 0$  takové, že  $p_i(\mathbf{a}) > c^2$  pro  $i = 1, \dots, k$ ,
3. Každá funkce  $p_i(\mathbf{a})$  má spojité derivace  $\frac{\partial p_i(\mathbf{a})}{\partial a_j}$  a  $\frac{\partial^2 p_i(\mathbf{a})}{\partial a_j \partial a_s}$ ,  $j, s = 1, 2, \dots, m$ .
4. Matice  $\left(\frac{\partial p_i(\mathbf{a})}{\partial a_j}\right)_{i=1}^k$   $\begin{matrix} k \\ j=1 \end{matrix}$  má hodnost  $m$ .

Nechť  $\mathbf{a}^0$  je vnitřním bodem nedegerovaného konečného intervalu  $A \in \mathbb{R}_m$ ,  $p_i^0 = p_i(\mathbf{a}^0)$  a vektor  $\mathbf{X} = (X_1, \dots, X_k)$  má multinomické rozdělení s parametry

$n, p_1^0, \dots, p_k^0$ . Pak soustava (7) má právě jeden kořen  $\mathbf{a}$  takový, že  $\mathbf{a}$  konverguje podle pravděpodobnosti pro  $n \rightarrow \infty$  k  $\mathbf{a}^0$ .

Náhodná proměnná

$$Z = \sum_{i=1}^k \frac{(X_i - np_i(\mathbf{a}))^2}{np_i(\mathbf{a})} \sim \chi_{k-1-m}^2$$

s odhadnutým parametrem  $\mathbf{a}$  je následně testovacím kritériem pro ověření hypotézy  $H_0$ , že výběr pochází z určitého rozdělení. Statistika  $Z$  má pro  $n \rightarrow \infty$  asymptoticky  $\chi_{k-m-1}^2$  rozdělení s  $k - m - 1$  stupni volnosti. Nulovou hypotézu tedy zamítáme tehdy, je-li hodnota testovací statistiky  $z \geq \chi_{k-m-1}^2(1 - \alpha)$ .

Uvedená soustava (7) se řeší iteračně, což je často poměrně komplikované, proto se jako odhad bere pouze vhodně zvolený počáteční krok iteračního procesu, nebo právě již zmíněné ML odhady parametrů. Druhého přístupu se přidržíme i v naší práci. Přes určitou „suboptimalitu“ tento postup ve většině praktických případů plně dostačuje. Dále viz skriptum Hron, Kunderová (2013), odst. před 7.3.1 (str. 197), od „Navíc, pokud...“. Každopádně opět je třeba mít na paměti podmínku na četnost tříd  $np_i \geq 5, \forall i = 1, \dots, k$ .

V další podkapitole si uvedenou teorii vyzkoušíme na testu normality, čímž si ukážeme postup, jak ověřit, že daný výběr pochází z normálního rozdělení.

## 4.4 Ověřování normálního rozdělení

Ověřování teoretických rozdělení je důležité zejména proto, že mnoho statistických testů předpokládá data pocházející z konkrétního známého rozdělení. V našem případě pak takto budeme chtít ověřit, že časová zátěž odpovídá určitému teoretickému rozdělení. V této kapitole takto provedeme test normálního rozdělení, protože je jedním z nejznámějších a nejčastěji se používá.

Předpokladem testu je náhodný výběr  $Y_1, \dots, Y_n$ , který přísluší námi zkoumané náhodné veličině  $Y$ . Testovanou nulovou hypotézou  $H_0$  je, že výběr pochází z normálního rozdělení  $N(\mu, \sigma^2)$ , přičemž parametry  $\mu$  a  $\sigma^2$  nejsou známy, tedy  $m = 2$ .

Postupujeme tak, že nejprve vytvoříme  $k$  třídících intervalů, které postupně označíme  $J_1, J_2, \dots, J_k$  pro  $k \geq 4$ ,

$$\underbrace{(-\infty, b_1)}_{J_1}, \underbrace{\langle b_1, b_2 \rangle}_{J_2}, \underbrace{\langle b_2, b_3 \rangle}_{J_3}, \dots, \underbrace{\langle b_{k-2}, b_{k-1} \rangle}_{J_{k-1}}, \underbrace{\langle b_{k-1}, \infty \rangle}_{J_k}.$$

Následně označíme středy příslušných tříd  $c_2, \dots, c_{k-1}$  tak, že

$$c_2 = \frac{b_1 + b_2}{2}, c_3 = \frac{b_2 + b_3}{2}, \dots, c_{k-1} = \frac{b_{k-2} + b_{k-1}}{2}.$$

„Krajní“ veličiny  $c_1$  a  $c_k$  volíme tak, aby  $c_1$  ležela v první třídě a měla od krajního bodu  $b_1$  stejnou vzdálenost jako střed druhé třídy, analogicky volíme i hodnotu  $c_k$ , aby ležela v poslední třídě a měla od krajního bodu  $b_k$  stejnou vzdálenost jako střed předposlední třídy.

Označíme-li  $X_i$  počet výběrových jednotek patřících do  $i$ -té třídy  $J_i$ , přičemž  $X_1 + X_2 + \dots + X_k = n$ , pak platí následující. Pochází-li výběr z rozdělení  $N(\mu, \sigma^2)$ , pak pravděpodobnost  $p_i$ , že náhodná veličina  $Y_j, j = 1, \dots, n$ , mající za platnosti nulové hypotézy stejné rozdělení jako zkoumaná náhodná veličina, padne do  $i$ -té třídy, je dána jako integrál z hustoty normálního rozdělení

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]$$

a tedy

$$p_i = p_i(\mu, \sigma^2) = P(Y \in J_i) = \int_{J_i} f(x) dx \quad i = 1, 2, \dots, k.$$

Rozepíšeme-li konkrétně každou pravděpodobnost  $p_1, p_2, \dots, p_k$  zvlášť, obdržíme následující

- $p_1 = p_1(\mu, \sigma^2) = P(Y \in J_1) = P(-\infty < Y \leq b_1) = \int_{-\infty}^{b_1} f(x) dx$
- $p_2 = p_2(\mu, \sigma^2) = P(Y \in J_2) = P(b_1 < Y \leq b_2) = \int_{b_1}^{b_2} f(x) dx$
- $\vdots$
- $p_{k-1} = p_{k-1}(\mu, \sigma^2) = P(Y \in J_{k-1}) = P(b_{k-2} < Y \leq b_{k-1}) = \int_{b_{k-2}}^{b_{k-1}} f(x) dx$

- $p_k = p_k(\mu, \sigma^2) = P(Y \in J_k) = P(b_{k-1} < Y \leq b_k) = \int_{b_{k-1}}^{\infty} f(x)dx$ .

Potom náhodný vektor  $\mathbf{X} = (X_1, \dots, X_k)$  má zřejmě multinomické rozdělení s parametry  $p_1(\mu, \sigma^2), \dots, p_k(\mu, \sigma^2)$ . Zároveň pro pravděpodobnosti platí následující derivace podle parametrů  $\mu$  a  $\sigma^2$

$$\begin{aligned}\frac{\partial p_i}{\partial \mu} &= \frac{1}{\sigma^2} \int_{J_i} (x - \mu) f(x) dx \\ \frac{\partial p_i}{\partial \sigma} &= \frac{1}{\sigma^3} \int_{J_i} (x - \mu)^2 f(x) dx - \frac{p_i}{\sigma}.\end{aligned}$$

Dosadíme-li derivace do vzorce (7) a upravíme, dostáváme parametry vyjádřené ve tvaru

$$\mu = \frac{1}{n} \sum_{i=1}^k \frac{X_i}{p_i} \int_{J_i} x f(x) dx \quad (8)$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^k \frac{X_i}{p_i} \int_{J_i} (x - \mu)^2 f(x) dx. \quad (9)$$

Pro získání odhadů  $\hat{\mu}, \hat{\sigma}^2$  využijeme iterační metody, přičemž za počáteční aproximace se volí

$$\mu_0 = \frac{1}{n} \sum_{i=1}^k X_i c_i, \quad \sigma_0^2 = \frac{1}{n} \sum_{i=1}^k X_i c_i^2 - \mu_0^2.$$

Následně vypočteme hodnoty pravých stran rovnic (8), (9). Pro výpočet hodnoty  $p_i$  můžeme využít vztahu pro výpočet pravděpodobností

$$p_i = \Phi\left(\frac{b_i - \mu}{\sigma}\right) - \Phi\left(\frac{b_{i-1} - \mu}{\sigma}\right),$$

vztahu pro vyjádření integrálu z (8)

$$\int_{J_i} x f(x) dx = \mu p_i + \frac{\sigma}{\sqrt{2\pi}} \left[ \exp\left\{-\frac{(b_{i-1} - \mu)^2}{2\sigma^2}\right\} - \exp\left\{-\frac{(b_i - \mu)^2}{2\sigma^2}\right\} \right]$$

a pro integrál z (9),

$$\int_{J_i} (x - \mu)^2 f(x) dx = \sigma^2 p_i + \frac{\sigma^2}{\sqrt{2\pi}} \left[ \frac{b_{i-1} - \mu}{\sigma} \exp \left\{ -\frac{(b_{i-1} - \mu)^2}{2\sigma^2} \right\} - \frac{b_{i-1} - \mu}{\sigma} \exp \left\{ -\frac{(b_i - \mu)^2}{2\sigma^2} \right\} \right].$$

Pomocí těchto uvedených vzorců získáme novou aproximaci  $\mu_1, \sigma_1^2$  a celý postup opakujeme s těmito hodnotami. Alternativně pak můžeme výpočet odhadů parametrů  $\mu, \sigma^2$  provést též pomocí metody maximální věrohodnosti z realizace vstupního výběru  $Y_1, \dots, Y_n$ . Nakonec stanovíme hodnotu  $Z$  ze vztahu (5). Vyjde-li hodnota  $z \geq \chi_{k-3}^2(\alpha)$ , zamítneme nulovou hypotézu o normálním rozdělení na hladině, která je asymptoticky  $\alpha$ .

## 5 Praktická část

Nyní už přejdeme k samotným výpočtům. Tato část diplomové práce je věnována aplikaci výše uvedené teorie na konkrétní data. Data byla poskytnuta Českým statistickým úřadem a odráží časovou zátěž v sekundách vybraných otázek v šetření Životní podmínky v roce 2013. Pro zpracování dat byl použit statistický software R, a to zejména proto, že je volně dostupný a v současné době patří k neoblíbenějším softwarovým nástrojům pro statistickou analýzu. Instalační soubory si můžeme zdarma stáhnout na internetové stránce <http://r-project.org>.

Cílem této kapitoly a současně cílem práce je najít skutečná pravděpodobnostní rozdělení časové zátěže deseti vybraných proměnných. Pro tento účel použijeme právě testy dobré shody, které jsme uvažovali v teoretické části práce, a u každé otázky takto otestujeme pět teoretických rozdělení pravděpodobnosti uvedených v kapitole 3. Výstupem každého z testů tak bude hodnota testovací statistiky, kterou lze porovnávat s výstupy testů pro další rozdělení. Vzhledem k velkému rozsahu výběru a jeho vlivu na statistiku  $Z$  bychom totiž nulovou hypotézu prakticky ve všech případech zamítali, aniž by měla tato skutečnost opravdu odpovídající hodnotu.

Proto, abychom mohli stanovit hodnotu testovacího kritéria  $Z$  (5), je třeba stanovit vektor empirických četností  $\mathbf{X}$  a teoretických četností  $n\mathbf{p}$ . V prvním případě (při stanovení skutečných četností) vyjdeme z intervalového rozdělení časové zátěže,  $n$  je počet měření, tedy zaznamenaných časových okamžiků jednotlivé proměnné, a  $\mathbf{p}$  (pravděpodobnosti realizace veličiny s hypotetickým rozdělením v daných intervalech) získáme jako rozdíl hodnot distribuční funkce uvažovaného rozdělení v krajních bodech intervalu.

Celý postup výpočtu, včetně načtení dat, si podrobněji ukážeme na první otázce A6\_UK\_VZD. Otázka se týká roku ukončení nejvyššího vzdělání. Je to tedy otázka objektivní, odpověď na ni souvisí s vypsáním číslice, respektive respondent odpoví vypsáním posledního dvojčíslí roku, v němž dosáhl nejvyššího vzdělání. U dalších otázek už se potom zaměříme zejména na výčet výsledků a okomentujeme až závěr testu.

## 5.1 Testování otázky A6\_UK\_VZD

Vzhledem k tomu, že chceme pracovat s reálnými daty, musíme si je nejprve načíst. Učiníme tak prostřednictvím jednoduchého příkazu `data=read.csv2("data.csv", na.strings="0")`. Funkci `read.csv2` použijeme proto, že chceme načíst datovou tabulku, která nese název *data*, je ve formátu `.csv` a sloupce tabulky jsou odděleny středníkem. Pomocí funkce `na.strings="0"` označíme nulové časové okamžiky jako chybějící hodnoty. Nuly, které vlastně odpovídají nerealizaci otázky, tedy z datového souboru záměrně vypustíme, abychom záhy mohli odhadovat parametry rozdělení metodou maximální věrohodnosti.

Nyní přiřadíme sloupce tabulky příslušným otázkám. K přiřazení otázky prvnímu sloupci tabulky použijeme následující příkaz `A6_UK_VZD<-data[,1][!is.na(data[,1])]`, který současně z dat odstraní chybějící hodnoty negací funkce `is.na`.

Jakmile máme načtena data a přiřazené otázky, můžeme s nimi samostatně pracovat. Nejprve je dobré udělat si o časové zátěži otázky přehled formou číselných statistik. Vypočteme si za tímto účelem průměr, rozptyl, resp. směrodatnou odchylku, a 99% kvantil.

```
> mean(A6\_UK\_VZD)
[1] 11.51654
> var(A6\_UK\_VZD)
[1] 1828.679
> sd(A6\_UK\_VZD)
[1] 42.76305
> quantile(A6\_UK\_VZD, probs=c(.99))
 99%
139.86
```

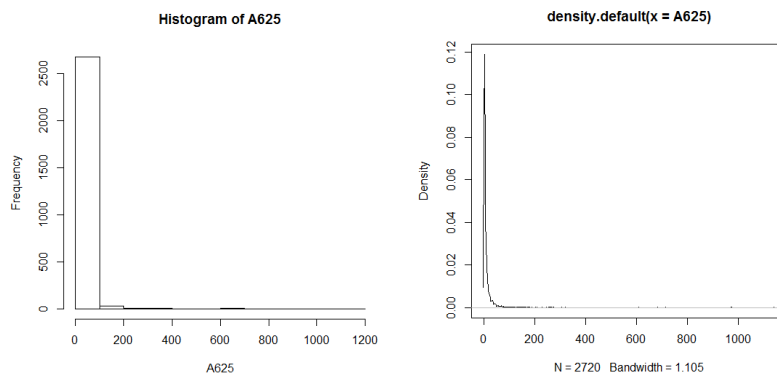
Čas zodpovězení otázky se v průměru pohybuje okolo 11,5 sekundy, směrodatná odchylka vyjadřuje variabilitu kolem průměru a vychází bezmála 43 sekund. Nakonec 99% kvantil je 139,86 sekundy, znamená to tedy, že 99% všech hodnot

má menší časovou zátěž než 139,86 sekund. Při následném zobrazení dat nám výpočty pomohou udělat si představu o případném výskytu extrémních hodnot.

Za tímto účelem vykreslíme histogram a můžeme přidat orientačně i graf hustoty.

```
> hist(A6\UK\VZD,main="Histogram of A6")
> plot(density(A6\UK\VZD),main="Density estimate of A6")
```

Z histogramu je patrné, že skutečně 99% všech hodnot se nachází v intervalu od nuly do sta, tedy před výše uvedeným kvantilem.



Obrázek 27: Histogram a hustota otázky A6\_UK\_VZD

Vzhledem k tomu, že testy dobré shody jsou obecně založeny na třídících intervalech, musíme je nejprve z dat vytvořit. Původně jsme postupovali tak, že jsme si pomocí Sturgesova pravidla spočítali počet tříd  $k$  jako

$$k = 1 + \log_2(n)$$

V tomto případě vyšlo  $k = 13$ , tedy sestavili jsme třináct třídících intervalů. Protože se v souboru vyskytují odlehlé hodnoty, použili jsme vypočtený kvantil  $k$  tomu, abychom hodnoty za touto hodnotou seskupili do jednoho koncového intervalu. Zbytek hodnot jsme roztřídlili do zbylých dvanácti intervalů, tak aby byla zachována stejná šířka všech intervalů. Ukázalo se, že při menším počtu intervalů, s tím, že nepatrně zvýšíme šířku, je hodnota testovacího kritéria u



většiny testů nižší, tedy test bude konzervativnější. Neubráníme se však tomu, že četnost posledního intervalu bude převyšovat četnost několika bezprostředně předcházejících intervalů.

Nakonec jsme zvolili toto kompromisní řešení. Sestavili jsme osm intervalů a šířku každého intervalu, kromě toho posledního, stanovili deset. Intervalové rozdělení vypadá takto.

$$\underbrace{\langle 1, 11 \rangle}_{X_1=2123}, \underbrace{\langle 11, 21 \rangle}_{X_2=287}, \underbrace{\langle 21, 31 \rangle}_{X_3=114}, \underbrace{\langle 31, 41 \rangle}_{X_4=66}, \underbrace{\langle 41, 51 \rangle}_{X_5=39}, \underbrace{\langle 51, 61 \rangle}_{X_6=16}, \underbrace{\langle 61, 71 \rangle}_{X_7=16}, \underbrace{\langle 71, 1139 \rangle}_{X_8=61}$$

Vektor četností vypadá následovně a budeme jej obecně značit  $\mathbf{x}$  vždy s přidáním čísla proměnné.

```
> xa6=c(2123, 287, 114, 66, 39, 16, 16, 61)
```

Vidíme, že s rostoucím časovou zátěží četnost klesá, což se dalo očekávat. Poslední interval seskupuje hodnoty od 71 sekund až po krajní hodnotu 1139 sekund a svou četností převyšuje předchozí. Vzhledem k tomu, že rozsah souboru je 2720, je téměř 90% hodnot obsaženo v prvních dvou intervalech. Tedy téměř 90% respondentů zodpoví otázku během 21 sekund.

Z krajních bodů intervalů sestavíme dva vektory  $\mathbf{a6\_1}$  a  $\mathbf{a6\_2}$ , které budeme potřebovat pro výpočet vektoru pravděpodobností  $p_i$ ,  $i = 1, \dots, 8$ .

```
> a6_1=c(11, 21, 31, 41, 51, 61, 71, 1139)
```

```
> a6_2=c(1, 11, 21, 31, 41, 51, 61, 71)
```

Protože vektor pravděpodobností už závisí na konkrétním rozdělení, budeme pro jeho výpočet potřebovat odhady parametrů příslušného rozdělení. K tomu, abychom získali odhady, můžeme využít dvou způsobů, výsledky oběma metodami se víceméně neliší. V prvním případě můžeme pro odhad parametrů použít funkci `fitdistr()`, která bere za parametr zvolená data, ze kterých chceme odhadovat, a zadáme testované rozdělení. Nejprve si ale musíme načíst knihovnu MASS. Například pro gamma rozdělení vypadá odhad následovně.

```

> library(MASS)
> fitdistr(A6\_UK\_VZD,"gamma")
      shape      rate
0.561784844  0.048780683

```

Zde `shape` odpovídá parametru  $p$  a `rate` parametru  $a$ .

Druhá možnost je, že parametry odhadneme přímo metodou maximální věrohodnosti, jak byla popsána v kapitole 4.1, a to pomocí funkce `mle`. Pro tyto účely nejprve musíme načíst knihovnu `stats4`.

```

> library(stats4)

```

Následně vyjádříme logaritmus věrohodnostní funkce s opačným znaménkem, tzv. *negative loglikelihood function*, označíme `ll`. Pro případ `gamma` rozdělení vypadá funkce takto

$$ll = -np \ln(a) + n \ln(\Gamma(p)) + (p - 1) \sum_{i=1}^n \ln(x_i) + a \sum_{i=1}^n x_i.$$

Jako startovací hodnoty parametrů uvažujeme hodnoty  $a = 1$ ,  $p = 2$ , rozsah výběru u otázky `A6\_UK\_VZD` označíme `nA6`, `x` jsou naše data pro otázku `A6`.

```

> nA6=2720
> x=A6\_UK\_VZD
> ll=function(a,p)
+ ll=-n*p*log(a)+n*log(gamma(p))-(p-1)*sum(log(x))+a*sum(x)
> mle(minuslog=ll,start=list(a=1,p=2))

```

Call:

```
mle(minuslogl = ll, start = list(a = 1, p = 2))
```

Coefficients:

```

      a      p
0.04878895 0.56184704

```

Takto také získáme požadované odhady. Vektor pravděpodobností vypočítáme jako rozdíl hodnot distribuční funkce v krajních bodech intervalů. Nyní tyto pravděpodobnosti pro všechny testy vypočteme a můžeme rovnou dosadit do testovací statistiky  $Z$ . Zde máme výsledky pro jednotlivé testy uvedeny.

### Normální rozdělení

```
> pa6n=pnorm(a6_1,mean=11.5165441,sd=42.7551908)
      -pnorm(a6_2,mean=11.5165441,sd=42.7551908)
> pa6n
[1] 0.09232804 0.09258812 0.08792799 0.07907684
[5] 0.06734752 0.05431803 0.04148742 0.08207376

> za6n=sum((xa6-sum(xa6)*pa6n)^2/(sum(xa6)*pa6n))
> za6n
[1] 14545.8
```

Pro zajímavost hodnotu testovacího kritéria porovnáme s kvantilem  $\chi^2$ -rozdělení o pěti stupních volnosti, které jsme stanovili jako  $k - m - 1$ . Počet tříd je totiž  $k = 8$ , počet parametrů je  $m = 2$ . Testujeme na hladině významnosti  $\alpha = 0.05$ , tj. příslušný kvantil  $\chi_{0.95}^2(5) = 11.09$ . Hodnota  $z$  díky velkému rozsahu výběru skutečně výrazně převyšuje hodnotu kvantilu,  $H_0$  bychom tak byli nuceni zamítnout. Je tak zřejmé, že bude mít smysl spíše porovnávat hodnoty  $Z$  pro různá rozdělení.

### Exponenciální rozdělení

```
> pa6e=pexp(a6_1,rate=0.08683160)-pexp(a6_2,rate=0.08683160)
> pa6e
[1] 0.532076127 0.223289821 0.093705283 0.039324140
[5] 0.016502676 0.006925475 0.002906328 0.002101627

> za6e=sum((xa6-sum(xa6)*pa6e)^2/(sum(xa6)*pa6e))
> za6e
```

```
[1] 1121.025
```

### Lognormální rozdělení

```
> pa6l=plnorm(a6_1,meanlog=1.33309577,sdlog=1.29906789)
      -plnorm(a6_2,meanlog=1.33309577,sdlog=1.29906789)
```

```
> pa6l
```

```
[1] 0.641396174 0.112355944 0.040933493 0.019472092
```

```
[5] 0.010716788 0.006478665 0.004184616 0.012056542
```

```
> za6l=sum((xa6-sum(xa6)*pa6l)^2/(sum(xa6)*pa6l))
```

```
> za6l
```

```
[1] 115.3961
```

### Gamma rozdělení

```
> pa6g=pgamma(a6_1,shape=0.561784844, rate=0.048780683)
      -pgamma(a6_2,shape=0.561784844, rate=0.048780683)
```

```
> pa6g
```

```
[1] 0.45898665 0.16240649 0.07970531 0.04227043
```

```
[5] 0.02326581 0.01309035 0.00747340 0.01039794
```

```
> za6g=sum((xa6-sum(xa6)*pa6g)^2/(sum(xa6)*pa6g))
```

```
> za6g
```

```
[1] 793.9433
```

### Weibullovo rozdělení

```
> pa6w=pweibull(a6_1,shape=0.669673555,scale=7.547438211)
      -pweibull(a6_2,shape=0.669673555,scale=7.547438211)
```

```
> pa6w
```

```
[1] 0.496228818 0.138646649 0.061366710 0.031324244
```

```
[5] 0.017314311 0.010091359 0.006113734 0.011260923
```

```

> za6w=sum((xa6-sum(xa6)*pa6w)^2/(sum(xa6)*pa6w))
> za6w
[1] 520.641

```

Výsledky shrneme do přehledné tabulky, v níž kromě hodnoty testovací statistiky též uvedeme příslušné relativní četnosti („empirické pravděpodobnosti“) a odpovídající teoretické pravděpodobnosti daného rozdělení v jednotlivých intervalech. Empirickou pravděpodobnost vypočítáme jako poměr empirické četnosti  $x_{A6}$  k celkovému rozsahu souboru  $n_{A6}$ .

	Empirické pravděpodobnosti	Teoretické pravděpodobnosti	$Z$
Normální rozdělení	0.09 0.09 0.08 0.08 0.07 0.05 0.04 0.08	0.78 0.11 0.04 0.02 0.01 0.01 0.01 0.02	14545.8
Lognormální rozdělení	0.64 0.11 0.04 0.02 .01 0.01 0.00 0.01	0.78 0.11 0.04 0.02 0.01 0.01 0.01 0.02	115.4
Exponenciální rozdělení	0.53 0.22 0.09 0.04 .02 0.01 0.00 0.00	0.78 0.11 0.04 0.02 0.01 0.01 0.01 0.02	1121.0
Gama rozdělení	0.46 0.16 0.08 0.04 0.02 0.01 0.01 0.01	0.78 0.11 0.04 0.02 0.01 0.01 0.01 0.02	793.9
Weibullovo	0.50 0.14 0.06 0.03 0.02 0.01 0.01 0.01	0.78 0.11 0.04 0.02 0.01 0.01 0.01 0.02	520.6

Tabulka 2: Výsledky testu pro otázku A6\_UK\_VZD.

## 5.2 Testy ostatních proměnných

V této podkapitole uvedeme výsledky ostatních testů, které jsme získali stejným způsobem jako v předchozím případě u otázky A6\_UK\_VZD. Za tímto účelem sestavíme deset tabulek. V první shrneme sestavené třídící intervaly a jejich četnosti. Další devět tabulek bude obsahovat vypočtené empirické a teoretické pravděpodobnosti a hodnotu testovacího kritéria  $Z$  pro jednotlivé testy pro zbylé proměnné.

	Intervalová rozdělení	Četnosti
B9	$\langle 1, 11 \rangle, \langle 11, 21 \rangle, \langle 21, 31 \rangle, \langle 31, 41 \rangle$ $\langle 41, 51 \rangle, \langle 51, 61 \rangle, \langle 61, 71 \rangle, \langle 71, 282 \rangle$	174, 148, 47, 24, 8, 7, 4, 11
B13	$\langle 1, 21 \rangle, \langle 21, 41 \rangle, \langle 41, 61 \rangle, \langle 61, 81 \rangle$ $\langle 81, 101 \rangle, \langle 101, 121 \rangle, \langle 121, 141 \rangle, \langle 141, 2113 \rangle$	2550, 505, 188, 103, 48, 31, 24, 82
B18	$\langle 1, 31 \rangle, \langle 31, 61 \rangle, \langle 61, 91 \rangle, \langle 91, 121 \rangle$ $\langle 121, 151 \rangle, \langle 151, 181 \rangle, \langle 181, 211 \rangle, \langle 211, 2990 \rangle$	2535, 621, 192, 95, 47, 32, 22, 55
B27	$\langle 1, 21 \rangle, \langle 21, 41 \rangle, \langle 41, 61 \rangle, \langle 61, 81 \rangle$ $\langle 81, 101 \rangle, \langle 101, 121 \rangle, \langle 121, 141 \rangle, \langle 141, 1496 \rangle$	3048, 269, 95, 42, 23, 14, 15, 50
C15	$\langle 1, 21 \rangle, \langle 21, 41 \rangle, \langle 41, 61 \rangle, \langle 61, 81 \rangle$ $\langle 81, 101 \rangle, \langle 101, 121 \rangle, \langle 121, 141 \rangle, \langle 141, 2684 \rangle$	947, 571, 160, 70, 30, 19, 12, 53
C17	$\langle 1, 11 \rangle, \langle 11, 21 \rangle, \langle 21, 31 \rangle, \langle 31, 41 \rangle$ $\langle 41, 51 \rangle, \langle 51, 61 \rangle, \langle 61, 71 \rangle, \langle 71, 1792 \rangle$	1565, 155, 44, 23, 10, 7, 5, 14
C21	$\langle 1, 21 \rangle, \langle 21, 41 \rangle, \langle 41, 61 \rangle, \langle 61, 81 \rangle$ $\langle 81, 101 \rangle, \langle 101, 121 \rangle, \langle 121, 141 \rangle, \langle 141, 1282 \rangle$	1456, 215, 79, 40, 11, 3, 2, 13
CM4	$\langle 1, 51 \rangle, \langle 51, 101 \rangle, \langle 101, 151 \rangle, \langle 151, 201 \rangle$ $\langle 201, 251 \rangle, \langle 251, 301 \rangle, \langle 301, 351 \rangle, \langle 351, 1985 \rangle$	1365, 835, 188, 68, 37, 20, 9, 34
CM8	$\langle 1, 21 \rangle, \langle 21, 41 \rangle, \langle 41, 61 \rangle, \langle 61, 81 \rangle$ $\langle 81, 101 \rangle, \langle 101, 121 \rangle, \langle 121, 141 \rangle, \langle 141, 5570 \rangle$	1243, 730, 236, 108, 68, 26, 21, 83

Tabulka 3: Sestavené intervaly a jejich četnosti pro ostatní otázky.

U všech otázek se jevílo jako nejlepší sestavit právě osm intervalů. Šířka intervalů je však pro jednotlivé otázky různá. U třech otázek (včetně A6\_UK\_VZD) nejvíce vyhovuje šířka deset, u poloviny dvacet, u otázky B18\_SLUZ\_VYR třicet a u otázky CM4\_POCITY dokonce padesát. Šířka intervalu třicet a padesát svědčí o tom, že jsou tyto otázky charakterizovány vyšší časovou zátěží. Četnosti mají potom všechny podobnou strukturu, první dvě četnosti jsou v řádech stovek nebo tisíců, následně klesají do desítek, četnost posledního širšího intervalu, který obsahuje koncová rozptýlená pozorování, převyšuje četnost předchozích třech, čtyřech intervalů.

Nyní uvedeme za sebou devět tabulek, v nichž pro každou proměnnou, kromě otázky A6\_UK\_VZD, zobrazíme vypočtené pravděpodobnosti a hodnoty testovací statistiky pro jednotlivá testovaná rozdělení. U každé otázky se pokusíme o krátký komentář výsledků.

### Otázka B9\_VYSE\_HYPO

	Empirické pravděpodobnosti	Teoretické pravděpodobnosti	Z
Normální rozdělení	0.14 0.17 0.16 0.13 0.09 0.05 0.02 0.01	0.41 0.35 0.11 0.06 0.02 0.02 0.01 0.03	14545.8
Exponenciální rozdělení	0.39 0.23 0.14 0.08 0.05 0.03 0.02 0.02	0.41 0.35 0.11 0.06 0.02 0.02 0.01 0.03	41.0
Lognormální rozdělení	0.39 0.35 .015 0.06 0.03 0.01 0.01 0.01	0.41 0.35 0.11 0.06 0.02 0.02 0.01 0.03	18.6
Gama rozdělení	0.35 0.30 0.17 0.09 0.04 0.02 0.01 0.01	0.41 0.35 0.11 0.06 0.02 0.02 0.01 0.03	49.9
Weibullovo rozdělení	0.37 0.26 0.15 0.09 0.05 0.03 0.01 0.01	0.41 0.35 0.11 0.06 0.02 0.02 0.01 0.03	40.3

Tabulka 4: Výsledky testu pro otázku B9\_VYSE\_HYPO

### Otázka B13\_ODH\_CEN

	Empirické pravděpodobnosti	Teoretické pravděpodobnosti	Z
Normální rozdělení	0.12 0.12 0.11 0.10 0.07 0.05 0.03 0.04	0.72 0.14 0.05 0.03 0.01 0.01 0.01 0.02	12050.3
Exponenciální rozdělení	0.54 0.24 0.10 0.05 0.02 0.010 0.00 0.00	0.72 0.14 0.05 0.03 0.01 0.01 0.01 0.02	956.7
Lognormální rozdělení	0.68 0.13 0.05 0.03 0.02 0.01 0.01 0.02	0.72 0.14 0.05 0.03 0.01 0.01 0.01 0.02	18.6
Gama rozdělení	0.51 0.18 0.09 0.05 0.03 0.02 0.01 0.01	0.72 0.14 0.05 0.03 0.01 0.01 0.01 0.02	481.1
Weibullovo rozdělení	0.55 0.16 0.07 0.04 0.02 0.01 0.01 0.01	0.72 0.14 0.05 0.03 0.01 0.01 0.01 0.02	247.7

Tabulka 5: Výsledky testu pro otázku B13\_ODH\_CEN

V tabulce vidíme pravděpodobnosti a hodnoty  $Z$  pro otázky B9 a B13. Při porovnávání hodnot statistiky pro různá rozdělení vidíme, že nejmenší je test pro lognormální rozdělení. Tento typ proměnné bychom tak snad mohli aproximovat právě lognormálním rozdělením.

### Otázka B18\_SLUZ\_VYR

	Empirické pravděpodobnosti				Teoretické pravděpodobnosti				Z
Normální rozdělení	0.14	0.14	0.13	0.10	0.71	0.17	0.05	0.03	9295.2
	0.07	0.05	0.03	0.02	0.01	0.01	0.01	0.02	
Exponenciální rozdělení	0.55	0.24	0.11	0.05	0.71	0.17	0.05	0.03	545.6
	0.02	0.01	0.00	0.00	0.01	0.01	0.01	0.02	
Lognormální rozdělení	0.67	0.21	0.07	0.03	0.71	0.17	0.05	0.03	115.3
	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.02	
Gama rozdělení	0.55	0.24	0.11	0.05	0.71	0.17	0.05	0.03	589.0
	0.02	0.01	0.00	0.00	0.01	0.01	0.01	0.02	
Weibullovo rozdělení	0.56	0.21	0.10	0.05	0.71	0.17	0.05	0.03	326.2
	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02	

Tabulka 6: Výsledky testu pro otázku B18\_SLUZ\_VYR

### Otázka B27\_PRIJ\_ZPRON

	Empirické pravděpodobnosti				Teoretické pravděpodobnosti				Z
Normální rozdělení	0.13	0.13	0.11	0.09	0.69	0.17	0.05	0.02	15521.6
	0.06	0.04	0.02	0.02	0.02	0.01	0.01	0.03	
Exponenciální rozdělení	0.64	0.21	0.07	0.02	0.69	0.17	0.05	0.02	2586.6
	0.01	0.00	0.00	0.00	0.02	0.01	0.01	0.03	
Lognormální rozdělení	0.83	0.12	0.03	0.01	0.69	0.17	0.05	0.02	690.6
	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.03	
Gama rozdělení	0.61	0.19	0.07	0.03	0.69	0.17	0.05	0.02	1342.0
	0.01	0.00	0.00	0.00	0.02	0.01	0.01	0.03	
Weibullovo rozdělení	0.63	0.16	0.06	0.02	0.69	0.17	0.05	0.02	760.0
	0.01	0.01	0.00	0.00	0.02	0.01	0.01	0.03	

Tabulka 7: Výsledky testu pro otázku B27\_PRIJ\_ZPRON

Další dvě testované otázky B18 a B27 jsou alternativní. Nejvyšší je hodnota statistiky opět pro test normálního rozdělení, nejvyšší naopak pro lognormální rozdělení.



### Otázka C15\_HL\_ZAM

Jediným zástupcem skupiny vypisovacích otázek je otázka C15, i zde je hodnota testu nejnižší pro lognormální rozdělení.

	Empirické pravděpodobnosti	Teoretické pravděpodobnosti	$Z$
Normální rozdělení	0.06 0.06 0.06 0.06 0.06 0.05 0.05 0.21	0.51 0.31 0.09 0.04 0.02 0.01 0.01 0.03	8371.7
Exponenciální rozdělení	0.38 0.23 0.14 0.09 0.05 0.03 0.02 0.03	0.51 0.31 0.09 0.04 0.02 0.01 0.01 0.03	303.0
Lognormální rozdělení	0.50 0.26 0.11 0.05 0.03 0.02 0.01 0.02	0.51 0.31 0.09 0.04 0.02 0.01 0.01 0.03	51.7
Gama rozdělení	0.39 0.22 0.13 0.08 0.05 0.03 0.02 0.35	0.51 0.31 0.09 0.04 0.02 0.01 0.01 0.03	297.8
Weibullovo rozdělení	0.44 0.20 0.11 0.07 0.04 0.03 0.02 0.04	0.51 0.31 0.09 0.04 0.02 0.01 0.01 0.03	243.1

Tabulka 8: Výsledky testu pro otázku C15\_HL\_ZAM

### Otázka C17\_POST\_ZAM

Stejným typem otázky, a to výběrové, jsou otázky C17 a modulová proměnná CM4, opět s preferencí lognormálního rozdělení.

	Empirické pravděpodobnosti	Teoretické pravděpodobnosti	$Z$
Normální rozdělení	0.08 0.08 0.07 0.07 0.06 0.05 0.04 0.12	0.86 0.09 0.02 0.01 0.01 0.00 0.00 0.01	15168.67
Exponenciální rozdělení	0.58 0.21 0.08 0.03 0.01 0.00 0.00 0.00	0.86 0.09 0.02 0.01 0.01 0.00 0.00 0.01	589.447
Lognormální rozdělení	0.81 0.13 0.03 0.01 0.00 0.00 0.00 0.00	0.86 0.09 0.02 0.01 0.01 0.00 0.00 0.01	322.934
Gama rozdělení	0.56 0.20 0.08 0.03 0.01 0.00 0.00 0.00	0.86 0.09 0.02 0.01 0.01 0.00 0.00 0.01	560.0639
Weibullovo rozdělení	0.56 0.16 0.06 0.03 0.01 0.01 0.00 0.00	0.86 0.09 0.02 0.01 0.01 0.00 0.00 0.01	439.5926

Tabulka 9: Výsledky testu pro otázku C17\_POST\_ZAM

### Otázka CM4\_POCITY

	Empirické pravděpodobnosti	Teoretické pravděpodobnosti	Z
Normální rozdělení	0.21 0.24 0.19 0.10 0.04 0.01 0.00 0.00	0.05 0.33 0.07 0.03 0.01 0.00 0.00 0.01	3230.0
Exponenciální rozdělení	0.51 0.25 0.12 0.06 0.03 0.01 0.01 0.01	0.05 0.33 0.07 0.03 0.01 0.00 0.00 0.01	201.3
Lognormální rozdělení	0.50 0.34 0.10 0.04 0.01 0.01 0.01 0.00	0.05 0.33 0.07 0.03 0.01 0.00 0.00 0.01	135.8
Gama rozdělení	0.50 0.35 0.14 0.05 0.01 0.00 0.00 0.00	0.05 0.33 0.07 0.03 0.01 0.00 0.00 0.01	1067.3
Weibullovo rozdělení	0.47 0.29 0.14 0.06 0.02 0.01 0.00 0.00	0.05 0.33 0.07 0.03 0.01 0.00 0.00 0.01	414.0

Tabulka 10: Výsledky testu pro otázku CM4\_POCITY

### Otázka C21\_ODPRAC\_HOD

	Empirické pravděpodobnosti	Teoretické pravděpodobnosti	Z
Normální rozdělení	0.18 0.18 0.14 0.09 0.04 0.02 0.01 0.00	0.8 0.12 0.04 0.02 0.01 0.00 0.00 0.01	4123.3
Exponenciální rozdělení	0.65 0.20 0.07 0.02 0.01 0.00 0.00 0.00	0.8 0.12 0.04 0.02 0.01 0.00 0.00 0.01	432.8
Lognormální rozdělení	0.79 0.15 0.04 0.01 0.01 0.00 0.00 0.00	0.8 0.12 0.04 0.02 0.01 0.00 0.00 0.01	66.4
Gama rozdělení	0.65 0.21 0.06 0.02 0.01 0.00 0.00 0.00	0.8 0.12 0.04 0.02 0.01 0.00 0.00 0.01	610.4
Weibullovo rozdělení	0.64 0.19 0.06 0.02 0.01 0.00 0.00 0.00	0.8 0.12 0.04 0.02 0.01 0.00 0.00 0.01	229.1

Tabulka 11: Výsledky testu pro otázku C21\_ODPRAC\_HOD

Otázka C21 se ptá na odpracované hodiny za týden. Je typově stejná jako otázka A6, proto zde nejnižší hodnota testovací statistiky pro lognormální rozdělení ani moc nepřekvapí.

## Otázka CM8\_DUVERA

	Empirické pravděpodobnosti				Teoretické pravděpodobnosti				$Z$
Normální rozdělení	0.06	0.06	0.06	0.06	0.49	0.29	0.09	0.04	10133.4
	0.06	0.05	0.05	0.20	0.03	0.01	0.01	0.03	
Exponenciální rozdělení	0.40	0.24	0.14	0.08	0.49	0.29	0.09	0.04	41.0
	0.05	0.03	0.02	0.02	0.03	0.01	0.01	0.03	
Lognormální rozdělení	0.47	0.28	0.12	0.06	0.49	0.29	0.09	0.04	60.7
	0.03	0.02	0.01	0.02	0.03	0.01	0.01	0.03	
Gama rozdělení	0.39	0.25	0.15	0.08	0.49	0.29	0.09	0.04	261.1
	0.05	0.03	0.02	0.02	0.03	0.01	0.01	0.03	
Weibullovo rozdělení	0.43	0.22	0.13	0.07	0.49	0.29	0.09	0.04	8.5
	0.05	0.03	0.02	0.03	0.03	0.01	0.01	0.03	

Tabulka 12: Výsledky testu pro otázku CM8\_DUVERA

Nakonec subjektivní škálová otázka CM8. Zde se řádově hodně odlišuje test Weibullova rozdělení. U tohoto typu proměnné bychom mohli, narozdíl od předěšlých případů, usuzovat na Weibullovo rozdělení. I tak ale všechny testy zamítáme.

### 5.3 Statistika $\phi$

Vzhledem k velkému rozsahu výběru není překvapením, že všechny testy byly zamítnuty ve prospěch alternativy. Proto se v takových případech pro porovnání hodnot testovací statistiky  $Z$ , navíc s eliminací konkrétního rozsahu výběru, užívá bezrozměrná statistika  $\phi$ , viz [3], jejíž hodnota je dána jako odmocnina z podílu testovací statistiky  $Z$  a rozsahu souboru  $n$  pro jednotlivou otázku,

$$\phi = \sqrt{\frac{Z}{n}}.$$

Pro každou proměnnou (včetně A6\_UK\_VZD) vypočteme hodnotu statistiky  $\phi$  a výsledky shrneme do tabulky 13. Zajímá nás přitom nejmenší její hodnota, jakožto nejmenší odchylka od testované distribuce, proto ji znázorníme tučně.

Pro většinu otázek je uvedená odchylka nejmenší pro lognormální rozdělení. U otázky CM8\_DUVERA je nejmenší u Weibullova rozdělení. Statistika  $\phi$  má

Rozdělení	Normální	Exponenciální	Lognormální	Gama	Weibullovo
Otázka A6	2.31	0.64	<b>0.21</b>	0.54	0.43
Otázka B9	0.92	0.31	<b>0.21</b>	0.34	0.31
Otázka B13	1.85	0.52	<b>0.07</b>	0.37	0.26
Otázka B18	1.62	0.39	<b>0.18</b>	0.41	0.30
Otázka B27	2.09	0.85	<b>0.44</b>	0.61	0.46
Otázka C15	2.12	0.40	<b>0.17</b>	0.40	0.36
Otázka C17	2.88	0.57	<b>0.42</b>	0.55	0.49
Otázka C21	1.51	0.49	<b>0.19</b>	0.58	0.35
Otázka CM4	1.12	0.28	<b>0.23</b>	0.65	0.40
Otázka CM8	2.01	0.31	0.16	0.32	<b>0.06</b>

Tabulka 13: Výsledky  $\phi$  statistiky.

sice pro tato rozdělení nejmenší hodnotu, avšak v porovnání s ostatními rozděleními, není rozdíl nikterak markantní, snad jen s výjimkou normálního rozdělení, což bylo znatelné i z hodnot testovací statistiky  $Z$ , která byla vždy v řádech desetitisíců. Můžeme tedy právě lognormální rozdělení označit jako nejvhodnější pro modelování časové zátěže jednotlivých otázek v šetření SILC.

## 5.4 Odhady parametrů lognormálního rozdělení

Závěrem využijeme zjištěný poznatek, že všechny z testovaných otázek šetření se dají velmi dobře aproximovat lognormálním rozdělením, a sestavíme tabulku 14, v níž uvedeme odhady parametrů lognormálního rozdělení  $\hat{\mu}$  a  $\hat{\sigma}$ , a také odhady základních číselných charakteristik - střední hodnotu ( $E$ ), rozptyl ( $var$ ), směrodatnou odchylku ( $sd$ ), modus ( $mode$ ), medián ( $med$ ), šikmost ( $skew$ ) a špičatost ( $kurt$ ). Poté vykreslíme grafy hustot pro jednotlivé skupiny otázek - alternativní otázky, výběrové otázky, vypisovací otázky (na číslici), otázky na částku, otevřené otázky, škálové otázky - a krátce okomentujeme. V každé skupině proti sobě znázorníme objektivní i subjektivní proměnnou, kromě posledních dvou skupin, které k sobě zmiňovaný „protějšek“ nemají. Objektivní proměnná bude v obrázku znázorněna vždy vpravo.

Připomeňme si, že pro lognormální rozdělení s parametry  $\mu$  a  $\sigma$  vypadají

uvedené číselné charakteristiky následovně.

Střední hodnota

$$E = e^\mu + \frac{\sigma^2}{2}$$

je hodnota, kolem které se nejčastěji koncentrují realizace náhodné veličiny, a tedy i hodnota, kde lze realizace s největší pravděpodobností očekávat.

Rozptyl

$$var = (e^{\sigma^2} - 1) * (e^{2*\mu + \sigma^2})$$

udává variabilitu realizací znaků kolem střední hodnoty. Vypočteme-li odmocninu z rozptylu, získáme směrodatnou odchylku

$$sd = \sqrt{var}.$$

Modus

$$mode = e^{\mu - \sigma^2}$$

lze pro spojitou náhodnou veličinu definovat jako bod, v němž dosahuje funkce hustoty lokálního maxima. Z toho je zřejmé, že modus nemusí existovat.

Medián

$$med = e^\mu$$

je  $\alpha$ -kvantilem náhodné veličiny pro hodnotu  $\alpha = 0.5$ , a rozděluje tak soubor hodnot na dvě poloviny, přičemž 50% hodnot se nachází před hodnotou mediánu, 50% hodnot za ním.

Charakteristiky šikmosti a špičatosti se často interpretují ve srovnání s normálním rozdělením. Šikmost určuje, je-li rozdělení symetrické (pro  $skew = 0$ ) nebo šikmé. Protáhlejší rozdělení vpravo je pro  $skew > 0$ , vlevo pro  $skew < 0$ . Kladná špičatost pro symetrické rozdělení znamená, že hustota náhodné veličiny je na svých koncích větší než hustota normálního rozdělení, pro zápornou špičatost je menší. Lze však použít i pro nesymetrická rozdělení.

Šikmost a špičatost vypočteme pomocí následujícího vzorce

$$skew = (e^{\sigma^2} + 2) * \sqrt{(e^{\sigma^2} - 1)}, \quad kurt = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3.$$

Dosadíme-li do vzorců za  $\mu$  a  $\sigma$  odhady těchto parametrů  $\hat{\mu}$ ,  $\hat{\sigma}^2$ , získáme odhady číselných charakteristik tohoto rozdělení, které následně dáme do tabulky 14.

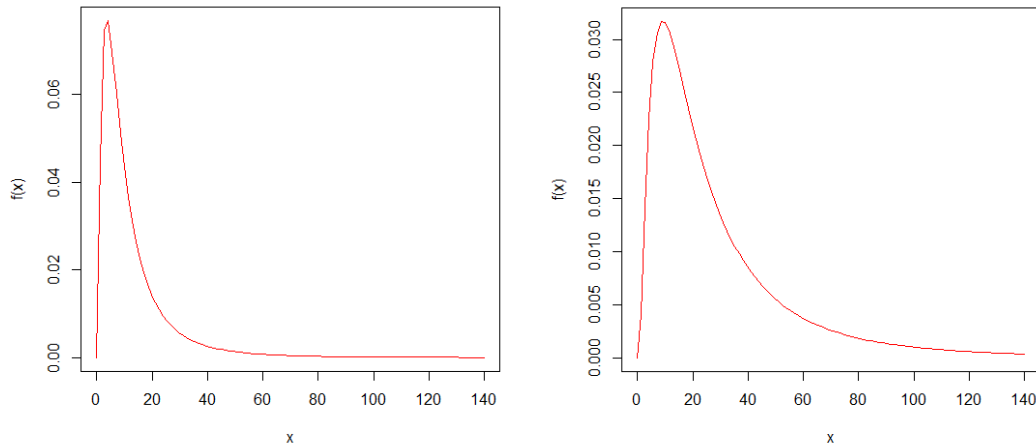
	$\hat{\mu}$	$\hat{\sigma}$	$E$	$var$	$sd$	$mode$	$med$	$skew$	$kurt$
B27	2.14	0.92	12.95	219.91	14.83	3.68	8.52	4.94	66.62
B18	3.04	0.91	31.68	1309.48	36.19	9.05	20.87	4.92	65.66
C17	1.64	0.78	7.02	41.87	6.47	2.80	5.17	3.55	31.56
CM4	3.93	0.68	64.11	2371.29	48.70	32.38	51.05	2.72	18.49
A6	1.33	1.30	8.82	342.69	18.52	0.70	3.79	15.55	1255.05
C21	2.33	0.89	15.18	275.52	16.60	4.67	10.25	4.59	55.84
B9	2.61	0.71	17.37	195.73	13.99	8.20	13.53	2.94	21.51
B13	2.14	1.40	22.62	3105.32	55.73	1.20	8.51	22.34	3348.76
C15	3.03	0.94	32.36	1493.61	38.65	8.56	20.77	5.29	77.92
CM8	3.11	0.88	32.79	1238.22	35.19	10.39	22.35	4.46	52.27

Tabulka 14: Odhady parametrů a základních číselných charakteristik lognormálního rozdělení v sekundách (s).

Z tabulky 14 je patrné, že rozdělení pravděpodobnosti všech otázek bude protáhlé napravo, z čehož lze usuzovat, že hodnoty jednotlivých rozdělení jsou na pravé straně asymetricky rozloženy. Stejně tak lze předpokládat, že rozdělení pravděpodobnosti všech otázek budou špičatější než-li v případě normálního rozdělení se stejnými parametry.

Na následujícím obrázku 28 vidíme dva grafy hustoty rozdělení pro objektivní a subjektivní alternativní proměnnou B27\_PRIJ\_ZPRON a B18\_SLUZ\_VYR. Je patrné, že rozdělení pravděpodobnosti pro objektivní i subjektivní otázku je asymetrické a zešikmené doprava. U objektivní otázky B27 pozorujeme hustotu rozdělení mírně špičatější, u dalších rozdělení je pak rozdíl ve špičatosti patrnější. To je koneckonců vidět i z hodnot vypočtených číselných charakteristik. Modus i medián jsou v případě objektivní otázky B27 menší. Modus je 3,68 sekund v porovnání s 9 vteřinami u subjektivní proměnné, medián 8,52 sekund proti téměř 21 sekundám. Vezmeme-li do úvahy i střední hodnotu, vidíme podobný charakter, a sice, že v případě objektivní otázky je k odpovědi zapotřebí menší časový limit, než u subjektivní proměnné.

Naopak s rostoucí špičatostí klesá rozptyl. Ten je u otázky B27 cca 220 sekund, u subjektivní otázky B18 přes 1 300 sekund. Rozptyl realizací se tak téměř zešestinasobil.

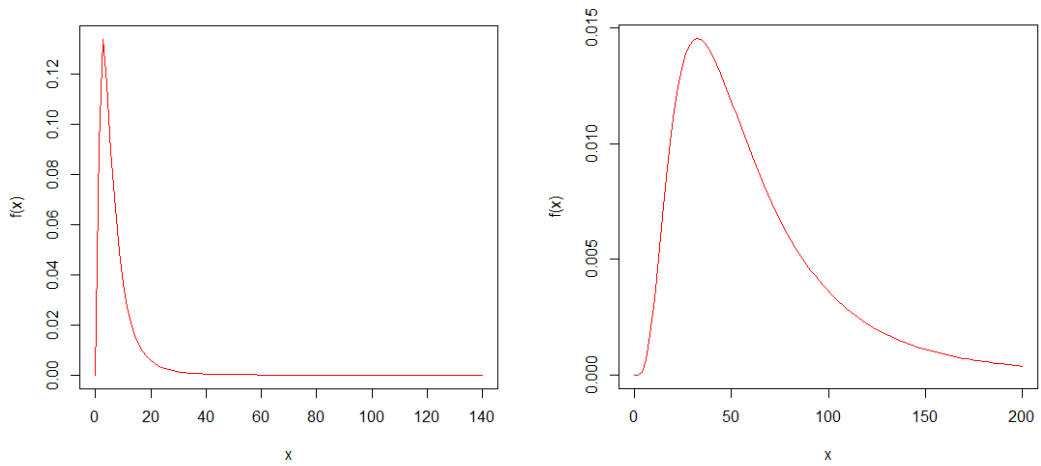


Obrázek 28: Alternativní otázky B27 a B18.

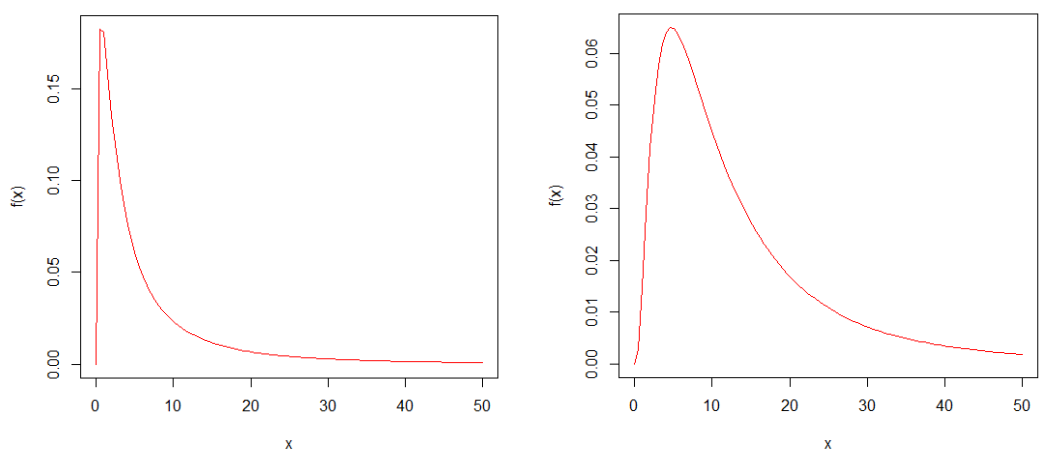
Podobný charakter jako v případě alternativních otázek lze sledovat i u dalších skupin proměnných viz obrázek 29 pro výběrové otázky, viz obrázek 30 pro vypisovací proměnné. Z toho lze usuzovat, že objektivní otázky dají respondentům obecně méně času, oproti tomu u subjektivní otázky záleží na daném jedinci, proto má otázka větší rozptyl (a tím menší špičatost).

U otázek na částku viz obrázek 31 se charakter naopak prohodí. Otázka B9\_VYSE\_HYPO se tak jeví být více problematickou v porovnání se subjektivní B13\_ODH\_CEN. Je tomu tak možná právě proto, že otázka B9, byť je objektivní, dotýká se velice citlivých údajů, a sice výše úvěru k financování bydlení.

Na obrázku 32 uvádíme také hustoty rozdělení pro otázky C15 a CM8. Tyto proměnné k sobě subjektivní nebo naopak objektivní protějšky nemají, proto podobnou dedukci provést nemůžeme.

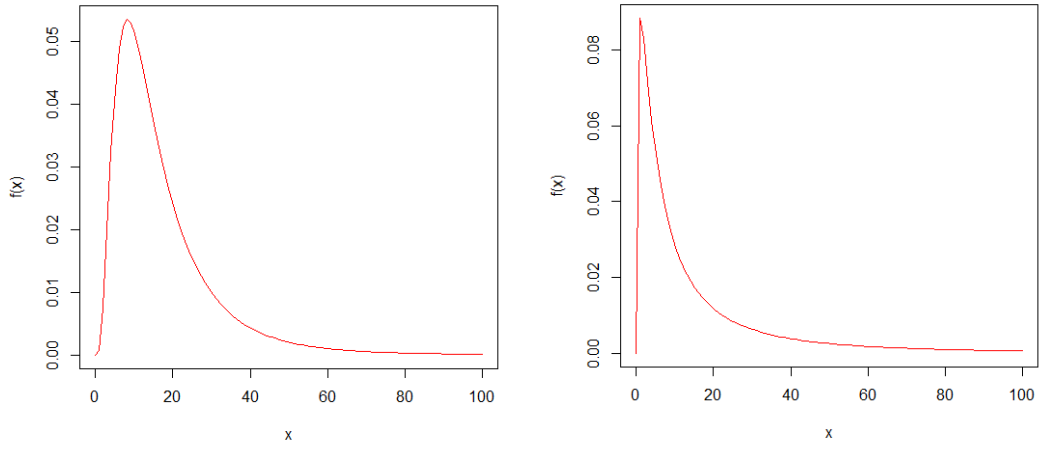


Obrázek 29: Výběrové otázky C17 a CM4.

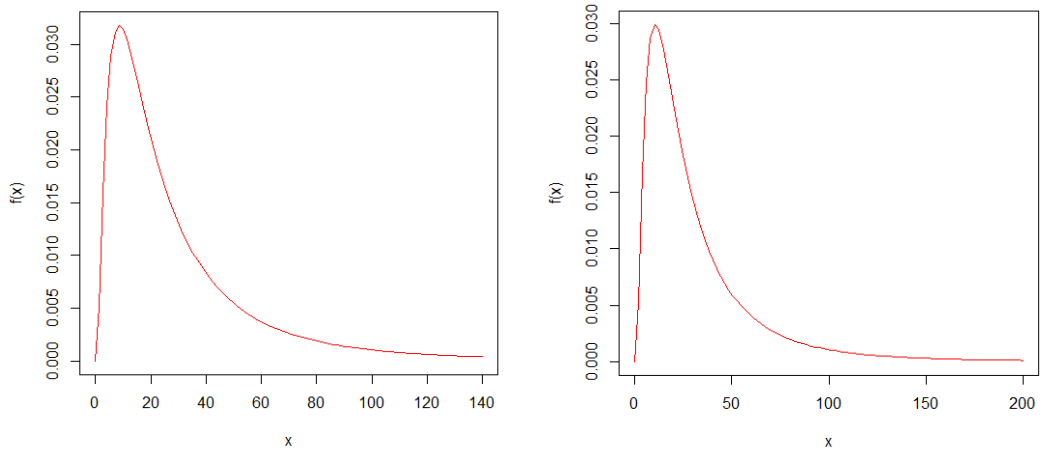


Obrázek 30: Vypisovací otázky A6 a C21.





Obrázek 31: Otázky na částku B9 a B13.



Obrázek 32: Otevřená otázka C15 (vlevo) a škálová otázka CM8 (vpravo).

## Závěr

Diplomová práce se svým obsahem opírá o dotazníkového šetření Životní podmínky 2013, které na našem území každoročně provádí Český statistický úřad s cílem vyšetřit životní úroveň v České republice. Šetření zajišťují tazatelé prostřednictvím čtyř druhů dotazníků, které obsahují přes sto různě tematicky zaměřených otázek, které jsou cíleně sestaveny tak, aby podávaly komplexní a hlavně skutečný obraz o situaci v České republice. Vzhledem k tomu, že vymezený časový rámec šedesáti minut je pro vyšetření jedné domácnosti v praxi nedostatečný, vznikl podnět pro vypracování této diplomové práce.

S cílem otestovat problematickou proměnnou v šetření, tedy časovou zátěž otázek, a určit skutečné pravděpodobnostní rozdělení času, mi byly poskytnuty četné konzultace na ČSÚ, které měly za cíl proniknout do pracovních postupů tazatelů v praxi, pochopit, jak takové šetření v domácnosti probíhá, na jaké překážky tazatelé při své práci naráží a co všechno časovou zátěž ovlivňuje. Současně mi byly předloženy materiály, které alespoň něco málo z těchto informací obsahují v textové podobě, vzorové dotazníky, které se na dlouhou dobu staly terčem mé pozornosti, a reálná data, na nichž jsem měla časovou zátěž otestovat.

K určení skutečného rozdělení pravděpodobnosti časové zátěže jsme použili testy dobré shody, které pro testování využívají Pearsonovu statistiku  $Z$ . Abychom mohli stanovit hodnotu testovacího kritéria, a tak rozhodnout o výsledku testu, bylo zapotřebí nalézt vhodné intervalového rozdělení časové zátěže vybraných otázek. Uvedeným testováním jsme došli k závěru, že k modelování časové zátěže jednotlivých otázek v šetření SILC můžeme použít lognormální rozdělení, což je vzhledem k dobrým vlastnostem tohoto rozdělení určitě příjemné zjištění. Tento výstup jsme následně použili k odhadu parametrů a výpočtu základních charakteristik.

Za účelem zpracování práce jsem se naučila pracovat s typografickým programem  $\text{T}_{\text{E}}\text{X}$ , který mi pomohl práci zpracovat po vizuální stránce, a se statistickým programem R, který mi ušetřil řadu těžkostí při matematických a statistických výpočtech s rozsáhlým souborem dat.

Nejtěžší a asi nejvíce časově náročné v teoretické části práce bylo projít poskytnuté dotazníky se všemi sta proměnnými a sestavit kritéria, na základě nichž by byl vybrán adekvátní vzorek proměnných, se kterým budu moci v průběhu práce pracovat a který by vhodně reprezentoval a zastupoval všechny proměnné v šetření. V praktické části práce zase bylo nejhorší sestavit vhodně intervalová rozdělení pro jednotlivé proměnné k tomu, abychom mohli použít testy dobré shody. Nicméně každá těžkost je zároveň zkušenost, a v tom všem mne tak práce obohatila. Dala mi možnost okusit, jak chutná práce s reálnými daty a jak lze spojit teoretické poznatky s praktickým využitím.

V práci by bylo zajímavé dále sledovat, k jakým výsledkům by vedlo testování širšího spektra teoretických rozdělení. Případně by stálo za to navázat zkoumáním rozdělení časové zátěže pro další šetření v rámci ČSÚ.

## Literatura

- [1] Anděl, J.: Matematická statistika, SNTL/ALFA, Praha, 1978.
- [2] Anděl, J.: Základy matematické statistiky. MATFYZPRESS vydavatelství Matematicko-fyzikální fakulty Univerzity Karlovy v Praze, Praha, 2005.
- [3] Bakytová, H. a kol.: Základy statistiky. Vydavatelstvo technickej a ekonomickej literatúry, Bratislava, 1975.
- [4] Disman M.: Jak se vyrábí sociologická znalost: příručka pro uživatele. Nakladatelství Karolinum, Praha, 2002.
- [5] Chráska, M.: Metody pedagogického výzkumu: základy kvantitativního výzkumu. Grada, Praha, 2007.
- [6] Kunderová, P: Základy pravděpodobnosti a matematické statistiky. UP, Olomouc, 2004.
- [7] Pokyny pro tazatele, SILC 2013. [Interní zdroj ČSÚ.]
- [8] Blaise-survey software for professionals [online]. Dostupné z: [www.blaise.com](http://www.blaise.com),  
[citováno 11. 2. 2014]
- [9] Characteristic of the Weibull Distribution [online]. Dostupné z: <http://www.weibull.com/hotwire/issue14/relbasics14.htm>,  
[citováno 8. 11. 2014]
- [10] Computer assisted personal interviewing [online]. Dostupné z: [http://www.unescap.org/stat/pop-it/pop-guide/capture\\_ch03.pdf](http://www.unescap.org/stat/pop-it/pop-guide/capture_ch03.pdf),  
[citováno 11. 2. 2014]
- [11] Estimation methods [online]. Dostupné z: [http://www.statlect.com/estimation\\_methods.htm](http://www.statlect.com/estimation_methods.htm),  
[citováno 30. 11. 2014]
- [12] EU-SILC a jeho metodologická úskalí: mezinárodní srovnatelnost a příjmové proměnné [online]. Dostupné z: [http://dav.soc.cas.cz/uploads/f92a71a76894abe16b748d2c698d557e929c5216\\_DaVp147-170%20Mysikova.pdf](http://dav.soc.cas.cz/uploads/f92a71a76894abe16b748d2c698d557e929c5216_DaVp147-170%20Mysikova.pdf)  
[citováno 1.8.2014]

- [13] Fitting distributions with R [online]. Dostupné z:  
<http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf>,  
[citováno 21.1. 2015]
- [14] Gallery of distributions [online]. Dostupné z:  
<http://www.itl.nist.gov/div898/handbook/eda/section3/eda366.htm>,  
[citováno 16. 11. 2014]
- [15] Maximálně věrohodný odhad [online]. Dostupné z:  
<http://home.zcu.cz/friessl/hpsb/ml.html>,  
[citováno 21. 11. 2014]
- [16] Maximum likelihood estimation [online]. Dostupné z:  
<http://www.itl.nist.gov/div898/handbook/apr/section4/apr412.htm>,  
[citováno 16. 11. 2014]
- [17] Některá rozdělení [online]. Dostupné z:  
<http://home.zcu.cz/friessl/hpsb/Irozd.html>,  
[citováno 14. 3. 2014]
- [18] Probability distributions [online]. Dostupné z:  
<http://www.statlect.com/distri.htm>,  
[citováno 14. 3. 2014]
- [19] The exponential distribution [online]. Dostupné z:  
<http://www.mast.queensu.ca/stat455/lecturenotes/set4.pdf>  
<http://www.bls.gov/osmr/pdf/st080180.pdf>,  
[citováno 11. 2. 2014]
- [20] Životní podmínky EU-SILC [online]. Dostupné z:  
[http://www.czso.cz/csu/redakce.nsf/i/zivotni\\_podminky\\_eu\\_silc/\\$File/zpc\\_silc.pdf](http://www.czso.cz/csu/redakce.nsf/i/zivotni_podminky_eu_silc/$File/zpc_silc.pdf),  
[citováno 4. 9. 2013]

## **Přílohy**

**Příloha č. 1:** Životní podmínky 2013: Dotazník A

**Příloha č. 2:** Životní podmínky 2013: Dotazník B

**Příloha č. 3:** Životní podmínky 2013: Dotazník C

**Příloha č. 4:** Modul 2013

# **Příloha č. 1**

Účast v šetření je dobrovolná. Zjištěné individuální údaje jsou chráněny podle zákona č. 89/1995 Sb., o státní statistické službě, ve znění pozdějších předpisů.

T

Tazatel zaznamená datum návštěvy a celkový čas zahájení a ukončení rozhovoru nad všemi dotazníky.	den <input type="text"/> <input type="text"/> měsíc <input type="text"/> <input type="text"/>
zahájení <input type="text"/> <input type="text"/> : <input type="text"/> <input type="text"/>	ukončení <input type="text"/> <input type="text"/> : <input type="text"/> <input type="text"/>

Adresa nešetřena - administrativní odpad (pouze 1. vlna)

### A1 Údaje o vyšetření bytu/domácnosti

	HD1	HD2	HD3	HD4
1. Domácnost vyšetřena	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Domácnost nevyšetřena</b>				
2. odmítnutí šetření (neochota sdělovat informace)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. domácnost nezastižena, dočasně nepřítomná	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. domácnost neschopna účasti (zdrav. důvody, vysoký věk)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. jiné důvody (jazyková bariéra)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
šetření nepřichází v úvahu (pouze 2. až 4. vlna)				
6. - celá HD odstěhovaná	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. - celá HD v kolektivní domácnosti nebo instituci	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. - celá HD v zahraničí	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. - žádný člen HD již nežije	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. - HD již neobsahuje panelovou osobu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. - HD sloučena s jinou HD v bytě	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### A2 Změny v domácnosti

(pouze vyšetřené domácnosti ze 2. až 4. vlny)

	HD1	HD2	HD3	HD4
1. některá panelová osoba se z HD odstěhovala	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. domácnost rozdělena v rámci bytu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. jiná změna nebo beze změny	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Poznámky:

A1 Identifikační údaje	1. Území	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	2. Pagina	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	3. Číslo SO	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	4. Vlna šetření	<input type="checkbox"/>
	5. Aktuální počet hospodařících domácností (jen u vyšetřeno bytu)	<input type="checkbox"/>
	6. Číslo tazatele	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
AIK. Kontrolní součet		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Obec:		

Vzory  
číslic

0123456789

Vyznačování



### A3 Druh domu

1. samostatně stojící rodinný dům	<input type="checkbox"/>
2. dvojdoměk, řadový dům	<input type="checkbox"/>
3. bytový dům s méně než 10 byty	<input type="checkbox"/>
4. bytový dům s 10 a více byty	<input type="checkbox"/>
5. jiný	<input type="checkbox"/>

### A4 Počet dotazníků

A byt	<input type="checkbox"/>	B HD	<input type="checkbox"/>	C osoby	<input type="text"/> <input type="text"/>
CM modul	<input type="text"/> <input type="text"/>	za byt celkem		<input type="text"/> <input type="text"/>	

### A5 Osoby zúčastněné na šetření

	Datum	Jméno	Podpis
Tazatel předal			
Kontroloval, vyznačil			

T



## A6 Údaje o osobách

Zapisují se následující osoby:

V domácnostech, které jsou zahrnuty do šetření **prvním rokem**, se zapisují osoby, které ve vybraném bytě obvykle bydlí, dále podnájemníci a hosté, jejichž zamýšlená délka pobytu v dané domácnosti je delší než 6 měsíců, a rovněž osoby dočasně nepřítomné, které však nejsou členy žádné jiné bytové domácnosti, mají jasnou finanční vazbu na vybranou domácnost a jejich nepřítomnost nepřekročí dobu 6 měsíců. Výjimku tvoří osoby studující nebo pracující mimo domov, u nichž nezáleží na délce nepřítomnosti, avšak nesmí mít žádnou jinou soukromou adresu a musí mít úzké finanční vazby na vybranou domácnost.

V domácnostech, které jsou zahrnuty do šetření **opakovaně**, se zapíše všechny osoby z Výpisu osob z předchozí vlny šetření, a to jak současní členové domácnosti, tak i odstěhované a zemřelé osoby. Zapisují se i nově přistěhovaní a narození, kteří jsou členy HD s alespoň jednou panelovou osobou.

**Nezapisují se** osoby, které jsou jen dočasně přítomné, ale mají svou vlastní domácnost jinde (návštěvy, osoby jinde v nájmu atd.), a dále osoby dlouhodobě nepřítomné bez existenčních vazeb na vybranou domácnost, jejichž doba nepřítomnosti je delší než 6 měsíců. Při opakované návštěvě domácnosti se nezapisují ani osoby, které spolu tvoří samostatnou HD složenou pouze z osob mimo panel, tj. členové HD, která neobsahuje žádnou panelovou osobu.

Pořadové číslo osoby	Identifikační číslo		Křestní jméno	Vztah k uživateli bytu, osobě v čele HD		Identifikační číslo (ze sl. 3)		
	HD	osoby				otce	matky	partnera
			Slouží ke snadnější identifikaci osob při rozhovoru. Jako první se zapíše uživatel/ka bytu, pak jeho/její partner/ka, děti a ostatní osoby.					
				BD	HD			
1	2	3	4	5	6	7	8	9
1.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
3.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
4.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
5.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
6.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
7.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
8.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
9.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
10.	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

## A7 Dohledané osoby - původní identifikace osob (z předchozí vlny šetření)

území	pagina	IČ osoby	
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

## Vztah k uživateli bytu, osobě v čele HD

- 1 uživatel/ka bytu, osoba v čele HD
- 2 manželka, družka (ve sl. 5 též manžel, druh)
- 3 syn, dcera
- 4 zeť, snacha
- 5 vnuk, vnučka
- 6 otec, matka, tchán, tchyně
- 7 bratr, sestra
- 8 jiná příbuzná/blízká osoba
- 9 podnájemník, členové jeho HD (pouze sl. 5)

## A8 Údaje o společném hospodaření

Uvedte prosím pořadová čísla osob, které spolu v rámci bytu/domu tvoří samostatně hospodařící domácnosti. Např. 1 + 2, 3 + 4 + 5, 6.

Pořadová čísla osob	HD č. 1	HD č. 2	HD č. 3	HD č. 4	HD č. 5

## A6 Údaje o osobách

Měsíc a rok narození		Pohlaví	Rodinný stav	Rok sňatku	Přítomnost/nepřítomnost v domácnosti (demografický pohyb, stěhování)								Kontrolní součet
měsíc	rok (poslední dvojčíslí)	1 muž 2 žena		Zapiše se dvojčíslí roku posledního sňatku*				Měsíc od	Rok od	Měsíc do	Rok do	Místo odstěhování	Součet číselných hodnot ve sloupcích 2, 3, 5 až 22
					1 panelová osoba 2 osoba mimo panel	Druh přítomnosti (2. až 4. vlna)	1 přítomný 2 dočasně nepřítomný						
10	11	12	13	14	15	16	17	18	19	20	21	22	23

<b>Rodinný/osobní stav</b> 1 svobodný(-á) 2 ženatý, vdaná, registrované partnerství 3 ovdovělý(-á), registrované partnerství zaniklé smrtí 4 rozvedený(-á), registrované partnerství zaniklé rozhodnutím	<b>Druh přítomnosti (pouze 2. až 4. vlna)</b> 1 osoba šetřena v tomto bytě i v minulém roce 2 přistěhovaný(-á) 3 narozený(-á) 4 odstěhovaný(-á) 5 zemřelý(-á) 6 bývalý člen domácnosti přítomný v minulém roce více než 3 měsíce	<b>Místo odstěhování</b> 1 do jiné soukromé domácnosti v ČR 2 do kolektivní domácnosti či instituce 3 do zahraničí 4 nezjištěno
--	--	---

\* U registrovaného partnerství se zapiše dvojčíslí roku poslední registrace

Poznámky:

## A6 Údaje o osobách

Pořadové číslo osoby	Nejvyšší dokončené vzdělání		Dvojití rok dokončení nejvyššího vzdělání	Současné studium		Dotazník C za osobu			Kontrolní součet
	24	25		26	27	28	29	30	
1.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Nejvyšší dokončené vzdělání / Současné studium
0 předškolní děti, neukončený 1. stupeň ZŠ (sl. 24)
1 první stupeň ZŠ
2 druhý stupeň ZŠ
3 vyučen(a), nižší střední
4 úplné střední s maturitou
5 nástavbové studium, pomaturitní kurzy, absolvování dvou nebo více SŠ
6 vyšší odborné (DiS.)
7 vysokoškolské bakalářské
8 vysokoškolské magisterské či inženýrské
9 doktorské (Ph.D., CSc., DrSc.)

Výsledek vyšetření dotazníku C
1 kompletní
2 částečně vyplněný - chybí příjmy
3 částečně vyplněný - příjmy jsou, chybí jiné údaje
4 nevyplněný
5 osoba mladší 16 let
6 osoba již není členem domácnosti

Způsob vyplnění
1 osobně s tazatelem
2 jinou osobou
3 samovyplnění

## **Příloha č. 2**

Při opětovné návštěvě tazatel nejprve dotazem u respondenta ověří, zda se jedná o byt stejné velikosti a uspořádání jako v předchozím roce šetření, tj. bez zásadních změn co do počtu místností, rozlohy bytu nebo jeho příslušenství. Dále se ujistí, že nedošlo ke změně právního vztahu domácnosti k tomuto bytu či zda se nezměnil počet hospodařících domácností v bytě (sloučení/rozdělení domácnosti).

taková změna nastala  ↓ stejný byt jako loni  → **B5**

**Bydlení**

**B1** Kolik obytných místností využívá Vaše domácnost?

počet místností

**B2** Jaká je celková plocha Vašeho bytu?

plocha v m<sup>2</sup>

**B3** Je ve Vašem bytě následující příslušenství?

	ano	ano, s další HD v bytě	ne
1. koupelna, sprchový kout	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. splachovací WC	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**B4** Kdy jste se nastěhovali do tohoto bytu? Pokud došlo ke změně právního vztahu Vaší domácnosti k tomuto bytu, uveďte rok této změny.

rok

**B5** Vnímáte některý z následujících problémů spojených s Vaším bydlením?

	ano	ne
1. zatékání střechou, vlhké zdi, podlahy, základy nebo shnilá okna, rámy, podlahy	<input type="checkbox"/>	<input type="checkbox"/>
2. příliš tmavý byt, nedostatek denního světla	<input type="checkbox"/>	<input type="checkbox"/>
3. příliš malý byt, nedostatek místa	<input type="checkbox"/>	<input type="checkbox"/>
4. hluk od sousedů nebo hluk z ulice (doprava, obchody, továrny atd.)	<input type="checkbox"/>	<input type="checkbox"/>
5. znečištění, špina nebo jiné problémy se životním prostředím v okolí	<input type="checkbox"/>	<input type="checkbox"/>
6. kriminalita, násilí nebo vandalismus v okolí	<input type="checkbox"/>	<input type="checkbox"/>

**B6** Jakou právní formou Vaše domácnost užívá tento byt?

Vlastnictví	1. byt ve vlastním domě	<input type="checkbox"/>	→ <b>B7</b>
	2. byt v osobním vlastnictví	<input type="checkbox"/>	
	3. družstevní byt (SBD)	<input type="checkbox"/>	
Pronájem a jiné formy užívání	4. pronajatý byt (nájem, podnájem)	<input type="checkbox"/>	→ <b>B11</b>
	5. služební, domovní byt	<input type="checkbox"/>	
	6. bydlení u příbuzných/známých apod.	<input type="checkbox"/>	

<b>BI</b> Identifikační údaje	1. Území	<input type="text"/>
	2. Pagina	<input type="text"/>
	3. Číslo HD	<input type="text"/>
	4. IČ osoby, která poskytla informace pro dotazník B	<input type="text"/>
BIK. Kontrolní součet		<input type="text"/>

Vzory číslíc

0123456789

Vyznačování



**Vlastnictví**

**B7** Splácíte nebo spláceli jste v minulém kalendářním roce na tento byt/dům hypotéku nebo úvěr ze stavebního spoření?

1. ano, alespoň jedna splátka byla v roce 2012
2. ano, splácíme až od roku 2013
3. ne  → **B13**

**B8** Můžete mi sdělit celkovou měsíční splátku Vaší hypotéky/úvěru zahrnující splátku jistiny i úroků?

měsíční splátka v Kč

**B9** Kolik činila počáteční výše Vaší hypotéky/úvěru, tj. výše jistiny?

částka v Kč

**B10** Ve kterém roce jste si Vaší hypotéku/úvěr sjednali a na kolik let byla Vaše hypotéka/úvěr sjednána?

rok sjednání

celková doba splácení v letech

→ **B13**

**Pronájem a jiné formy užívání**

**B11** Platíte nájemné, resp. úhradu za užívání bytu (bez energií)?

1. ano, tržní
2. ano, snížené
3. ne

**B12** Kdo Vám poskytl možnost užívat tento byt/část bytu?

1. obec, stát
2. vlastník bytu/domu
3. člen bytového družstva
4. nájemce bytu

**BK1.** Kontrolní součet za otázky B1, B2, B4, B8 až B10

<input type="text"/>
----------------------



### Spotřeba z vlastního hospodářství nebo podniku

B24 Odhadněte prosím množství, resp. hodnotu spotřebovaných výrobků, které Vaše domácnost spotřebovala z vlastního hospodaření nebo podniku, který vlastníte nebo provozujete (nezahnujte spotřebu pro krmení zvířat). Uveďte množství, případně částku za celý minulý kalendářní rok.

1. maso a masné výrobky (kg)	<input type="text"/>	6. dřevo z vlastního lesa (Kč)	<input type="text"/>
2. vejce (ks)	<input type="text"/>	7. ostatní potraviny a nápoje včetně stravného (Kč)	<input type="text"/>
3. brambory (kg)	<input type="text"/>	8. průmyslové výrobky a služby (Kč)	<input type="text"/>
4. ovoce (kg)	<input type="text"/>		
5. zelenina (kg)	<input type="text"/>	nemáme	<input type="checkbox"/>
B24K. Kontrolní součet za tab. B24			<input type="text"/>

### Transfery mezi domácnostmi

B25 Mnoho lidí poskytuje peněžní nebo naturální výpomoc osobám žijícím v jiné domácnosti nebo v nějaké instituci. Uveďte prosím částku příjmů/výdajů, kterou jste Vy nebo někdo jiný z Vaší domácnosti v minulém kalendářním roce dostával/poskytoval pravidelně (opakovaně) a u jednorázových částek uveďte součet za celý rok. Pokuste se také odhadnout hodnotu přijatých, resp. věnovaných darů za celý minulý kalendářní rok.

	transfery nebyly		přijaté v Kč	vydané v Kč
1. výživné (na děti i bývalého partnera)	<input type="checkbox"/>	<input type="checkbox"/> M <input type="checkbox"/> R	<input type="text"/>	<input type="text"/>
2. další opakované peněžní transfery (podpora studentů, blízkých osob v jiných domácnostech)	<input type="checkbox"/>	<input type="checkbox"/> M <input type="checkbox"/> R	<input type="text"/>	<input type="text"/>
3. jednorázové a mimořádné částky	<input type="checkbox"/>	R	<input type="text"/>	<input type="text"/>
4. naturální transfery (přijaté/darované produkty z vlastního hospodářství nebo podniku, bezplatné stravování u příbuzných, přijaté/darované výrobky a služby)	<input type="checkbox"/>	R	<input type="text"/>	<input type="text"/>
B25K. Kontrolní součet za tab. B25			<input type="text"/>	<input type="text"/>

### Dávky státní sociální podpory a sociální péče

B26 Pobírala Vaše domácnost v minulém kalendářním roce některý z těchto sociálních příjmů? Uveďte prosím počet měsíců pobírání a měsíční částku, resp. částku za celý rok.

	nepobírala		počet měsíců	Kč	počet měsíců	Kč
1. přídatky na děti	<input type="checkbox"/>	M	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2. dávky pěstounské péče	<input type="checkbox"/>	M	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
3. příspěvek na bydlení	<input type="checkbox"/>	M	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
4. pomoc v hmotné nouzi	<input type="checkbox"/>	M	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
5. porodné	<input type="checkbox"/>	R		<input type="text"/>		
6. pohřebné	<input type="checkbox"/>	R		<input type="text"/>		
7. jiné sociální dávky	<input type="checkbox"/>	R		<input type="text"/>		
B26K. Kontrolní součet za tab. B26			<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Poznámky:



### Příjmy z pronájmu

**B27** Měl(a) jste Vy nebo někdo jiný z Vaší domácnosti v minulém kalendářním roce příjem z pronájmu nemovitosti (bytu nebo jeho části, domu, nebytových prostor, chaty, pozemku), popř. movitých věcí (auta, strojů atd.)?

1. ano  2. ne  → **B30**

**B28** Můžete prosím uvést, kolik činil příjem z tohoto pronájmu po odečtení nákladů (údržba, opravy, úroky z úvěru, pojištění a jiné poplatky)? Označte, zda uvádíte hrubou nebo čistou částku a zda jde o měsíční nebo roční příjem.

příjem z pronájmu v Kč









1. hrubá částka  měsíční  → **B30**  
 2. čistá částka  roční   
 3. neznám přesnou výši  ↓

**B29** Odhadněte prosím alespoň interval, do kterého by patřil hrubý roční příjem Vaší domácnosti z tohoto pronájmu.

A. méně než 20 000 Kč  D. 100 001 – 200 000   
 B. 20 001 – 50 000  E. 200 001 – 500 000   
 C. 50 001 – 100 000  F. 500 001 a více



### Daně z nemovitostí

**B30** Platil(a) jste Vy nebo někdo jiný z Vaší domácnosti v minulém kalendářním roce daň z nemovitostí?

1. ano  2. ne  → **B33**

**B31** Uvedte prosím částku, kterou jste musel(a) zaplatit.

částka v Kč







→ **B33**

neznám přesnou výši  ↓

**B32** Odhadněte prosím alespoň interval, do kterého by zaplacená částka pravděpodobně patřila.

A. méně než 1 000 Kč  D. 3 001 – 4 000   
 B. 1 001 – 2 000  E. 4 001 – 5 000   
 C. 2 001 – 3 000  F. 5 001 a více

**BK4.** Kontrolní součet za otázky B28 a B31









### Péče o děti do 12 let (narozené v roce 2000 a později)

**B33** Jakým způsobem je ve Vaší domácnosti zajištěna péče o děti do 12 let (kromě péče samotných rodičů, případně pěstounů)? Uvedte prosím, zda je Vaše dítě předškolák či školák a kolik hodin týdně tráví v uvedených zařízeních nebo v péči jiné osoby.

Identifikační číslo dítěte	předškolák	školák	1. předškolní zařízení	2. povinná školní docházka	3. školní družina, dětská centra	4. denní zařízení, stacionáře	5. chůva, au-pair	6. prarodiče, příbuzní a známí	
<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	
<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	
<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	
<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	
<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	
v HD nejsou děti do 12 let		<input type="checkbox"/>	B33K. Kontrolní součet za tab. B33					<input type="text"/>	<input type="text"/>

Poznámky:



Mnohokrát děkujeme za účast na tomto rozhovoru



## **Příloha č. 3**



Tazatel zaznamená datum, čas zahájení a ukončení rozhovoru nad dotazníkem C.	den <input type="text"/> <input type="text"/> měsíc <input type="text"/> <input type="text"/>
zahájení <input type="text"/> <input type="text"/> : <input type="text"/> <input type="text"/>	ukončení <input type="text"/> <input type="text"/> : <input type="text"/> <input type="text"/>
Vzory číslic <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Vyznačování <input checked="" type="checkbox"/>

<b>CI</b> Identifikační údaje	1. Území <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	2. Pagina <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	3. IČ osoby <input type="text"/> <input type="text"/>
	CIK. Kontrolní součet <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>

**Pracovní aktivita**

C1 Uveďte prosím Vaši pracovní aktivitu v jednotlivých měsících minulého kalendářního roku a v současnosti. V případě, že jste v měsíci pracoval(a) na částečný úvazek, vyznačte do hlavní činnosti jeden z kódů 1 až 3 a zaškrtněte políčko v řádku částečný úvazek.

	2012												celý rok 2012	aktuální stav
	leden	únor	březen	duben	květen	červen	červenec	srpen	září	říjen	listopad	prosinec		
	1	2	3	4	5	6	7	8	9	10	11	12		
<b>A. hlavní činnost</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
částečný úvazek	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>souběžné činnosti</b>														
B. zaměstnanec	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C. samostatně činný(-á)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D. žák, učeň, student (denní studium)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>důchody a sociální dávky</b>														
E. starobní důchod	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F. invalidní důchod	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
G. rodičovský příspěvek	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
													C1K. Kontrolní součet za tab. C1 <input type="text"/> <input type="text"/> <input type="text"/>	

Pracovní aktivita - hlavní činnost	
1 zaměstnanec	6 žák, učeň, student (denní studium)
2 samostatně činný(-á)	7 ve starobním důchodu
3 na placené mateřské dovolené	8 v invalidním důchodu
4 pobírající rodičovský příspěvek	9 v domácnosti, péče o děti nebo péče o blízkou osobu
5 nezaměstnaný(-á)	0 ostatní ekonomicky neaktivní

<b>CT1</b> Na základě odpovědí na otázku <b>C1A</b> zaškrtněte odpovídající možnost.	
1. v C1A odpověděl v posledním sloupci kódy 1 až 3	<input type="checkbox"/> → <b>C13</b>
2. v C1A odpověděl v posledním sloupci kódem 5	<input type="checkbox"/> → <b>C2</b>
3. v C1A odpověděl v posledním sloupci kódy 4, 6 až 9 nebo 0	<input type="checkbox"/> → <b>C4</b>



## Nezaměstnaní

<b>C2</b> Jak dlouho jste nezaměstnaný(-á)?	
1. méně než 1 rok <input type="checkbox"/> počet měsíců <input type="text"/>	
2. 1 rok a déle <input type="checkbox"/>	

<b>C3</b> Jste evidován(a) na úřadu práce a pobíráte podporu v nezaměstnanosti?	
1. jsem evidován(a) a pobírám podporu <input type="checkbox"/>	
2. jsem evidován(a) a nepobírám podporu <input type="checkbox"/>	
3. nejsem evidován(a) <input type="checkbox"/>	

## Nepracující

<b>C4</b> Vykonával(a) jste během posledních 7 dní nějakou placenou práci nebo podnikatelskou činnost, i kdyby se jednalo pouze o jednu hodinu?	
1. ano, pravidelnou činnost <input type="checkbox"/>	
2. ano, jednorázovou činnost <input type="checkbox"/>	
3. ne <input type="checkbox"/>	

<b>C5</b> Hledal(a) jste v posledních 4 týdnech práci?	
1. ano <input type="checkbox"/>	
2. ne, vyřizuji založení vlastní firmy <input type="checkbox"/>	→ <b>C8</b>
3. ne, mám již sjednané zaměstnání a nejpozději do 3 měsíců do něj nastoupím <input type="checkbox"/>	
4. ne, z jiných důvodů <input type="checkbox"/>	

<b>C6</b> Jakým způsobem jste v uplynulých 4 týdnech hledal(a) práci?		
	ano	ne
1. nabídka pracovních míst na úřadu práce	<input type="checkbox"/>	<input type="checkbox"/>
2. žádost o místo přímo u zaměstnavatele	<input type="checkbox"/>	<input type="checkbox"/>
3. pomoc zprostředkovatelské agentury	<input type="checkbox"/>	<input type="checkbox"/>
4. dotaz u příbuzných a přátel	<input type="checkbox"/>	<input type="checkbox"/>
5. inzerce na internetu nebo v novinách	<input type="checkbox"/>	<input type="checkbox"/>
6. účast na pohovoru, zkouškách či testu	<input type="checkbox"/>	<input type="checkbox"/>
7. čekání na výsledek žádosti o místo nebo na výsledek konkurzu	<input type="checkbox"/>	<input type="checkbox"/>
8. jiný aktivní způsob hledání	<input type="checkbox"/>	<input type="checkbox"/>

<b>C7</b> Jste připraven(a) nastoupit do zaměstnání během 2 týdnů?	
1. ano <input type="checkbox"/>	2. ne <input type="checkbox"/>

## Bývalé hlavní zaměstnání

<b>C8</b> Byl(a) jste už někdy zaměstnán(a) nebo jste podnikal(a)? (Mělo by se jednat o <u>pravidelnou činnost</u> trvající alespoň 6 měsíců.)	
1. ano <input type="checkbox"/>	
2. pracuji ve svém prvním zaměstnání <input type="checkbox"/>	→ <b>C13</b>
3. ne <input type="checkbox"/>	→ <b>C25</b>

<b>C9</b> Jaké bylo Vaše poslední ukončené hlavní zaměstnání (profese)? Popište prosím co nejpodrobněji práci, kterou jste vykonával(a).	
→ dvoumístný kód CZ-ISCO: <input type="text"/>	

<b>C10</b> Jaké postavení jste měl(a) ve svém posledním hlavním zaměstnání?	
1. zaměstnanec <input type="checkbox"/>	→ <b>C13</b>
2. společník, jednatel s.r.o. <input type="checkbox"/>	
3. osoba samostatně výdělečně činná se zaměstnanci <input type="checkbox"/>	
4. osoba samostatně výdělečně činná bez zaměstnanců <input type="checkbox"/>	
5. pomáhající rodinný příslušník <input type="checkbox"/>	

<b>C11</b> Jaký typ pracovní smlouvy jste měl(a) ve svém posledním hlavním zaměstnání?	
1. smlouva na dobu neurčitou <input type="checkbox"/>	
2. smlouva na dobu určitou <input type="checkbox"/>	
3. dohoda o pracovní činnosti nebo o provedení práce <input type="checkbox"/>	
4. práce bez smlouvy <input type="checkbox"/>	

<b>C12</b> Měl(a) jste ve svém posledním hlavním zaměstnání nějaké podřízené zaměstnance?	
1. ano <input type="checkbox"/>	2. ne <input type="checkbox"/>

<b>C13</b> Kdy jste nastoupil(a) do svého prvního řádného zaměstnání, popř. začal(a) podnikat?	
první zaměstnání v roce	<input type="text"/>

<b>C14</b> Kolik let jste od té doby odpracoval(a)?	
počet roků	<input type="text"/>

TAZATEL: Pokud respondent odpověděl v otázce **C4** kódem 2 (ano, jednorázovou činnost) nebo kódem 3 (ne), přejděte na otázku **C24**, jinak pokračujte následující otázkou.

<b>CK2.</b> Kontrolní součet za otázky C2, C9, C13 a C14	
<input type="text"/>	

**Současné hlavní zaměstnání**

C15 Jaké je Vaše současné hlavní zaměstnání (profese)?  
Popište prosím co nejpodrobněji práci, kterou vykonáváte.

→ dvoumístný kód CZ-ISCO:

C16 Popište hlavní činnost místní jednotky firmy  
nebo organizace, kde pracujete.

→ dvoumístný kód CZ-NACE:

C17 Jaké postavení máte ve svém současném hlavním  
zaměstnání?

- |  |                          |       |
|--|--------------------------|-------|
| 1. zaměstnanec   | <input type="checkbox"/> | → C20 |
| 2. společník, jednatel s.r.o.                          | <input type="checkbox"/> |       |
| 3. osoba samostatně výdělečně činná<br>se zaměstnanci  | <input type="checkbox"/> |       |
| 4. osoba samostatně výdělečně činná<br>bez zaměstnanců | <input type="checkbox"/> |       |
| 5. pomáhající rodinný příslušník                       | <input type="checkbox"/> |       |

C18 Jaký typ pracovní smlouvy máte ve svém zaměstnání?

- |   |                          |
|---|--------------------------|
| 1. smlouva na dobu neurčitou                            | <input type="checkbox"/> |
| 2. smlouva na dobu určitou                              | <input type="checkbox"/> |
| 3. dohoda o pracovní činnosti nebo<br>o provedení práce | <input type="checkbox"/> |
| 4. práce bez smlouvy                                    | <input type="checkbox"/> |

C19 Máte ve svém současném zaměstnání nějaké podřízené  
zaměstnance?

- |                                 |                                |
|---------------------------------|--------------------------------|
| 1. ano <input type="checkbox"/> | 2. ne <input type="checkbox"/> |
|---------------------------------|--------------------------------|

C20 Kolik lidí má místní jednotka firmy nebo organizace, v níž  
pracujete? (Do počtu zahrňte také sebe.)

- |   |
|---|
| 1. 10 a méně osob: přesně <input type="text"/> <input type="text"/> nevíم přesně <input type="checkbox"/> |
| 2. 11 až 19 osob <input type="checkbox"/>   |
| 3. 20 až 49 osob <input type="checkbox"/>   |
| 4. 50 a více osob <input type="checkbox"/>  |

CI Identifikační údaje	1. Území	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	2. Pagina	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	3. IČ osoby	<input type="text"/> <input type="text"/>
	CIK. Kontrolní součet	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>

C21 Kolik hodin odpracujete ve svém zaměstnání/podnikání  
nebo při práci pro rodinnou firmu? Uveďte prosím obvyklý  
počet hodin týdně. (U zaměstnanců se zahrne  
i neplacená práce přesčas.)

hlavní zaměstnání

další zaměstnání (přichází-li v úvahu)

**TAZATEL:** Pokud je součet odpracovaných hodin v předchozí  
otázce C21 menší než 30 hodin, pokračujte následující otázkou  
C22, jinak přejděte na otázku C23.

C22 Uveďte důvod, proč pracujete méně než 30 hodin týdně.

- |  |                          |
|--|--------------------------|
| 1. vlastní nemoc nebo invalidita   | <input type="checkbox"/> |
| 2. chtěl(a) bych pracovat více hodin,<br>ale nemohu najít práci na plný úvazek | <input type="checkbox"/> |
| 3. studium nebo proškolení   | <input type="checkbox"/> |
| 4. nechci pracovat více hodin  | <input type="checkbox"/> |
| 5. péče o děti nebo jiné osoby   | <input type="checkbox"/> |
| 6. počet hodin za všechna moje zaměstnání<br>je považován za plný úvazek       | <input type="checkbox"/> |
| 7. jiné důvody   | <input type="checkbox"/> |

C23 Ukončil(a) jste v období od 1. 1. 2012 do současnosti  
své hlavní zaměstnání a začal(a) jste pracovat v jiném?  
Pokud ano, uveďte měsíc a rok ukončení předchozího  
zaměstnání. (V případě více změn uveďte poslední.)

1. ano

měsíc   rok

2. ne  → C25

CK3. Kontrolní součet za otázky C15, C16, C20, C21 a C23

TAZATEL: Na otázku **C24** odpovídá také respondent, který pracoval (C8 kód 1), ale nemá současné hlavní zaměstnání (C15 až C21 není vyplněné).

**C24** Proč jste změnil(a) své předchozí zaměstnání, resp. ukončil(a) své poslední zaměstnání?

1. získání nebo hledání lepšího místa	<input type="checkbox"/>	5. potřeba pečovat o děti nebo o blízkou osobu	<input type="checkbox"/>
2. konec zaměstnání na dobu určitou	<input type="checkbox"/>	6. přestěhování kvůli zaměstnání partnera nebo z důvodu sňatku	<input type="checkbox"/>
3. donucen(a) k odchodu zaměstnavatelem (propuštění z práce, uzavření firmy atd.)	<input type="checkbox"/>	7. vlastní nemoc nebo invalidita	<input type="checkbox"/>
4. prodej nebo uzavření vlastní či rodinné firmy	<input type="checkbox"/>	8. všechny ostatní důvody (odchod do starobního nebo předčasného důchodu, změna bydliště, další studium, rozhodnutí žít z úspor atd.)	<input type="checkbox"/>

## PŘÍJMY ZA ROK 2012

### Příjmy ze zaměstnání

**C25** Měl(a) jste v minulém kalendářním roce příjem ze zaměstnání? Vedle příjmu z hlavního pracovního poměru uveďte případný příjem z dalších pracovních poměrů, prací na dohody a ostatních jednorázových a příležitostných prací.

1. ano  2. ne  → **C32**

**C26** Uveďte prosím, kolik činil v minulém kalendářním roce Váš hrubý, resp. čistý příjem ze zaměstnání (dle uvedených položek). Můžete uvést buď pravidelný měsíční příjem (průměrný) nebo celkový roční příjem zahrnující i veškeré příplatky a mimořádné příjmy.

			počet měsíců	Kč	počet měsíců	Kč
hlavní pracovní poměr	hrubý <input type="checkbox"/>	M	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	čistý <input type="checkbox"/>	R	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
další pracovní poměr	hrubý <input type="checkbox"/>	M	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	čistý <input type="checkbox"/>	R	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
dohoda(-y) o provedení práce	hrubý	R	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
dohoda(-y) o pracovní činnosti	hrubý	R	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

**C27** Obdržel(a) jste v minulém kalendářním roce některou z následujících položek jako příjem navíc?  
- náhrada za přesčasy; 13. a 14. plat; mimořádné odměny, prémie; podíly na výsledku hospodaření firmy, bonusy; odstupné; spropitné; jiné platby (ošatné, diety, provize atd.)

1. ano   
2. ne  → **C29**

### Požítky od zaměstnavatele

**C29** Poskytoval Vám Váš zaměstnavatel v minulém kalendářním roce automobil, dodávku či jiné motorové vozidlo, které jste mohl(a) využívat i pro soukromé účely? Pokud ano, uveďte kolik měsíců jste jej využíval(a).

1. ano, využíval(a) jsem  počet měsíců   
2. neposkytoval / nevyužíval(a) jsem

**C28** Pokud jste nezahmul(a) některou z těchto položek do výše uvedených příjmů, uveďte prosím hrubou, resp. čistou částku těchto dodatečných plateb za celý minulý kalendářní rok.

hrubá   
 čistá částka v Kč

všechny platby jsou již zahrnuty

CK4. Kontrolní součet za otázky C23, C26, C28 a C29

C30 Poskytoval Vám Váš zaměstnavatel v minulém kalendářním roce příspěvky na stravování? Pokud ano, uveďte prosím počet měsíců, po které jste tyto příspěvky pobíral(a), jejich počet za měsíc a dále jejich hodnotu a cenu, za jakou jste je koupil(a).

1. **závodní stravování**  2. **stravenky**

počet měsíců   ks/měsíc

hodnota jídla / hodnota stravenky    za cenu

3. **neposkytoval / nevyužíval(a) jsem**

C31 Poskytoval Vám Váš zaměstnavatel některé další výhody a nepeněžní služby, ať už zdarma nebo za částečnou úhradu? Odhadněte prosím **roční výši** (v Kč) těch, které jste v minulém kalendářním roce využíval(a).

1. mobilní nebo pevný telefon

2. jazykové kurzy

3. příspěvek na benzín, dopravu

4. slevu na firemní zboží nebo služby

5. příspěvek na sportovní vyžití, dovolenou

6. příspěvek na penzijní nebo životní pojištění

**Příjmy z podnikání a jiné samostatně výdělečné činnosti**

C32 Měl(a) jste v minulém kalendářním roce nějaké příjmy z podnikání nebo jiné samostatně výdělečné činnosti? Pokud máte příjem jako spolupracující osoba, uveďte prosím částku rozepsanou na Vaši osobu jako podíl z úhrnné společné částky.

1. ano  3. ne  → C36

2. ano, jako spolupracující osoba  částka v Kč      → C36

C33 Jaký byl Váš zisk/ztráta z podnikání (tj. rozdíl mezi příjmy a výdaji) v minulém kalendářním roce? Vyberte jednu z následujících možností pro vyčíslení Vašich příjmů z podnikání.

1. **daňové přiznání** - hrubý zisk/ztráta (základ daně) v Kč

2. **vlastní odhad** hrubý

čistý

C34 Účastnil(a) jste se v minulém kalendářním roce některého z uvedených pojištění? Uveďte prosím výši Vašich záloh na sociální (důchodové a nemocenské) a zdravotní pojištění a počet měsíců jejich placení v minulém kalendářním roce.

	neúčastnil(a)	počet měsíců	minimální možné zálohy		výše měsíční zálohy v Kč
			ano	ne	
1. důchodové pojištění	<input type="checkbox"/>	<input type="text"/> <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/> →	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
2. nemocenské pojištění	<input type="checkbox"/>	<input type="text"/> <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/> →	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
3. zdravotní pojištění	<input type="checkbox"/>	<input type="text"/> <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/> →	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>

C35 Využíval(a) jste opakovaně, případně použil(a) jste jednorázově peněžní prostředky z příjmů z Vašeho podnikání na soukromé účely? Pokud ano, v případě opakovaných výběrů uveďte prosím počet měsíců a pravidelné měsíční částky, v případě jednorázových uveďte jejich roční úhrn.

	výběry nebyly	počet měsíců	částka v Kč
1. opakované	<input type="checkbox"/>	<input type="text"/> <input type="text"/>	měsíční <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
2. jednorázové	<input type="checkbox"/>		roční <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>

**Další příjmy z pracovní činnosti**

C36 Měl(a) jste v minulém kalendářním roce příjmy za příspěvky do novin, časopisů, rozhlasu nebo televize jako drobné autorské honoráře nepřesahující 7 000 Kč měsíčně od téhož plátce? Pokud ano, uveďte prosím jejich roční úhrn.

1. ano  2. ne

celková roční částka v Kč

C37 Měl(a) jste v minulém kalendářním roce nějaké další příjmy z pracovní činnosti?

1. ano  2. ne

celková roční částka v Kč

CK5. Kontrolní součet za otázky C30 až C37

## Ostatní příjmy

C38 Měl(a) jste v minulém kalendářním roce některé z následujících příjmů, ať už pravidelných nebo jednorázových? Uvedte prosím vždy čistý celoroční příjem.

	neměl(a)	roční částka v Kč
1. <b>příjmy z kapitálového majetku</b> úroky z účtů v bance, z vkladů a vkladových listů; podíly na zisku kapitálových společností a družstev, dividendy z akcií; výnosy z cenných papírů, dluhopisy, obligace; příjmy ze zdrojů v zahraničí	<input type="checkbox"/>	<input type="text"/>
2. <b>příjmy z prodeje</b> příležitostná domácí samovýroba; prodej zemědělských výrobků jako přebytků z osobního hospodářství	<input type="checkbox"/>	<input type="text"/>
3. <b>příjmy ze životního pojištění</b> pojistné plnění pro případ dožití určitého věku nebo v případě úmrtí jiné osoby	<input type="checkbox"/>	<input type="text"/>
4. <b>příjmy z neživotního pojištění</b> plnění z jiných druhů pojištění (úrazové, nemovitosti, domácnosti, motor. vozidel atd.)	<input type="checkbox"/>	<input type="text"/>
5. <b>stipendia, kapesné učňů</b> prospěchové, sociální i ubytovací stipendium	<input type="checkbox"/>	<input type="text"/>
6. <b>jiné příjmy</b> náhrady majetkových křivd (restituce; totální nasazení, PTP); výhry z loterií, sázek, sportovních a jiných soutěží; dědictví, odstupné za uvolnění bytu; státní podpora a úroky ze stavebního spoření (pouze v případě jednorázové výplaty)	<input type="checkbox"/>	<input type="text"/>
C38K. Kontrolní součet za tab. C38		<input type="text"/>

## Sociální příjmy

C39 Pobíral(a) jste v minulém kalendářním roce některý z uvedených druhů sociálních příspěvků a dávek? Uvedte prosím počet měsíců pobírání a měsíční částku. U vybraných dávek uveďte také počet dní nemoci, resp. OČR, pokud se jednalo o dobu kratší než 1 měsíc.

		nepobíral(a)	počet měsíců	Kč	počet měsíců	Kč
1. podpora v nezaměstnanosti	rekvalifikace <input type="checkbox"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2. rodičovský příspěvek		<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
3. příspěvek na péči		<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
4. výsluhový příspěvek, odchodné (čistá částka)		<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<b>dávky nemocenského pojištění</b>						
5. nemocenské	počet dní nemoci <input type="text"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
6. ošetřovné	počet dní OČR <input type="text"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
7. peněžitá pomoc v mateřství		<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
8. vyrovnávací příspěvek v těhotenství a mateřství		<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<b>dávky důchodového pojištění</b>						
9. starobní důchod		<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
10. vdovský/vdovecký důchod		<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
11. invalidní důchod	stupeň invalidity <input type="text"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
12. sirotčí důchod	počet dětí <input type="text"/>	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
13. příspěvky pro těžce zdravotně postižené občany (na opatření zvláštních pomůcek, na úpravu bytu, na zakoupení a provoz motorového vozidla atd.)		<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
14. jiné sociální dávky		<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
C39K. Kontrolní součet za tab. C39						<input type="text"/>

**Daně z příjmu**

C40 Měl(a) jste v minulém kalendářním roce nějaké zdanitelné příjmy, ze kterých Vy resp. Váš zaměstnavatel odvádíte daň?

1. ano
2. ne  → C42

C41 Uplatňujete za rok 2012 u svého zaměstnavatele nebo ve svém daňovém přiznání nárok na následující úlevy na dani?

1. na vyživované dítě počet dětí
2. na vyživované dítě ZTP-P počet dětí
3. na manželku/manžela
4. na manželku/manžela, který je držitelem ZTP-P
5. na poživatele invalidního důchodu (1. a 2. stupeň)
6. na poživatele invalidního důchodu (3. stupeň)
7. na držitele ZTP-P
8. na studium
9. hodnota daru
10. odečet úroků
11. penzijní připojištění
12. životní pojištění
13. členský příspěvek odborů
14. odpočet ztráty
15. další položky podle §34 (reinvestice, věda, výzkum)

Poznámky:

**Penzijní připojištění**

I

C42 Platil(a) jste si v minulém kalendářním roce příspěvky na penzijní připojištění? Pokud ano, uveďte prosím počet měsíců a měsíční částku Vašich (vlastních) příspěvků v minulém kalendářním roce.

1. ano  2. ne
- počet měsíců  měsíční částka v Kč

C43 Pobíral(a) jste v minulém kalendářním roce pravidelný důchod z Vašeho penzijního připojištění? Pokud ano, uveďte prosím počet měsíců a měsíční částku pobíraného důchodu.

1. ano  2. ne
- počet měsíců  měsíční částka v Kč

**Biografické informace**

TAZATEL: Otázky C44 až C46 je nutné vyplnit **v první vlně**, dále u přistěhovaných osob, osob narozených v roce 1996 a osob, u kterých došlo ke změně.

C44 Ve které zemi jste se narodil(a)?

1. ČR (území dnešní ČR)
2. jiná země:

C45 Jaká je Vaše státní příslušnost? Pokud máte dvojí občanství, uveďte prosím obě.

1. státní příslušnost 1:
- ČR
- jiná země:
2. státní příslušnost 2:
- země:

C46 Jestliže jste pobýval(a) dlouhodobě mimo ČR, uveďte prosím rok, ve kterém jste se do ČR přistěhoval(a) nebo vrátil(a).

1. rok přistěhování
2. netýká se mě

L

CK7. Kontrolní součet za otázky C41 až C46



## Zdraví

**C47** Jak celkově hodnotíte svůj zdravotní stav?

- |                 |                          |
|-----------------|--------------------------|
| 1. velmi dobrý  | <input type="checkbox"/> |
| 2. dobrý        | <input type="checkbox"/> |
| 3. přijatelný   | <input type="checkbox"/> |
| 4. špatný       | <input type="checkbox"/> |
| 5. velmi špatný | <input type="checkbox"/> |

**C48** Máte nějakou dlouhodobou nemoc nebo dlouhodobý zdravotní problém? (Problém, který již trvá nebo bude trvat 6 a více měsíců.)

- |        |                          |
|--------|--------------------------|
| 1. ano | <input type="checkbox"/> |
| 2. ne  | <input type="checkbox"/> |

**C49** Byl(a) jste kvůli zdravotním problémům nejméně po dobu posledních 6 měsíců omezen(a) v činnostech, které lidé obvykle dělají?

- |                         |                          |
|-------------------------|--------------------------|
| 1. ano, velmi omezen(a) | <input type="checkbox"/> |
| 2. ano, omezen(a)       | <input type="checkbox"/> |
| 3. neomezen(a)          | <input type="checkbox"/> |

**C50** Kolikrát jste za posledních 12 měsíců navštívil(a) praktického lékaře nebo specialistu s výjimkou zubaře a očního lékaře?

počet návštěv

**C51** Potřeboval(a) jste během posledních 12 měsíců navštívit zubaře, a přesto jste k němu nešel(-a)?

- |   |                                       |
|---|---------------------------------------|
| 1. ano, taková situace minimálně jednou nastala a k zubaři jsem nešel(-a) | <input type="checkbox"/> ↓            |
| 2. ne, taková situace nenastala   | <input type="checkbox"/> → <b>C53</b> |

**C52** Proč jste k zubaři nešel(-a)? Uveďte hlavní důvod.

- |   |                          |
|---|--------------------------|
| 1. nemohl(a) jsem si to dovolit (příliš drahé, nehradí pojišťovna)      | <input type="checkbox"/> |
| 2. čekací seznamy, nutnost objednávat se ve velkém předstihu            | <input type="checkbox"/> |
| 3. nemohl(a) jsem se uvolnit z práce nebo od péče o děti či jinou osobu | <input type="checkbox"/> |
| 4. daleké cestování, nebyl vhodný způsob dopravy                        | <input type="checkbox"/> |
| 5. strach ze zubaře, vyšetření, léčby apod.                             | <input type="checkbox"/> |
| 6. chtěl(a) jsem počkat, zda se zdravotní problém sám nezlepší          | <input type="checkbox"/> |
| 7. neznal(a) jsem žádného dobrého zubaře                                | <input type="checkbox"/> |
| 8. z jiného důvodu  | <input type="checkbox"/> |

**C53** Potřeboval(a) jste během posledních 12 měsíců navštívit lékaře (kromě zubaře), a přesto jste k němu nešel(-a)?

- |   |   |
|---|---|
| 1. ano, taková situace minimálně jednou nastala a k lékaři jsem nešel(-a) | <input type="checkbox"/> ↓              |
| 2. ne, taková situace nenastala   | <input type="checkbox"/> → <b>konec</b> |

**C54** Proč jste k lékaři nešel(-a)? Uveďte hlavní důvod.

- |   |                          |
|---|--------------------------|
| 1. nemohl(a) jsem si to dovolit (příliš drahé, nehradí pojišťovna)      | <input type="checkbox"/> |
| 2. čekací seznamy, nutnost objednávat se ve velkém předstihu            | <input type="checkbox"/> |
| 3. nemohl(a) jsem se uvolnit z práce nebo od péče o děti či jinou osobu | <input type="checkbox"/> |
| 4. daleké cestování, nebyl vhodný způsob dopravy                        | <input type="checkbox"/> |
| 5. strach z lékařů, nemocnic, vyšetření, léčby apod.                    | <input type="checkbox"/> |
| 6. chtěl(a) jsem počkat, zda se zdravotní problém sám nezlepší          | <input type="checkbox"/> |
| 7. neznal(a) jsem žádného dobrého lékaře                                | <input type="checkbox"/> |
| 8. z jiného důvodu  | <input type="checkbox"/> |

Poznámky:

## **Příloha č. 4**



CM5	Máte někoho, komu se můžete svěřit s osobními záležitostmi?
1.	ano <input type="checkbox"/>
2.	ne <input type="checkbox"/>
3.	nedokážu posoudit <input type="checkbox"/>

CM6	Myslíte si, že byste mohl(a) v případě potřeby požádat své příbuzné, přátele nebo sousedy o pomoc?
1.	ano <input type="checkbox"/>
2.	ne <input type="checkbox"/>
3.	nemám příbuzné, přátele, sousedy <input type="checkbox"/>
4.	nedokážu posoudit <input type="checkbox"/>

CM7	Do jaké míry byste řekl(a), že se lidem dá důvěřovat? Při odpovědi použijte, prosím, stupnici od 0 ( <i>vůbec</i> ) do 10 ( <i>zcela</i> ).	
1.	důvěra <input type="checkbox"/> <input type="checkbox"/>	2. nedokážu posoudit <input type="checkbox"/>

CM8	Do jaké míry důvěřujete: Při odpovědi použijte, prosím, stupnici od 0 ( <i>vůbec nedůvěřuji</i> ) do 10 ( <i>zcela důvěřuji</i> ).	
	důvěra	nedokážu posoudit
1.	státnímu systému ČR <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>
2.	právnímu systému ČR <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>
3.	Policii ČR <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>

CM9	Jak bezpečně se cítíte v okolí Vašeho bydliště po setmění?
1.	zcela bezpečně <input type="checkbox"/>
2.	spíše bezpečně <input type="checkbox"/>
3.	nepříliš bezpečně <input type="checkbox"/>
4.	vůbec ne bezpečně <input type="checkbox"/>
5.	nedokážu posoudit <input type="checkbox"/>

CM10	Jak předpokládáte, že by se mohla vyvíjet Vaše celková životní situace? Uvažujte <u>příštích 12 měsíců</u> .
1.	zlepší se <input type="checkbox"/>
2.	nezmění se <input type="checkbox"/>
3.	zhorší se <input type="checkbox"/>
4.	nedokážu posoudit <input type="checkbox"/>

Poznámky:

Mnohokrát děkujeme za účast na tomto rozhovoru