



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Možnosti implementace nového informačního systému o terciárním vzdělávání

Autoři:

doc. Ing. Štěpán Jurajda, Ph.D.

Mgr. Jan Mysliveček, Ph.D.

Ing. Filip Pertold, M.A.

Mgr. Pavla Nikolovová

Mgr. Mário Vozár, M.A.



Seznam použitých zkratek

Bc.	Bakalářský
ČR	Česká republika
ČSSZ	Česká správa sociálního zabezpečení
ČVUT	České vysoké učení technické
ESF	Evropský sociální fond
KZAM	Kategorie zaměstnání
ILR	Individualized Learner Record (individualizovaný záznam žáka)
ISA	Informationssystem Studienwahl und Arbeitsmarkt (Informační systém pro výběr studia a pracovní trh)
LSC	Learning and Skills Council (Rada pro učení a dovednosti)
Mgr.	Magisterský
MŠMT	Ministerstvo školství, mládeže a tělovýchovy
MŠ SR	Ministerstvo školství Slovenské republiky
N.Mgr	navazující magisterský
OECD	Organizace pro hospodářskou spolupráci a rozvoj
Ph.D.	Doktorský
SBD	Statistische Bundesamt Deutschland (Německý statistický úřad)
SIMS	Sdružené informace matrik studentů
ÚIPŠ	Ústav informací a prognóz školství
ÚIV	Ústav pro informace ve vzdělání
UK	Univerzita Karlova
VŠ	Vysoká škola
VŠCHT	Vysoká škola chemicko-technologická
VUT	Vysoké učení technické v Brně

Hlavní závěry a doporučení

- Hlavní cíl nového informačního systému je zlepšit informovanost široké veřejnosti o terciárním vzdělávání v České republice
- Pomocí dat ze Sdružených informací matrik studentů (SIMS) navrhujeme rozšířit široce dostupný počet indikátorů až do úrovně programů na vysokých školách.
- Navržené indikátory by měly především popisovat průchod studentů studiem na v daném programu a škole a jejich úspěšnost dokončování studia.
- Navrhujeme zavést povinnost pro většinu nepovinných údajů dnes sbíraných v SIMS a zvážit sběr ještě detailnějších informací o průchodu studiem - například o absolvovaných předmětech, včetně informací o předmětech samotných.
- Současnou databázi SIMS navrhujeme rozšířit o některé socioekonomické charakteristiky studentů, aby mohl být plnohodnotně zachycen rozsah selektivity u jednotlivých vysokých škol a vztah k prezentovaným indikátorům.
- Navrhujeme využít probitový model, případně model hazardní míry, k zachycení vlivu selekce a k modelování úspěšnosti studentů při průchodu studiem.
- Navrhujeme propojit SIMS s externími databázemi, které by popisovaly studium před vysokou školou a případně působení a úspěšnost absolventů na trhu práce. V současné době doporučujeme propojení s existující databází Uchazeč.
- Výsledné indikátory, spočítané na základě této práce, by se měly prezentovat veřejnosti srozumitelným a transparentním způsobem s důrazem na omezení, která vyplývají z charakteru dat.
- Databáze SIMS by měla být zpřístupněna v anonymizované podobě pro akademický výzkum, aby se rozšířil záběr a možnosti analýz terciárního vzdělávání.

Shrnutí

V této práci se zabýváme možnostmi rozšíření současných dostupných informací o terciárním vzdělávání v České republice. V úvodu konstatujeme, že v současné době je v České republice relativně malá informovanost veřejnosti o tom, jak probíhá studium na vysokých školách a to i přes vysokou poptávku po vysokoškolském studiu a po informacích o kvalitě a charakteru studia na různých vysokých školách.

V současné době zprostředkovává nejvíce informací Ústav pro informace ve vzdělávání (ÚIV), který přináší řadu statistických indikátorů o vysokých školách a vybraných programech. V této práci navrhujeme výrazně rozšířit dostupnou řadu indikátorů s cílem podpořit informované rozhodování uchazečů o studiu na vysoké škole. Navíc by tento krok zlepšil celkovou informovanost veřejnosti, soukromého a veřejného sektoru o terciárním vzdělávání.

Text je rozdělen do šesti základních částí (*Přehled informací vedených v SIMS, Seznam a popis možných indikátorů ze stávající SIMS, Možnosti rozšíření SIMS, Návrhy na propojení SIMS, Informace sbírané v zahraničí a indikátory z nich generované a Využití indikátorů pro potřeby veřejnosti*) a přílohy.

V první části popisujeme všechny informace obsažené v současné formě Sdružených informacích matrik studentů (SIMS), což je základní zdroj dat pro naše navrhované indikátory. Jedná se o řadu charakteristik studentů a poměrně detailní informace o průběhu jejich studia: *jméno a rodné číslo, adresa, předchozí vzdělání, vysoká škola, program, zápis, délka studia, postup studiem, sociální údaje, ukončení studia, studijní pobyt, jazyk výuky, obor a jeho akreditace, místo výuky, forma studia, financování a přerušování studia*. U každého údaje udáváme, zda je povinný nebo nepovinný, v jaké formě je zapsaný a v čem případně spočívá jeho případná užitečnost. U většiny nepovinných údajů též doporučujeme zavést povinnost jejich vyplňování.

V části *Seznam a popis možných indikátorů ze stávající SIMS* navrhujeme řadu indikátorů, které lze ze současné databáze SIMS počítat nebo odhadovat. Indikátory jsou

rozděleny na jednoduché a pokročilé a to dle náročnosti na využití statistické metody. Jednoduché indikátory popisují především počty studentů dle řady kritérií: *počty studentů rozdělené na geografické jednotky, počet duplicitních studentů, počty absolventů, počet neúspěšných studentů-studií, počty přerušení, počet studentů nad standardní dobu studia, počty přestupů, počet studentů s trvalým bydlištěm v kraji působnosti školy a počet studentů dle absolvované střední školy*. K jednoduchým indikátorům jsme zařadili také délky studia (*úspěšných studií, relativní délka studia vzhledem k akreditované délce, studia na zahraničním pobytu, délka přerušení*), (ne)úspěchy při studiu (*průměrná délka neúspěšného studia, „úmrtnost“ po ročnících, poměr počtu neúspěšných a úspěšných studentů, míra přestupu na magisterský program, přestupy mezi školami, studia delší než standardní doba*) a indikátory o jazyku výuky (*počet programů akreditovaných v cizím jazyce, počty studentů a absolventů programu v cizím jazyce*). Všechny tyto indikátory se samozřejmě mohou navzájem doplňovat a zdaleka nemusí být prezentovány všechny. Přes navržené velké množství indikátorů navrhujeme i další tzv. pokročilé indikátory, které se detailně popisují pravděpodobnosti úspěšnosti dokončení studia nebo jeho části.

Část *Pokročilé indikátory ze současné databáze SIMS* obsahuje dvě části – *Pravděpodobnosti úspěšnosti a Regresní analýza v současné databázi SIMS*. První část detailně rozebírá problém správného výpočtu *Pravděpodobnosti dokončení studia, Podmíněné pravděpodobnosti dokončení studia po dokončení daného ročníku, Pravděpodobnosti dosažení daného ročníku a Pravděpodobnosti dokončení ročníku X za předpokladu dokončení ročníku X-1*. Ve všech případech zdůrazňujeme, že výpočty musí vycházet z tzv. „kohortního“ přístupu, kdy se sleduje prostup jedné generace studentů celým studiem. V části *Regresní analýza v současné databázi SIMS* ukazujeme jisté možnosti pro tuto analýzu se současnou SIMS. Vzhledem k omezenosti dostupných informací o charakteristikách studentů tuto problematiku rozvádíme až pro případ rozšíření SIMS o další individuální charakteristiky.

Další část této práce navrhuje možnosti jak současnou databázi SIMS rozšířit o další důležité charakteristiky studentů a studijních programů. Nejdříve se zabýváme rozšířením o *Socioekonomické charakteristiky, jako je Vzdělání a příjem rodičů, Sourozenci, Místo pobytu a dojíždění a Příjem a pracovní činnost*. Poté navrhujeme rozšíření o *Studijní*

výsledky a zaplacené poplatky a Charakteristiky programu, které se týkají především nákladů na financování (*nákladový Koeficient programu, Náklady na absolventa a Zdroje financování*).

Z navržených charakteristik dále popisujeme i možné indikátory, které se z těchto charakteristik dají vyjádřit. V případě jednoduchých indikátorů navrhujeme soubor průměrů jednotlivých individuální charakteristik za jednotlivé školy a programy (*Průměrné vzdělání a příjem rodičů, Průměrný počet sourozenců a jejich vzdělání, Průměrná vzdálenost dojíždění*). Ne všechny tyto indikátory musí být nutně přímo dostupné široké veřejnosti a mohou sloužit pouze akademické veřejnosti. Následně se zabýváme pokročilými indikátory, které by mohly být výsledkem jednoduché regresní analýzy.

Popisujeme základní probitový model, který umožňuje identifikovat vliv individuálních charakteristik (například výsledky u maturity, typ střední školy, nebo vzdělání rodičů) na určitou vysvětlující proměnnou (například na pravděpodobnost dokončení studia). Regresní analýza by měla vyústit v porovnání hrubé (pozorované) pravděpodobnosti dokončení studia s pravděpodobností dokončení studia pro určitý typ studenta (například pro charakteristiky průměrného studenta, který nastupuje do programu chemie).

V kapitole *Návrhy na propojení SIMS* ukazujeme možnosti jejího napojení na externí zdroje a získání dodatečných informací o předchozích aktivitách studentů nebo o jejich uplatnění na trhu práce. Nejdříve popisujeme databázi Uchazeč, která by mohla přinést cenné informace o střední škole, kterou student navštěvoval a o tom, na kolik a na jaké vysoké školy si podal přihlášky. Dále popisujeme případné využití *Předchozích studijních výsledků*, což by mohla být důležitá informace o selekci studentů na jednotlivé programy škol.

V kapitole zabývající se uplatněním absolventů na trhu práce nejdříve diskutujeme *Možné datové zdroje* a upozorňujeme na momentální hlavní problémy, které by hrály roli při propojování SIMS s informacemi o mzdách absolventů. Jako hlavní potenciální zdroj dat o mzdách absolventů navrhujeme záznamy z České správy sociálního zabezpečení při zajištění nezbytně anonymizace. Následně navrhujeme některé jednoduché indikátory:

Průměrná mzda čerstvých absolventů podle programů a škol, Průměrná mzda čerstvých absolventů podle regionů, Vývoj mezd absolventů během kariéry a Mzdy přijatých uchazečů. Všechny tyto indikátory lze samozřejmě modifikovat a upravovat dle potřeby a dostupnosti dat.

Další propojení, které navrhujeme, je zamýšleno především pro potřeby monitoringu *nezaměstnanosti* a *zaměstnanosti* absolventů. Nezaměstnanost absolventů je již dnes prezentována na internetovém portálu www.infoabsolvent.cz. Problém zaměstnanosti absolventů podle vystudované školy a programu však v současné době není monitorován, přestože to může být, podobně jako mzdy, důležitá informace pro uchazeče. Zde by bylo možné vycházet i z případného propojení s daty České správy sociálního zabezpečení nebo s daty Informačního systému o průměrném výdělku (s určitým omezením – tj. pokud k tomuto externímu zdroji dat bude připojen identifikátor odbožený z rodného čísla). Zaměstnanost i nezaměstnanost by mohla být, podobně jako mzdy, prezentována ve struktuře podle programu, regionu a vývoji během kariéry.

V návaznosti na navržené indikátory uvádíme i možnost vytvoření *Indexu uplatnění absolventů na trhu práce*, který kombinuje údaje o mzdách, zaměstnanosti a nezaměstnanosti do jednoho souhrnného indexu. Nedomníváme se však, že by měl být tento indikátor hlavním výstupem informačního systému a to zejména proto, že jeho konstrukce podléhá určitým předpokladům o kvalitě uplatnění, které mohou být velice subjektivní.

V kapitole *Pokročilé indikátory z propojené databáze SIMS* se věnujeme regresní analýze, která by se zaměřila na vysvětlení přínosu školy na mzdu absolventa. Regresní model by se mohl zaměřit na to, aby dokázal oddělit vliv dané vysoké školy od ostatních individuálních charakteristik (například úspěšnost u maturitní zkoušky, rodinné zázemí, daný region a podobně). Podobně by se mohlo postupovat i při vysvětlování nezaměstnanosti absolventů škol.

V šesté části se zabýváme problematikou *Informace sbírané v zahraničí a indikátory z nich generované*. Ukazujeme, jaké zahraniční zdroje jsou v současné době k dispozici a co z nich se blíží potřebám České republiky. Hlavní závěr shrnutí zahraničních

dostupných zdrojů je to, že v evropském měřítku není dosud vytvořen dostatečně přehledný a transparentní informační systém, který by byl plně a snadno dostupný široké veřejnosti a který by bylo možné beze zbytku následovat. Nejpodrobněji popisujeme informační systém v *Německu*, které je k nám historicky s systémově relativně blízko. Na rozdíl od současného stavu v České republice, německý informační systém poskytuje řadu deskriptivních statistik v poměrně dlouhých časových řadách. Navíc tam lze nalézt data o průměrné délce studií a o úspěšnosti dokončování studií. Tyto informace však nejsou na úrovni vysoké školy, ale pouze na úrovni poměrně široce definovaného oboru nebo regionu. To však z našeho pohledu není dostatečné, protože to nezlepšuje pozici uchazeče při rozhodování o studiu na různých vysokých školách.

Dále se zabýváme informačními systémy nebo studii o terciárním školství na *Slovensku, Velké Británii, Nizozemí, Finsku a mezinárodními studii*. Z toho informační systém ve Finsku nabízí zajímavé informace o postupu studentů ve studiu prostřednictvím stránek statického úřadu.

Na závěr hlavní části diskutujeme *Využití indikátorů pro potřeby veřejnosti*. Upozorňujeme na základní pravidla, která by se měla dodržovat při představování výsledků laické i odborné veřejnosti, případně MŠMT, akreditační komisi a zahraničním institucím. Především zdůrazňujeme potřebu srozumitelnosti jednotlivých ukazatelů a statistik a důraz na jejich správnou interpretaci. V tomto kontextu je nutné upozorňovat na omezení, která jsou dána především tím, že data zachycují minulý vývoj a tudíž nemohou zcela věrohodně odhadovat vývoj budoucí.

Hlavním uživatelem informačního systému by měl být uchazeč o studium, kterému by výsledky měly usnadnit rozhodování o vysoké škole a programu. Z tohoto důvodu by měl být dáván důraz na možnost srovnávat různé vysoké školy dle zadaných kritérií a celkově tak zlepšit i informovanost široké veřejnosti o terciárním vzdělávání.

Tato práce obsahuje též řadu příloh, které doplňují základní text. V příloze 1 představujeme podrobný popis SIMS, tak jak je v současné době k dispozici i s podrobnými vysvětlivkami k jednotlivým informacím v něm obsaženým. V druhé části přílohy poskytujeme *Přehled indikátorů prezentovaných Ústavem pro informace ve*

vzdělání. Třetí část přílohy se zabývá podrobným popisem modelovacích technik pro odhad pravděpodobnosti dokončení studia (*Modelování pravděpodobnosti úspěšného dokončení studia – komplexní pohled*).

V této příloze komplexně diskutujeme regresní analýzu popisující a vysvětlující pravděpodobnost úspěšného dokončení studia. Podrobně rozebíráme typy vstupních dat, které model může využívat a hlavně problémy, které se mohou vyskytnout při odhadování, a to zejména výběrové zkreslení. Do technických detailů též popisujeme probitový model, který má jako vysvětlovanou proměnnou binární veličinu (dokončil nebo nedokončil VŠ). Samostatně se věnujeme správné interpretaci výsledků probitového modelu.

Alternativní modelovací technika, kterou lze též použít je *hazardní míra*, které by byla vhodná pro odhad kontinuální pravděpodobnosti „úmrtnosti“ studentů. Výhoda měř hazardu oproti probitovému modelu spočívá v tom, že vykreslují celý průchodu studentů daným vzdělávacím programem.

Na závěr přílohy předkládáme souhrn *Současné akademické literatury* na téma předčasné ukončení vysokoškolského studia. Jeden ze zajímavých výsledků je ten, že pravděpodobnost dokončení studia je závislá na věku při začátku studia. Čím je věk vyšší, tím se zvyšuje pravděpodobnost nedokončení vysokoškolského vzdělání.

Obsah

1	Úvod a motivace	12
2	Přehled údajů vedených v SIMS	15
2.1	Úvod do SIMS (zákonost sběru dat, mechanismus sběru, správce)	15
2.2	Stručný přehled sbíraných dat	15
2.2.1	Jméno, rodné číslo, stávající tituly	15
2.2.2	Adresa	16
2.2.3	Předchozí vzdělání, SŠ, Rok maturitní zkoušky	16
2.2.4	Studium (VŠ, Fakulta, Program, Zápis)	16
2.2.5	Délka studia	16
2.2.6	Postup studiem	17
2.2.7	Sociální údaje	17
2.2.8	Ukončení studia	17
2.2.9	Studijní pobyt	17
2.2.10	Jazyk výuky	17
2.2.11	Obor a jeho aprobace	18
2.2.12	Místo výuky	18
2.2.13	Forma studia	18
2.2.14	Financování	18
2.2.15	Přerušování studia	18
3	Seznam a popis možných indikátorů ze stávající SIMS	18
3.1	Jednoduché indikátory ze stávající matriky studentů	19
3.1.1	Počty studentů	19
3.1.2	Délky studia	22
3.1.3	(Ne)úspěchy při studiu	23
3.1.4	Indikátory o jazyku výuky	25
3.2	Pokročile indikátory ze současné databáze SIMS	25
3.2.1	Pravděpodobnosti úspěšnosti	26
3.2.2	Regresní analýza v současné databázi SIMS	29
4	Možnosti rozšíření SIMS	30
4.1	Charakteristiky studenta	30
4.1.1	Socioekonomické charakteristiky	30
4.1.2	Studijní výsledky a zaplacené poplatky	32
4.2	Charakteristiky programu	32
4.2.1	Předměty a jejich obsah	33
4.2.2	Převis poptávky po daném programu	33
4.2.3	Koeficient programu	33
4.2.4	Náklady na absolventa	33
4.2.5	Zdroje financování	34
4.3	Indikátory z rozšířené databáze SIMS	34
4.3.1	Jednoduché indikátory	34
4.3.2	Pokročilé indikátory z rozšířené SIMS – regresní analýza	35
5	Návrhy na propojení SIMS	40
5.1	Předchozí studium	41
5.1.1	Propojení s databází Uchazeč	41

5.1.2	Výsledky maturit a jiných testů	42
5.1.3	Výsledky z předchozího studia na vysoké škole.....	42
5.2	Úspěšnost absolventa na trhu práce – mzdy	43
5.2.1	Možné datové zdroje - diskuse	43
5.2.2	Průměrná mzda čerstvých absolventů podle programů a škol.....	44
5.2.3	Průměrná mzda čerstvých absolventů podle regionů.....	44
5.2.4	Vývoj mezd absolventů během kariéry.....	44
5.2.5	Mzdy přijatých uchazečů	45
5.3	Zaměstnanost absolventů	45
5.4	Nezaměstnanost absolventů.....	45
5.5	Index celkové kvality uplatnění absolventů na trhu práce	46
5.6	Pokročilé indikátory z propojené databáze SIMS.....	46
5.6.1	Regresní analýza vysvětlující vliv školy na mzdy.....	46
5.6.2	Regresní analýza vlivu školy na zaměstnanost a nezaměstnanost.....	47
6	Informace sbírané v zahraniční a indikátory z nich generované.....	48
6.1	Německo	48
6.2	Finsko.....	49
6.3	Nizozemí	50
6.4	Slovensko.....	51
6.5	Velká Británie	53
6.6	Mezinárodní studie.....	56
7	Využití indikátorů pro potřeby uživatelů	57
7.1	Laická veřejnost	58
7.2	Odborná veřejnost.....	59
7.3	Ministerstvo školství, mládeže a tělovýchovy	59
7.4	Akreditační komise MŠMT	59
7.5	Zahraničí instituce.....	59
8	Závěr	60
9	Reference a internetové zdroje.....	62
10	Přílohy.....	65
P.1	Úplný seznam vedený záznamů v SIMS	66
P.2	Přehled indikátorů prezentovaných Ústavem pro informace ve vzdělání	76
P.3	Základní údaje obsažené v databázi Uchazeč	79
P.4	Modelování pravděpodobnosti úspěšného dokončení studia.....	80
P.5	Akademické studie.....	97

1 Úvod a motivace

Údaje sbírané o studentech VŠ a jejich studiích představují mimořádně hodnotný zdroj informací nejen pro veřejnou správu a soukromý sektor, ale i pro vysoké školy samotné studenty a především zájemce o studium. V současné době jsou tyto bohaté informace využívány jen v omezené míře pro určení vzdělávacích dotací školám a pro výpočet velice základních statistických ukazatelů o systému terciárního vzdělávání. Přitom je z médií zřejmé, že poptávka veřejnosti po informacích o terciárním vzdělávání je velická. To ostatně plyne i z toho, že po roce 1989 se podle řady empirických studií výrazně zvýšily výnosy ze vzdělání a z toho plynoucí vysoký nárůst poptávky po studiu na vysoké škole. V současné době mívá za studiem na vysoké škole již kolem poloviny maturitního ročníku a celkový počet přijatých dosahuje velikosti jedné věkové kohorty a to díky starším generacím, které se vstupují do systému později nebo se vrací doplnit si vzdělání.

Vysokou poptávku po informacích o vysokých školách můžeme pozorovat na zájmu médií prezentovat řadu žebříčků vysokých škol v rámci určitých, široce definovaných, oborů. Takovéto pokusy byly za poslední dva roky prezentovány v několika celostátních denících (Lidové noviny, Hospodářské noviny, Mladá fronta Dne)s a na internetové stránce www.aktualne.cz. Všechny tyto žebříčky vycházejí z různě věrohodných informačních zdrojů a dokumentují poptávku veřejnosti po podobných informacích.

Co se týká dalšího uplatnění absolventů na trhu práce, což může mít vypovídací hodnotu i o dané vysoké škole, existuje poměrně informativní internetový server www.infoabsolvent.cz, vytvořený Národním ústavem odborného vzdělávání. Ten ukazuje poměrně podrobné statistiky dle zaměstnanosti a nezaměstnanosti absolventů podle jednotlivých oborů.

Tato práce nejvíce navazuje na nejkomplexnější informační zdroj, které v České republice existuje - Ústav pro informace ve vzdělání (ÚIV), který publikuje řadu

užitečných deskriptivních statistik, zjištěných mimo jiné též ze SIMS, podle jednotlivých programů¹ a jednotlivých vysokých škol. Přesný přehled již dostupných indikátorů je uveden v příloze tohoto textu. Většinu těchto indikátorů představují absolutní počty studentů v dělení podle kritérií (programy, školy, regiony).

Podstata a hlavní výhoda námi navržených indikátorů spočívá v tom, že popisují studium jednotlivých programů na vysokých školách. To usnadní rozhodování uchazečů o studium na konkrétní vysoké škole. Indikátory tohoto typu jsou zatím poměrně nedostupné a pokud existují tak v nízké komparativní kvalitě. Námi navrhované indikátory by též měly být prezentovány v časových řadách, což v některých případech doplní již známé indikátory prezentované ÚIV, prezentované v nesourodé podobě.

Tato studie též analyzuje užitečnost statistických informací, které lze z existujících informačních zdrojů spočítat nebo odhadnout a identifikuje pro koho tyto informace mohou být zajímavé. Kromě toho prezentujeme několik možností na rozšíření sbíraných informací a výhody, které by to přineslo. Navrhovaná rozšíření vycházejí z dostupnosti daných informací a míry jejich užitečnosti pro uživatele. Tato práce respektuje omezenost zdrojů na získání dodatečných informací a proto je rozdělena do třech základních částí popisující jednotlivé indikátory. V první části pracujeme pouze se současným stavem sbíraných informací, jak je představuje databáze SIMS a na základě toho navrhujeme dostupné statistiky a indikátory, které by mohly být prezentovány veřejnosti. V druhé části navrhujeme možné rozšíření o další charakteristiky studentů a škol, které by bylo možné získat pouhým „došetřením“ studentů o další informace nebo případně jednoduchým dotazníkovým šetřením. Vytvořili jsme seznam možných dodatečně zjišťovaných charakteristik s možnostmi jejich využití pro uživatele. V třetí části prezentujeme možnosti propojení informačního systému s dalšími státem spravovanými informačními databázemi. Jelikož SIMS obsahuje identifikátor každého jednotlivce

¹ V ideálním případě by se navrhované mohly uvádět i na úrovni akreditovaných oborů. V současné době však není jisté, že by to databáze SIMS dovolovala, a zda by v důsledku mohlo dojít k do statečně transparentnímu srovnání mezi školami, vzhledem k velké heterogenitě studijních oborů.

(rodné číslo) je možné ji propojit s dalšími již existujícími databázemi a využívat dodatečné informace z nich.

V textu dále popisujeme různé informační systémy o terciárním vzdělávání v zahraničí. Nejpodrobněji se zabýváme případem Německa, které je systémem školství poměrně podobné České republice a kde v některých aspektech prezentují podobné indikátory, které navrhuje.

V tomto textu se však k žádnému vzorovému informačnímu systému ani publikacím o terciárním vzdělávání neuvážeme, jelikož žádný plnohodnotný systém dle našeho názoru v současnosti neexistuje. Podle dostupných informací je v současné době podobný informační systém předmětem veřejných diskusí v Německu.

Všechny informace, které se již v SIMS sbírají nebo se potenciálně mohou sbírat, lze využít pro výpočet indikátorů, které mají rozdílnou vypovídací schopnost a náročnost na výpočet. Proto rozdělujeme tyto statistické indikátory do dvou skupin – jednoduché a pokročilé.

Jednoduché indikátory se dají triviálně spočítat z dat již existujících nebo dodatečně sesbíraných, bez nutnosti využití pokročilejších statistických metod. Jedná se obvykle o absolutní počty (například počet studentů na daném programu v dané škole, případně počty studentů v daném kraji atd.) nebo o jednoduché průměry (například průměrná délka studia, průměrný počet vystudovaných studentů za rok apod.).

Jelikož tyto statistické údaje by samy o sobě nemusely stačit na přesné zachycení některých významných jevů, navrhuje též skupinu takzvaných pokročilých indikátorů. Ty vyžadují již znalosti ze statistiky a ekonometrie, aby mohly být správně spočítány, případně odhadnuty. Vždy se snažíme prezentovat jejich základní využití, interpretaci a výhody oproti jednoduchým indikátorům.

Využívání (a tedy i zveřejňování) navržených indikátorů by bylo prospěšné nejen pro studenty, management škol, ale i pro efektivnější využívání veřejných prostředků. Studenty může zajímat náročnost studia a tudíž vědět, kdy a jak dochází k neúspěšnému ukončení studia. Školy mohou zdůraznit prostupnost svých programů na základě

informací o tom, kolik studentů k nim přišlo nebo lákat studenty na dobré šance vycestovat do zahraničí. MŠMT a veřejná správa obecně, může využít informace o nákladech, délkách studií a šancích na úspěšné dokončení pro lepší alokaci finančních prostředků.

2 Přehled údajů vedených v SIMS

2.1 Úvod do SIMS (zákonost sběru dat, mechanismus sběru, správce)

Současný systém sběru informací o studentech a jejich studiích je zakotvena v zákoně č. 111/1998 Sb. o vysokých školách v pozdějších zněních. Na základě tohoto zákona vede každá vysoká škola vlastní matriku studentů. Zákon stanovuje minimum informací, které tyto matriky obsahují a také opravňuje Ministerstvo školství, mládeže a tělovýchovy ČR (dále jen MŠMT) ke sdružování těchto informací.

Informace jsou vedeny v elektronické databázi, jejíž podoba byla vypracována ve spolupráci s Ústavem výpočetní techniky Masarykovy univerzity v Brně. Její název je Sdružené informace matrik studentů (zkráceně SIMS, <http://sims.ics.muni.cz/>). Správce matriky v každé vysoké škole je zodpovědný za správnost a úplnost informací vedených v SIMS. Sběr informací pro SIMS probíhá zpravidla čtyřikrát ročně.

2.2 Stručný přehled sbíraných dat

Základní sbíranou jednotkou je záznam o studentovi (rodné číslo), VŠ/Fakultě, Studijním programu a datu zápisu. Tyto údaje jednoznačně identifikují každého studenta a jeho studium. O každém studiu je zpravidla čtyřikrát do roka (k 31.3., 30.6., 31.10. a k 31.12.) sbírána řada údajů, z nichž nejdůležitější uvádíme v následujících podkapitolách. Kompletní přehledová tabulka se nachází v první příloze tohoto textu. Jak uvádíme níže, některé údaje jsou sbírány nepovinně. Pro budoucí analytické použití SIMS by bylo dle našeho názoru důležité zavedení povinnosti při jejich získávání a tím by se rozšířily analytické možnosti analýz dat.

2.2.1 Jméno, rodné číslo, stávající tituly

O každém studentovi je evidováno jeho jméno, příjmení, rodné číslo a stávající tituly (nepovinné). Zejména rodné číslo je důležité pro potenciální propojení databáze

s ostatními databázemi. Rodné příjmení může být rovněž evidováno, je však položkou nepovinnou.

2.2.2 Adresa

SIMS teoreticky eviduje podrobnou trvalou adresu studenta, včetně identifikace obce, poštovní směrovací číslo. Tento údaj je možné využít ke studiu pohybu studentů mezi obcemi a kraji a pro analýzu vztahu geografické dostupnosti škol a úspěšnosti u studia, případně pobytu na koleji, samozřejmě se zřetelem k některým problémům s tímto údajem spojeným. Například, že školy mají povinnost uvádět pouze okres a stát a údaj proto nemusí být kompletní.

2.2.3 Předchozí vzdělání, SŠ, Rok maturitní zkoušky

Školy mají povinnost uvádět nejvyšší předchozí dosažené vzdělání („úplné střední vzdělání“ apod.). Údaj o konkrétní střední škole je nepovinný, stejně jako rok maturitní zkoušky. Pokud budou v databázi tyto informace dostupné v dostatečném rozsahu, bude možné analyzovat vztah mezi vystudovanou střední školou a úspěšností v dostudování, což může být cenná informace. Případně by bylo možné sledovat na jakých vysokých školách se vyskytují studenti z konkrétních středních škol.

2.2.4 Studium (VŠ, Fakulta, Program, Zápis)

Studium je identifikováno VŠ, její fakultou, studovaným programem a datem zápisu. Jde o povinné údaje. Spojením záznamů z různých období lze získat podrobný přehled toho, jak se studium daného studenta vyvíjelo v průběhu času. Samotný jediný zápis představuje stav daného studia, nikoliv historii, jak k tomuto stavu student a jeho studium dospěli.

2.2.5 Délka studia

Tento údaj eviduje standardní dobu studia daného studijního programu. Tento údaj je důležitý pro vytvoření indikátorů, které porovnávají skutečnou dobu studia daného studenta s teoretickou dobou studia.

2.2.6 Postup studiem

SIMS eviduje, zda jde o studium, do kterého je student nově přijat (tj. nebyl evidován záznam v předchozím období o tomtéž studiu a datum zápisu následuje datum předchozího sběru dat), a zda jde o navazující studijní program. Dále se sleduje, zda se studium dělí na ročníky a pokud ano, ve kterém ročníku se student právě nachází. Tyto údaje jsou ale nepovinné.

2.2.7 Sociální údaje

Do toho patří, zda student bydlí na koleji a zda má nárok na sociální stipendium. Tento údaj může být využit pro orientaci uchazečů, do jaké míry jsou na určité škole dostupné koleje.

2.2.8 Ukončení studia

Pokud studium bylo ve sledovaném období ukončeno, SIMS eviduje datum ukončení, způsob ukončení a případně udělený titul. Způsoby ukončení zahrnují ukončení studia na základě žádosti studenta, vyloučení ze studia, úspěšným vykonáním státní závěrečné zkoušky, přestoupením na jiný program/VŠ atd. Tato informace bude zásadní pro indikátory o (ne)úspěšném ukončení dané školy.

2.2.9 Studijní pobyt

Školy mají možnost do matriky uvádět informaci o tom, zda student ve sledovaném období vycestoval nebo ukončil studijní pobyt. Pokud tuto informaci uvedou, musejí uvést typ studijního pobytu. Pro uchazeči o studiu, to může být důležitá informace o tom, jakou má šanci vycestovat v rámci studia do zahraničí.

2.2.10 Jazyk výuky

SIMS eviduje informaci o jazyku výuky. Pokud škola poskytuje výuku v cizím jazyce, může to být informace o kvalitě této školy. Tento údaj je důležitý nejen z těchto věcných důvodů, ale také proto, že i veřejné VŠ mají právo a povinnost vybírat poplatky za studium vedeno v cizím jazyce.

2.2.11 Obor a jeho aprobace

V SIMS je rovněž evidován obor studia (až 4, bez ohledu na pořadí) a případná aprobace těchto oborů. Jde o nepovinné údaje.

2.2.12 Místo výuky

Tento údaj eviduje obec, ve které je příslušná etapa studia vykonávána. Jde o povinný údaj, podstatný zejména v případě, že výuka probíhá mimo sídlo školy. V kombinaci s informací o trvalém pobytu může být zajímavé sledovat, zda škola láká studenty z celé republiky, či se jedná pouze o vysokou školu s regionálním dosahem.

2.2.13 Forma studia

Možné formy studia jsou prezenční, distanční a kombinované.

2.2.14 Financování

Tato položka shrnuje, zda studium je ve standardní době, či zda je tato doba překročena o jeden či dva a více let. Dále je evidováno, zda jde o studium na soukromé VŠ, souběžné či navazující.

2.2.15 Přerušení studia

Informace o tom, zda studium probíhá nebo je přerušeno. Pokud studenti často přerušují studium, může to být informace o tom, že na dané škole nemusí být vše v pořádku.

3 Seznam a popis možných indikátorů ze stávající SIMS

V této kapitole uvádíme seznam a podrobný popis indikátorů, které SIMS umožňuje vypočítat. Pro lepší přehlednost dělíme indikátory do několika tématicky podobných skupin. Všechny uvedené indikátory je možné počítat pouze pro veřejné VŠ, pouze pro soukromé VŠ, a nebo souhrnně.

Většinu níže uváděných indikátorů je také možné prezentovat ve formě srovnání dvou či více vybraných škol, případně pro dané programy. Výstupem může být srovnání průměrné „úmrtnosti“ po ročnících na ČVUT a VUT v Brně. Vzhledem k tomu, že tento přístup znásobuje počet možných indikátorů a srovnání, je otázkou jaké vhodného

(webového) rozhraní zvolit pro přehlednost těchto indikátorů. Design prezentace jde výrazně nad rámec tohoto dokumentu.

Nejdříve popisujeme takzvané jednoduché indikátory, které nevyžadují výrazné statistické dovednosti. Tyto indikátory se mohou částečně překrývat s těmi, které již publikuje Ústav pro informace ve vzdělání, nicméně tento překryv je velmi omezený, jak lze poznat z přílohy textu, kde indikátory z ÚIV uvádíme. Následně se soustředíme na některé pokročilé indikátory, které je možné spočítat nebo odhadnout již pouze za pomoci statistických metod.

3.1 Jednoduché indikátory ze stávající matricy studentů

3.1.1 Počty studentů

Následující kapitola představuje základní statistické indikátory popisující počty studentů podle různých charakteristik. Ve všech případech je možné rozlišení indikátorů i podle pohlaví, což již dále nezmiňujeme. Dále považujeme za vhodné, aby se vždy tyto indikátory prezentovaly za co nejdější období, aby uživatelé měli představu o dlouhodobém vývoji. V této kapitole nejvíce navazujeme na informace prezentované ÚIV. Naše návrhy indikátorů jsou obvykle myšleny tak, že budou prezentovány za jednu školu.

3.1.1.1 Počty studentů rozdělené na geografické jednotky

Například počet studentů na určité vysoké škole v kraji a okrese. Pro potřeby výzkumu může být důležitý agregovaný počet studentů v daném regionu. Tato informace může vypovídat o koncentraci lidského kapitálu v daném regionu. Případně pomocí konfrontace s údaji o průměrném vzdělání v populaci je možno zjistit, zda absolventi zůstávají v daném regionu nebo odcházejí.

3.1.1.2 Počty studentů po ročnících a programech a školách

Celkový přehled studentů podle ročníku studia, rozděleno na bakalářské, magisterské, navazující magisterské a doktorské studijní programy. Tyto indikátory by bylo možné dále členit po programech (tj. kombinace ročník a program) pro získání přehledu prostupu studentů jejich studiem. Již dnes jsou relativně dobře dostupné (ze zdrojů

MŠMT) informace o počtech studentů v jednotlivých programech. Samozřejmostí jsou i indikátory o počtu studentů na dané škole či její součásti.

3.1.1.3 Počty duplicitních studentů

Zajímavý indikátor představují i počty studentů, kteří aktivně studují dva či více studijních programů. Tyto počty studentů pro danou VŠ, fakultu a program představují jistou informaci o náročnosti a propojenosti různých studií. V indikátorech o neúspěšnosti toto bude provázáno s úspěchy a neúspěchy těchto studentů.

3.1.1.4 Počty absolventů

Počet absolventů jednotlivých oborů, programů, fakult a VŠ za posledních (např.) 5 let. Mohou jednoduše ukázat, které školy mají největší počet studentů, nemůže to být však indikátor, který by něco ukazoval o nadprodukcii studentů.

3.1.1.5 Počty neúspěšných studentů-studií

Za posledních 5 let spočítané průměrné počty neúspěšných studií v jednotlivých ročnících (opět po oborech/programech/fakultách) a to i relativně k celkovým počtům studentům v daných ročnících (viz kapitola 3.1.3 (Ne)úspěchy, sekce o „úmrtnosti“)

3.1.1.6 Počty přerušení

Počet studentů, kteří přerušili studium po ročnících, oborech/programech/fakultách. Jejich velký počet bude naznačovat jisté problémy na dané škole.

3.1.1.7 Počty studentů nad standardní dobou studia

Absolutní i relativní počet studentů, kteří přesahují v daném programu standardní dobu studia. Kromě toho je vhodné spočítat i počet studentů studující déle než je standardní doba studia plus jeden rok, tedy těch, co již mohou mít povinnost platit poplatky. Z hlediska uchazečů jde bezesporu o velmi cennou informaci vztahující se k organizaci studia na dané škole. Velký počet studujících nad standardní dobu bude naznačovat jisté problémy s nastavením studijního programu.

3.1.1.8 Počty přestupů (a změn škol/programů mezi Bc a N.Mgr)

Počet studentů, kteří přestoupili mezi programy či školami. Počet studentů, kteří na dané škole úspěšně dokončili Bc. studium a přešli na studium navazující magisterské. Počet studentů, kteří přišli z jiné školy na navazující Mgr. po úspěšném absolvování studia Bc. na jiné škole. V tomto případě je možné, že velký počet studentů, kteří odcházejí z bakalářského studia a pokračující v magisterském stupni na jiné škole, může ukazovat určité (ne nutně špatné) zaměření školy. Toto zaměření může některým uchazečům vyhovovat a některým nikoli, záleží na jejich studijních plánech.

3.1.1.9 Počty studentů z kraje působnosti školy

Jedním z důležitých ukazatelů kvality VŠ (či programu na ní) může být schopnost přilákat studenty ze vzdálenějších oblastí ČR. To lze postihnout indikátorem, který by využil informací o bydlišti studenta a sídle školy. Indikátor by byl roven podílu studentů, kteří mají trvalé bydliště v kraji působnosti dané VŠ či její části, a počtu všech studentů.

Alternativně, na základě bydliště a sídla školy, je možné spočítat průměrnou dojezdovou vzdálenost studentů na dané VŠ.

3.1.1.10 Počty studentů dle absolvované střední školy

Tento indikátor popisuje jaký typ střední školy absolvovali současní studenti dané VŠ, její fakulty a programu. Vysoký podíl studentů z odborných středních škol může naznačovat techničtější zaměření programu (i na netechnické VŠ) apod. Tento indikátor může být přesněji vypočítán po připojení databáze Uchazeč, jak uvádíme v kapitole 5.1.1

3.1.1.11 Počty mužů a žen

Relativní a absolutní počet mužů a žen na určitém programu nebo škole, může být zajímavý indikátor z několika důvodů. Jednak pro uchazeče o studiu to může být užitečná informace o budoucí očekávané struktuře jejich spolužáků. Z pohledu zkoumání trhu práce je informace o počtech žen a mužů důležitá pro zkoumání mzdových rozdílů. Existující studie totiž ukazují, že mzdové rozdíly mohou začínat již segregací do určitých programů.

3.1.1.12 Počty studentů dle věkových kategorií

Věk studentů se může výrazně lišit mezi jednotlivými programy. Je pravděpodobné, že věk studentů na školách se bude dále prodlužovat s tím, jak starší generace pracovníků se budou vracet do vzdělávacího procesu a doplňovat si vzdělání, které z nějakých důvodů nezískali nebo neměli šanci získat dříve. Je také zřejmé, že některé školy se budou na tyto studenty jako klienty více specializovat a některé nikoliv. Z toho mohou vyplývat i pozorované věkové rozdíly studentů. Tento indikátor je možné i doplnit průměrným věkem, jak je to prezentované německém informačním systému, popsany v šesté kapitole tohoto textu.

3.1.2 Délky studia

V této kapitole diskutujeme různé indikátory vztahující se k délkám úspěšných a neúspěšných studií, přerušení apod.

3.1.2.1 Délky úspěšných studií

Užitečné indikátory jsou i průměrné délky úspěšných studiích pro různé typy studia (Bc, navazující Mgr, Ing, Mgr,Phd) pro různé VŠ, fakulty i o programy. Tyto indikátory je možné spočítat i pro neúspěšná studia, jak se zmíníme v kapitole o indikátorech neúspěchů. Další možný indikátor je i průměrný (absolutní) přesah průměrné délky úspěšného studia nad standardní dobu studia.

3.1.2.2 Relativní délka studia vzhledem k akreditované délce

SIMS obsahuje informaci o standardní době studia (dle akreditace). Užitečný indikátor je i relativní průměrná délka studia, tj. poměr průměrné délky úspěšných studií a standardní doby studia. Tato informace naznačuje, jak obvyklé jsou přesahy standardní doby studia na dané škole nebo programu.

3.1.2.3 Průměrná délka studia na zahraničním pobytu

Informace v SIMS lze použít i k výpočtům průměrné délky zahraničního studijního pobytu. Užitečné mohou být indikátory průměru přes všechny studenty, tak i průměr přes studenty, kteří na nějaký pobyt vyjeli. První číslo, počítané pro absolventy, udává dobu,

kteřou průměrně studenti v zahraničí strávili. Druhé číslo popisuje průměrnou délku pobytu studenta, který vycestoval.

3.1.2.4 Průměrná délka přerušeni

Další indikátor může být průměrná délka přerušeni u absolventů (počítáno přes všechny absolventy za posledních několik let), a podobně jako v předchozím případě i průměrná doba přerušeni u těch, kteří měli během studia své studium přerušili.

3.1.3 (Ne)úspěchy při studiu

3.1.3.1 Průměrná délka neúspěšného studia

Tento indikátor představuje průměrnou délku neúspěšných a zahájených studií. Alternativně je tento indikátor možné počítat jako průměrnou délku všech neúspěšných studií, což ovšem podhodnocuje skutečnou průměrnou délku neúspěšných studií, protože některá nedávno zahájená studia v budoucnu skončí neúspěšně. Doplňuje indikátor o počtu neúspěšných studií uváděný v kapitole 3.1.1.7.

Jde o poměrně významný indikátor pro zájemce o studium, protože jim umožňuje odhadnout, kdy daná škola „vyhazuje“. Kratší hodnota znamená, že většina neúspěšných studií končí brzy a tak jen relativně málo studentů skončí bez diplomu po delším studiu. Indikátor je významný i pro administrativu (jak vysokých škol, tak MŠMT), protože vysoké hodnoty indikátoru znamenají relativně neefektivně vynaložené náklady.

3.1.3.2 „Úmrtnost“ po ročnících

Tento indikátor je o něco podrobnější než předchozí indikátor. Ze studií, u kterých lze očekávat že jich většina byla již ukončena, lze spočítat průměrnou úmrtnost po ročnících, tedy počet studentů, kteří dané studium ukončili v daném ročníku, vztahený k celkovému počtu nastupujících studentů.

Význam je obdobný jako v předchozí kapitole, i když tento indikátor umožňuje lépe odhadnout, které ročníky jsou ve studiu kritické. Tato informace je rovněž potenciálně užitečná pro školy samotné, neboť jim umožňuje lépe (rovnoměrněji) naplánovat zátěž studia. Nicméně, jak ukazujeme v části 3.2 o pokročilých indikátorech, spočítaná

úmrtnost prostým průměrem za minulé roky a ročníky nemusí nutně být totožná s pravděpodobností dokončení, například prvního ročníku. Jak ukazujeme dále, je nutné brát v potaz různé „kohorty“ vstupující do školy a počítat úmrtnost pouze v rámci těchto „kohort“.

3.1.3.3 Poměr počtu neúspěšných a úspěšných studentů

Jednoduchý indikátor shrnující počet úspěšných a neúspěšných studentů (vzhledem k počtu zapsaných), opět z dostatečně „starých“ studií. Toto číslo představuje apriorní odhad šancí uchazeče o studium vybrané studium úspěšně dokončit.

Nelze ale říci, zda vysoké či nízké hodnoty jsou lepší, protože např. nízká úspěšnost může indikovat jak vysokou kvalitu studia, tak přehnanou náročnost či nevyváženost studia a špatný přístup ke studentům. Naopak vysoká úspěšnost může indikovat příliš snadné (a proto nepříliš užitečné) studium.

Pro interpretaci takto jednoduchého indikátoru je však nutné brát v potaz to, že mícháme mnoho generací najednou. Proto pro výpočet skutečně informativního ukazatele o pravděpodobnosti dokončení studia budeme počítat až v pokročilých indikátorech.

3.1.3.4 Přestupnost (pokračování) na N.Mgr.

Indikátor lze definovat jako počet studentů na Bc. stupni, kteří nastoupili do navazujícího magisterského programu.

Tento indikátor informuje primárně studenty o tom, zda vybraný program je vhodný v případě jejich ambic pokračovat po jeho ukončení ve studiu. Nízké hodnoty naznačují, že většina studentů po jeho ukončení dále nepokračuje v navazujícím programu. Tento indikátor má informační hodnotu i pro samotné školy, protože je upozorňuje na potenciální problémy v programech, u kterých se očekává pokračování studentů dále, ale fakticky k němu nedochází. Do určité míry může indikátor rovněž popisovat kvalitu studia – lepší program studenty lépe připraví na následující studium, ale je potřeba k tomuto významu přistupovat opatrně. Některé programy mohou být cíleně prakticky zaměřené a vést absolventa k dobrému vstupu do praxe. V případě propojení databáze s informacemi o tom, kam se student hlásí, by bylo samozřejmě možné přidat indikátor o

ambicích pokračovat (relativní počet studentů, kteří po ukončení daného Bc. programu podají přihlášku na N.Mgr. program) a jejich úspěšnost (původní indikátor).

3.1.3.5 Přestupy mezi školami

Relativní počet studentů, kteří přestoupili z obdobného programu jiné VŠ a relativní počet studentů N.Mgr. studia, kteří mají Bc. titul z jiné VŠ představuje zajímavý signál o kvalitě a prostupnosti daného studia.

Větší počet studentů příchozích „z venku“ obvykle značí vyšší kvalitu (možná motivace studentů pro přestup) a také větší otevřenost, která obvykle vyžaduje zavedení jasných a otevřených pravidel. Například jde i o signál toho, zda přijímací zkoušky na N.Mgr. program představují „malou domů“, tedy testy snadno splnitelné pro domácí studenty, ale nepřekonatelné pro studenty jiných škol. Podobně také jde o signál toho, nakolik je škola flexibilní v uznávání částí studií absolvovaných na jiných VŠ.

3.1.4 Indikátory o jazyku výuky

3.1.4.1 Počet programů (oborů) akreditovaných v cizím jazyce

V rámci akreditací se stanovuje, zda daný předmět je vyučován v cizím jazyce (primárně angličtina). Pro studenty může být zajímavé znát celkový počet programů (případně oborů) na dané VŠ vyučovaných v cizím jazyce, jakožto určitý indikátor kvality. Může také ale představovat informaci o tom, na jaké studenty se daná VŠ primárně soustředí.

3.1.4.2 Počty studentů a absolventů programů (oborů) v cizím jazyce

Hodnotnější indikátor již je počet studentů a absolventů studií v cizím jazyce, protože nedokumentuje jen počet akreditovaných programů, ale také počet studentů v nich studujících, tedy zda jde o marginální, nenavštěvané, či až formálně existující programy, nebo skutečně běžně využívanou nabídku studia.

3.2 Pokročile indikátory ze současné databáze SIMS

V této kapitole diskutujeme indikátory, které lze spočítat na základě informací obsažených v SIMS, ale které vyžadují pokročilejší statistické metody. Z důvodů relativně omezených dostupných socio-demografických charakteristikách tyto indikátory

budou mít relativně omezený rozsah a vypovídací schopnost. V dalších kapitolách a v příloze podrobněji naznačíme, jak by rozšíření SIMS pomohlo k získání přesnějších indikátorů.

Hlavní indikátory v této kapitole se zabývají prostupností studia a jsou proto přirozeně primárně zajímavé pro stávající a potenciální studenty. Pro odbornou veřejnost, ať již ze samotných VŠ, MŠMT či výzkumných institucí, mohou být užitečné pro další detailnější analýzy nastavení podmínek průchodu studiem, obtížnosti jednotlivých programů a efektivity využití finančních prostředků. Tyto indikátory tak mohou například naznačit, kde by bylo vhodné přesunout obtížnější předměty směrem k začátku studia, aby nedocházelo ke zbytečnému a nákladnému prodlužování studia, které pak stejně nebude úspěšné.

3.2.1 Praviděpodobnosti úspěšnosti

Navrhované pokročilé indikátory jsou zaměřené především na pravděpodobnosti postupu studiem. Níže uvádíme určité příklady těchto indikátorů. Jelikož se jedná o pravděpodobnosti, jejich vyjádření by mělo být vždy v procentech, což je zřejmě uživatelsky nejsrozumitelnější. Ve skutečnosti jich je možné navrhnout a spočítat téměř libovolný počet. Domníváme se však, že ty námi uvedené patří k těm nejzajímavějším.

Všechny tyto indikátory lze, za předpokladu dostatečného počtu pozorování, počítat jak pro jednotlivé VŠ, úrovně studia a programy (případně obory). Všechny tyto indikátory představují různý popis téhož – jakým způsobem studenti procházejí studiem a které ročníky představují největší překážky. Indikátory mají význam pro studenty samotné – informace o šancích na dokončení může ovlivnit výběr školy apod. Využijí je ale i samotné VŠ pro optimalizaci zátěže kladené na studenty a prostředků na studenty vynaložených.

Optimální sledovanou jednotkou je kohorta, tedy soubor studentů zapsaných do studia ve stejném roce či semestru. Pokud je cílem výpočet pravděpodobnosti pro studenty daného programu, školy apod., je soubor studentů těmito podmínkami zúžen. Na takto definovaném souboru je pak sledován průchod studiem. Pokud je cílem například spočítat pravděpodobnosti úspěšného ukončení prvního ročníku, je nutné sledovat studenty, kteří

alespoň začnou druhý ročník. Podobně pro postup do vyšších ročníků je nutné vycházet ze souborů studentů nastupujících v dřívějších letech. Pro analýzu pravděpodobnosti úspěšného studia je tak nutné omezit se na studenty, kteří zahájili studium před více než standardní dobou, potenciálně dvojnásobkem standardní doby, aby naprostá většina studií byla již ukončena, ať už úspěšně či neúspěšně.²

Takto získané pravděpodobnosti samozřejmě vypovídají jen o posledním sledovaném ročníku. Pro zpřesnění požadované informace lze stejně postupovat u všech ročníků předcházejících, pro která jsou data dostupná. Výsledkem pak může být zobrazení takto získaných pravděpodobností pro všechny jednotlivé ročníky. Tím je nejen omezen vliv náhodných faktorů omezených na jednotlivé ročníky, ale také je umožněn alespoň základní pohled na trend vývoje. Je rovněž možné spočítat vážený průměr každého indikátoru, přičemž jako váhy je nutné použít počty studentů v každé kohortě.

Při publikování tohoto indikátoru veřejnosti je potřeba upozornit na to, že jde o průměrný ukazatel, a proto nemůže postihnout všechny charakteristiky jednotlivce. Ani hodnoty indikátoru na úrovni 100% neznamenaají, že je dosažení studovaného cíle (postupu do dalšího ročníku, dokončení studia apod.) jisté. A to především proto, že všechna vstupní data logicky vycházejí z minulého vývoje, který nemusí nutně být shodný s vývojem budoucím.

Podobně nelze indikátory jako pravděpodobnost dokončení studia považovat za indikátor kvality. Některé programy se mohou vyznačovat nízkou úspěšností jen proto, že jsou špatně strukturované, pro studenty nezajímavé, či prostě představují „přestupní stanici“ k vysněnému programu. Rovněž je možné, že některé programy mohou mít nízkou pravděpodobnost úspěchu proto, že na nich je umožněno studium velkému počtu studentů (nízká selektivita), zatímco jinde je výběr studentů soustředěn do přijímacích zkoušek a větší část těch, které přijímacím procesem projdou, následně školu i dokončí.

² Možností jak omezit tento problém je několik. Lze například vyloučit všechna studia přesahující standardní dobu studia daného programu o více než dva roky, nebo je považovat za neúspěšná. Případně lze stanovit délku sledovaného období (a tedy stáří posledního sledovaného ročníku) tak, aby méně než je určitý stanovený počet (v procentech) studií nebylo ukončeno (ať již úspěšně či neúspěšně).

Přes všechny tyto nedostatky či omezení spojená s interpretací této skupiny indikátorů, je jejich význam nesporný. Tyto indikátory umožňují poměrně důkladný vhled do procesu studia, dostupnosti a reálných šancí na dokončení.

Později se v tomto textu budeme zabývat tím, jak získat přesnější odhady uvedených pravděpodobností. Nejprve však diskutujeme, čeho lze dosáhnout s využitím existujících informací v SIMS. Později uvádíme, jaké další údaje by bylo potřebné sbírat pro další vylepšení odhadů indikátorů.

3.2.1.1 Pravděpodobnost dokončení studia

Tento indikátor odpovídá na otázku jakou šanci má student dokončit studium při vstupu do programu. Tento indikátor bude asi nejatraktivnější pro zájemce o studium.

3.2.1.2 Podmíněná pravděpodobnost dokončení studia po dokončení daného ročníku

Indikátor vyjadřuje šanci dokončit studium po dokončení určitého ročníku. Je zřejmé, že se tento indikátor bude v praxi zpravidla zvyšovat s vyššími ročníky.

3.2.1.3 Pravděpodobnosti dosažení daného ročníku studia

Tento indikátor by měl ukázat, jakou má přijatý student šanci dosáhnout určitého ročníku. V zásadě jde o vyjádření (nepodmíněné) úmrtnosti studentů během studia.

3.2.1.4 Pravděpodobnosti dokončení ročníku X za předpokladu dokončení ročníku X-1.

V tomto případě jde o vyjádření šance dokončit například 3. ročník, za předpokladu, že student dokončil 2. ročník. Jde v zásadě o nejpresnější vyjádření toho, jak obtížně procházejí studenti celým studiem. Neboli v jakém ročníku jich nejvíce ukončuje studium. Pro 1. ročník bude tento indikátor logicky shodný s indikátorem *Pravděpodobnosti dosažení daného ročníku studia* a pro poslední ročník by měl být teoreticky shodný s indikátorem *Pravděpodobnost dokončení studia po dokončení daného ročníku*, pokud by celá kohorta dokončovala studium ve stejném ročníku, což v praxi nemusí být vždy splněno.

3.2.2 Regresní analýza v současné databázi SIMS

Předcházející kapitola představila několik pokročilejších, ale stále deskriptivních indikátorů. I když je možné například spočítat, jakou pravděpodobnost dokončení mají studenti na různých školách, programech či úrovních studia, tyto indikátory neumožňují vysvětlit, jak a čím jsou tyto pravděpodobnosti určeny. Například nízká pravděpodobnost dokončení může být způsobena jak jeho náročností, tak kvalitou přihlášených studentů. Při popisném pohledu tyto rozdíly není možné adekvátně postihnout.

Ideálním cílem je proto získat dostatek informací k tomu, aby rozdíly v pravděpodobnostech dokončení bylo možné vysvětlit.

Metodou, které toho lze dosáhnout, je statistická regrese pomocí probit nebo logit modelů. Ta si klade za cíl zjistit roli různých veličin (vysvětlujících proměnných) na určitou binární proměnnou (jako je dokončení či nedokončení studia). Řadu možných vysvětlujících proměnných, stejně jako podrobnější vysvětlení fungování této metody, poskytujeme v příloze k tomuto textu.

Cílem je získat model, který umožní odhadnout pravděpodobnost například dokončení studia na základě relevantních charakteristik studentů. Mezi ty patří jak individuální socio-demografické charakteristiky, tak ale i některé obecné charakteristiky jako je například dosavadní vzdělání (lze aproximovat například typem střední školy). Model poté pro vybranou skupinu studentů (například studenti dané univerzity, daného programu apod.) spočítá vliv oněch charakteristik (vysvětlujících proměnných) na studovaný indikátor (např. pravděpodobnost dokončení studia). Vliv různých charakteristik je odražen v odhadnutých koeficientech. Tyto koeficienty jsou samy o sobě zajímavé pro odbornou veřejnost – umožňují například studovat vliv prostředí, rodiny, dosavadního vzdělání na další studijní úspěchy.

Výsledky lze ale použít i pro získání individuálních odhadů studovaného indikátoru. Model umožňuje po dosažení konkrétních hodnot za dané charakteristiky (tedy typ střední školy, vzdělání rodičů atd.) spočítat hodnotu indikátoru – například pravděpodobnosti dokončení studia v daném programu či dané VŠ.

Musíme ale v tuto chvíli upozornit, a v příloze diskutujeme tento problém podrobněji, že SIMS podle našeho názoru zatím neobsahuje dostatečné množství údajů na to, aby bylo možné tento odhad udělat dostatečně přesně. Pro dobrý odhad je nutné znát dostatečné množství relevantních individuálních charakteristik studenta, z nichž některé v SIMS nejsou. Proto nyní budeme diskutovat možnosti rozšíření SIMS a uvádíme, jaké indikátory by určité rozšíření přineslo.

4 Možnosti rozšíření SIMS

SIMS v současné době neobsahuje řadu charakteristik studentů a programů, které mohou být podstatné pro výpočet indikátorů a statistik užitečných pro rozhodování uchazečů, ale také pro socioekonomický výzkum zaměřený na školství a úspěšnost jednotlivců na vysokých školách. V této sekci některé tyto charakteristiky navrhuje a u každé charakteristiky popisujeme její význam. V principu jde o informace, které škola může relativně snadno sama sbírat buď jednoduchým dotazníkovým šetřením, nebo z vlastních informačních databází. Jedná se o základní socioekonomické charakteristiky a studijní výsledky na dané škole. Dále v této kapitole popisujeme možnost připojení ke každému studentovi další informace o programu, který student studuje. Jak rozebíráme níže, potenciálně by to značně usnadnilo odhadnutí vztahu mezi úspěšností studentů a finanční nákladností programů. V první části této kapitoly se zabýváme především možnostmi rozšíření SIMS o další charakteristiky studentů. V druhé části se zabýváme charakteristikami programu.

4.1 Charakteristiky studenta

4.1.1 Socioekonomické charakteristiky

4.1.1.1 Vzdělání a příjem rodičů

Rodinné zázemí je vždy důležitý faktor, který ovlivňuje ať již rozhodnutí o samotném studiu, ale také úspěšnost dokončení. Pokud budeme mít bližší představu o jeho podobě, budeme moci například statisticky zjistit, na jakých vysokých škol má rodinné zázemí větší vliv pro úspěšnost dokončení.

Obě navrhované charakteristiky - vzdělání a příjem rodičů jsou podstatné a mohou hrát velkou roli při studiu. Vzdělání rodičů je obvykle relativně jednodušší zjišťovat, jelikož respondenti obvykle pravdivě odpovídají a málo chybují. U zjišťování příjmu rodičů je však potřeba postupovat obezřetněji, jelikož je tento údaj obvykle považován za velmi citlivý. V oblasti zjišťování příjmů se proto se obvykle postupuje tak, že se určí intervaly výše příjmů a respondent se „zařadí“ do příslušného intervalu. Minimalizuje se tím problém neochoty odpovídat na daný dotaz a respondenti mají relativně menší tendenci zkreslovat odpovědi.

4.1.1.2 Sourozenci

Počet sourozenců, případně jejich věk a vzdělání (i nedokončené) je též potenciálně důležitá informace. Vzhledem k možnému rozpočtovému omezení v rodině, větší počet sourozenců může být bariérou pro podporu rodiny ve studiu potomků. Pokud má jednotlivec staršího sourozence, znalost informace o jeho sociálně ekonomické pozici může například dopomoci ke zjištění, zda sourozenci mohou ovlivnit například výběr programu, motivaci dokončit studium a podobně.

4.1.1.3 Místo pobytu a dojíždění

Vzdálenost vysoké školy od domova může být důležitý faktor v rozhodování při přihlašování na vysokou školu. V současné době se v SIMS sleduje adresa trvalého bydliště, který mimo jiné hraje roli v přidělování ubytování v rámci ubytovacích zařízení na vysokých školách. Tento údaj díky tomu ovšem nemusí nutně odpovídat skutečnosti. Proto navrhuje, aby ve zvláštním dotazníkovém šetření bylo zjištěno skutečné místo pobytu studentů na vysokou školu v okamžiku podávání přihlášky a dále v průběhu studií.

4.1.1.4 Příjem a pracovní činnost

Stále více studentů při studiu pracuje, ať již na dočasných brigádách, částečných pracovních úvazcích, nebo na praxích. Tato činnost může výrazným způsobem ovlivnit jejich pravděpodobnost dostudování a to jak v pozitivním, tak negativním smyslu. Bylo by vhodné, aby se zjišťovalo zda student pracuje a pokud ano, kolik hodin v týdnu věnuje práci. Následně kolik přibližně vydělává a jaký podíl z jeho celkových příjmů (včetně

kapesného) tvoří tyto příjmy. Příjmy, podobně jako v případě rodičů lze sledovat v intervalovém vyjádření.

Dotazníkové šetření u celé populace studentů však může být technicky a administrativně náročné. Proto si myslíme, že by bylo vhodnější dělat toto šetření v delších než ročních intervalech a na náhodném menším vzorku populace studentů na dané škole.

Pokud pracující studenti odvádějí sociální a zdravotní pojištění, mohla by být tato informace zjištěna z externí databáze, jak podrobně rozebíráme v páté části.

4.1.2 Studijní výsledky a zaplacené poplatky

Základním výstupem SIMS je úspěšnost absolvování ročníku studia a následně celého studia. To samo však může být příliš hrubý (i když nejpodstatnější) ukazatel úspěšnosti studenta. Pro zpřesnění informace o postupu studiem proto navrhuje, aby se k základním údajům připojily i získané kredity za daný rok, včetně srovnání s požadovaným počtem kreditů. Případně se tento údaj může doplnit o průměrný studijní prospěch a zda plní požadavky rovnoměrně během studia nebo převládá dohánění předmětů na poslední chvíli a podobně. Nedostatkem indikátoru však bude, že ne všechny kredity škola započítává do objemu požadovaných kreditů.

Poplatky zaplacené během studia mohou představovat důležitý údaj, který doplní další informace, například o době studia nad rámec standardní doby studia a náklady na studenta v programu. V tomto případě předpokládáme, že jednotlivé školy doplní tento údaj z vlastních informačních zdrojů. Sledování poplatků za studium dostane jiný rozměr v případě, že bude zavedeno školné, ať již v odložené nebo jiné podobě. Informační systém bude nutné dle toho upravit a počet informací nutných ke sledování se jistě bude muset značně rozšířit.

4.2 Charakteristiky programu

V této kapitole se budeme zabývat několika indikátory, které se dotýkají přímo daných programů a nákladů, které stát na ně vydává. Většina srovnání bude užitečná zejména mezi stejnými programy různých škol neboť náklady na různé programy se značně liší

(jak je odraženo v normativních koeficientech) a srovnání nákladů mezi programy by tak bylo značně zavádějící.

4.2.1 Předměty a jejich obsah

Součástí charakteristiky studijního programu jsou i předměty, které je nutné absolvovat pro úspěšné ukončení programu. Mohlo by být proto užitečné sledovat jaké předměty jsou nejobtížnější pro studenty, a které musí nejčastěji opakovat. Navíc by se mohlo uvažovat o propojení se stručným obsahem jednotlivých předmětů, přestože tento cíl by jistě nebyl technicky snadno realizovatelný. Z hlediska uživatelů by však tento informační systém byl jistě velmi vítaný a značně by snížila asymetrii informací mezi uchazečem o studiu a vysokou školou jako poskytovatele vzdělání.

4.2.2 Převis poptávky po daném programu

Jako užitečný údaj by mohl sloužit poměr přijatých a nepřijatých studentů. Tato charakteristika do značné míry ukáže, jak náročné je pro uchazeče se na daný program dostat. V současné době je možné tuto informaci již poměrně spolehlivě získat z různých zdrojů, takže její připojení k ostatním by neměl být problém.

4.2.3 Koeficient programu

V současné době nejsou informace o nákladových koeficientech programů v databázi SIMS evidovány. I když nejde o informace zásadní, jsou velmi důležité pro následující indikátory. Propojení existující databáze s databází normativních koeficientů je však poměrně snadné a bezproblémové, a proto jej uvádíme již v této části.

4.2.4 Náklady na absolventa

Spolu s normativním nákladovým koeficientem programu umožňuje databáze SIMS spočítat náklady na jednoho absolventa. Výpočet musí vycházet z již ukončených studií (ať již úspěšně či neúspěšně), které začaly dostatečně dávno. Vynásobením počtu vystudovaných semestrů či ročníků normativním koeficientem a základním normativem lze získat celkové náklady. Po dělení počtem úspěšných absolventů lze získat průměrné náklady na absolventa.

Při srovnání stejných programů toto číslo efektivně shrnuje to, kolik peněz daná škola v daném programu-oboru vynaloží na studenta, který studium nedokončí. Významně vyšší hodnoty průměrných nákladů tak indikují jisté plýtvání ve srovnání s ostatními školami. Jde o indikátor významný jako pro management škol (upozorňuje na potenciální problémy v organizaci studia), tak pro MŠMT, které může z rozdílů hodnot vyvodit informaci o efektivitě vynaložených prostředků.

Je ale potřeba upozornit na to, že nízké průměrné náklady nemusejí nutně odrážet kvalitní studium, ale naopak „rozdávání“ diplomů bez podstatných požadavků na studenty. Příliš vysoké číslo ale naznačuje, že daná škola v daném programu neprovádí odpovídající selekci studentů, nabírá a delší dobu nechává studovat studenty, kteří následně dané studium nedokončí.

4.2.5 Zdroje financování

Tento indikátor nelze ze současných dat v SIMS počítat. Do budoucna by ale informace o finančních zdrojích daných vysokých škol a jejich součástí mohli představovat užitečný indikátor kvality. Školy schopné diverzifikovat příjmy mezi platby státu, školné a jiné poplatky, příjmy ze sponzorství, vědy, patentů a vedlejší činnosti, obvykle představují školy s kvalitnějšími pedagogy a také vyšší kvalitou studia, zejména na úrovni magisterských a doktorských programů. Školy získávající své příjmy převážně z dotací a příspěvků státu a poplatků od studentů mohou indikovat „masové“ programy, kde vyučující nejsou schopni věnovat každému studentovi potřebnou pozornost.

4.3 Indikátory z rozšířené databáze SIMS

Podobně jako v třetí kapitole i z dodatečných charakteristik se dají počítat různé deskriptivní indikátory základního a pokročilejšího typu.

4.3.1 Jednoduché indikátory

V této kapitole v zásadě navazujeme na indikátory navržené v kapitole 3.1. Z rozšířené databáze lze samozřejmě spočítat jak ty předchozí, tak i nové, které navrhuje níže a které vycházejí z dodatečných charakteristik přidanych k základní verzi SIMS.

4.3.1.1 Socioekonomické charakteristiky

Podle jednotlivých charakteristik, které se podaří sesbírat lze sestavit základní sadu deskriptivních indikátorů.

4.3.1.1.1 Průměrné vzdělání a příjem rodičů

Tento indikátor spočítaný pro jednotlivé vysoké školy a případně programy by ukázal, zda dochází k výrazné segregaci studentů s různým sociálním zázemím na různé školy a programy. Pokud by v budoucnu mělo dojít k zavedení školného dle předem definovaného klíče, tento indikátor by mohl napovědět, zda se bude výrazně díky tomu bude měnit socioekonomická struktura studentů.

4.3.1.1.2 Průměrný počet sourozenců a jejich vzdělání

Podobně jako v případě rodičů jde o doplnění informací o tom, jací studenti s rodinným zázemím se na různých školách vyskytují.

4.3.1.1.3 Průměrná vzdálenost dojíždění

Skutečné místo pobytu a průměrný čas dojíždění do školy by mohla být zajímavá informace z mnoha důvodů. Především by se tím daly odhadnout skutečné náklady na dojíždění, což nelze zjistit přesně z údaje o trvalém bydlišti. Dodatečným anonymním dotazníkem by mohlo být zjištěno, jaká je skutečná dojezdová vzdálenost každý den do školy na výuku, případně jaké jsou s ní spojeny náklady.

4.3.2 Pokročilé indikátory z rozšířené SIMS – regresní analýza

Jak jsme již uvedli v kapitole 3.2.2 databáze SIMS potřebuje rozšíření, aby s ní mohla být realizována základní regresní analýza. Regresní analýza je fakticky sofistikovanější verze korelační tabulky s více proměnnými. Jelikož jsme v předchozí kapitole již navrhli možné charakteristiky, bylo by možné s nimi pracovat i v níže popsané v regresní analýze.

Základní model by mohl například vysvětlovat proměnnou, zda student dokončil studium v závislosti na jeho demografických charakteristikách, jako jsou věk, pohlaví, místo trvalého bydliště, vzdělání a příjem rodičů, počet sourozenců a například typ vystudované vysoké školy. Potenciálně lze také testovat, zda na dokončení studenta mají vliv náklady

na studenta. Pro jednoduchost pracujeme pouze s charakteristikami studenty. Přestože detaily regresní analýzy probíráme komplexně a s technickými detaily v příloze tohoto textu, uvádíme již zde podrobnější intuitivní popis, aby bylo zřejmé k čemu je dobré rozšířit stávající databázi SIMS.

V kapitole 3.2.1 jsme představili indikátory průměrné pravděpodobnosti úspěšnosti studia, například dokončení studia na určitém programu. Obecně však můžeme říci, že pravděpodobnost úspěšnosti je ovlivněna mnoha faktory, které mohou být různé na různých vysokých školách. Pokud bychom si dovolili zjednodušení, tak rozdělíme tyto charakteristiky na dva typy. První jsou spojené se studenty na těchto školách (jejich rodinné zázemí, studijní dispozice, ekonomické zázemí atd.). Jinými slovy každá vysoká škola může lákat jiné typy studentů, mít jiný systém přijímacích zkoušek, a to vše může ovlivnit pozorovatelné rozdíly v pravděpodobnosti dokončení daného studijního programu. Druhý typ je spojen čistě s konkrétní školou, její politikou v oblasti organizaci výuky, náročností pedagogů, schopnost pracovat se studenty a podobně. Právě regresní analýza s využitím individuálních charakteristik studenta nám částečně dovolí očistit vliv dané školy od vlivu individuálních charakteristik. Navíc nám ukáže (s jistou mírou přesnosti) jak konkrétní individuální charakteristiky (například vzdělání a příjem rodičů) ovlivňují pravděpodobnost dokončení studia.

Aby měla taková analýza smysl, musí se porovnávat studenti v podobných programech. Konkrétně si představme, že analýzu provádíme pro bakalářské studijní programy zaměřené na chemii a chemické inženýrství. Náš jednoduchý model bude nyní vypadat následovně:

$$(1) Y_{i,s}^* = \alpha_0 + S_s \alpha_1 + X_i \alpha_2 + \varepsilon_{i,s}$$

V této rovnici je několik proměnných. Vysvětlovaná proměnná $Y_{i,s}^*$ vyjadřuje, s jakou pravděpodobností jednotlivec i dostuduje vysokou školu s . Ve skutečnosti do odhadu vstupuje tato proměnná ve formě tzv. binární proměnné, která nabývá hodnoty nula nebo jedna (dokončil(a), nedokončil(a) tak, jak to v datech pozorujeme), jak je vysvětleno

podrobněji v příloze tohoto textu, kde je též technicky vysvětleno jak odhad bude probíhat.

Dále tato rovnice obsahuje vysvětlující proměnné S_s a X_i . První z nich je vektor binárních proměnných, nabývající hodnot nula nebo jedna, pokud student je identifikován na škole s . V praxi to znamená to, že ke studentovi je přiřazená právě jedna škola a vliv této školy na pravděpodobnost dokončení studia je odhadnut koeficienty α_1 . Vždy se určí jedna škola jako referenční a pro každou další školu je odhadnut koeficient, který ukazuje jak ta konkrétní škola ovlivňuje úspěšnost například dokončení studia ve srovnání s referenční školou. Poslední vektor proměnných je X_i , což jsou charakteristiky studentů. V našem případě to jsou právě zmíněné demografické a socioekonomické charakteristiky, o které navrhujeme databázi SIMS rozšířit (například vzdělání a příjem rodičů). Ke každé charakteristice odhadneme její vliv, který je vyjádřen koeficientem α_2 . Těchto koeficientů je přesně tolik, kolik je charakteristik.

Důležitý element tohoto modelu je člen ε_i , který má náhodný charakter a je nepozorovatelný. Tento člen zahrnuje nepozorovatelné charakteristiky studentů a školy. Pro jednoduchost o něm budeme zatím předpokládat, že má vlastnosti, které nevytvářejí výchylku našich odhadů. Jak ukazujeme v příloze tohoto textu, tento předpoklad je velmi silný, zejména pokud máme k dispozici velmi omezený počet vysvětlujících proměnných X_i .

Vraťme se nyní interpretaci výsledků. Jestliže jsme tento model odhadli na relevantních datech, které obsahují dostatek pozorování (studentů) pro každou školu, získali jsme koeficienty α_0 , α_1 a α_2 , které je nutné správně interpretovat. Podstatné jsou především poslední dva vektory koeficientů. Koeficienty α_1 , které jsou odhadnuty za jednotlivé školy, ukazují jak velký má v průměru daná škola vliv na dostudování studentů ve srovnání s referenční školou, bez ohledu na vliv ostatních kontrolovaných charakteristik X_i . Tento indikátor v principu ukazuje relativní pravděpodobnost dokončení studia, očištěnou o do modelu zahrnuté charakteristiky studentů.

Podobně koeficienty α_2 , které jsou odhadnuty pro každou charakteristiku studentů zvlášť, měří vliv těchto charakteristik na dostudování studentů, bez ohledu na to, na jaké studují škole.

Výhoda této vícenásobné regrese tkví v tom, že dokáže očistí vliv určitého faktoru na dokončení studia od jiných, pozorovatelných faktorů. V tomto příkladě jsme oddělili vliv dané školy od jiných demografických faktorů, které mohou mít též vliv na dostudování studentů.

Na tomto jednoduchém příkladu lze ukázat jak je nutné při analýze tohoto typu obecně postupovat. Za prvé je nutné si ujasnit pro koho se dané odhady provádějí, tj. v našem případě pro ty, kteří chtějí studovat chemii a rozhodují se na jaké škole jít nebo dát přihlášku. Musíme tedy do analýzy zahrnout porovnatelné programy, z čehož též vyplývá i relevantní výběr studentů z databáze. Za druhé je potřeba si ujasnit co vlastně chceme vysvětlit. Nejdříve je nutné určit vysvětlovanou proměnnou: tj. zda nás zajímá dokončení studia, dokončení prvního ročníku, délka studia a podobně. Poté je nutné přemýšlet o vhodných vysvětlujících proměnných, které dávají v kontextu vysvětlované proměnné smysl a které nemohou například způsobit opačnou kauzalitu, která výsledky znehodnotí (viz diskuze v příloze). Za třetí je nutné si ujasnit jaké charakteristiky studentů (případně škol) jsou v datech k dispozici a tomu i přizpůsobit odhad a nakonec interpretaci výsledků.

Z hlediska interpretace odhadnutých výsledků je nutné též brát v potaz, že ekonometrická teorie věnuje speciální pozornost k potenciálním nepozorovatelným charakteristikám, které mohou ovlivňovat jak pravděpodobnost dokončení studia (v tomto případě), tak některé pozorovatelné charakteristiky studentů a nakonec i to, na jaké škole jednotlivce pozorujeme (tzv. problém selekce studentů do programů). Technické vysvětlení tohoto jevu je uvedeno v příloze. Eliminace tohoto problému jde však nad rámce běžné analýzy tohoto typu, a proto doporučujeme tento jev zahrnout pouze do diskuse výsledků.

4.3.2.1 Návrh na prezentaci výsledků regresní analýzy

Pokud se podaří získat určitý dostatečný počet charakteristik (jako přibližný minimální počet může být pět), které by dovolily udělat regresní analýzu, je nutné zajistit, aby

prezentace výsledků byla dostatečně srozumitelná pro širší veřejnost. Zejména jde o to, aby uživatelé měli možnost srovnat hrubé pravděpodobnosti dokončení studia prezentované 3.2.1.1 s pravděpodobnostmi dokončení, která je očištěna o charakteristiky přijatých studentů. Jak jsme uvedli v předchozí kapitole výsledek regresní analýzy, výsledek regrese právě dopomůže oddělit vliv školy od vlivu individuálních charakteristik. Jak tuto očištěnou pravděpodobnost spočítat? Odhadneme model (1) a získáme koeficient α_0 , α_1 a α_2 . Díky těmto koeficientům získáme pro každého studenta odhadnutou pravděpodobnost $\hat{Y}_{i,s}^*$, že dokončí studium.

$$(2) \hat{Y}_{i,s}^* = \hat{\alpha}_0 + S_s \hat{\alpha}_1 + X_i \hat{\alpha}_2.$$

Abychom však spočítali průměrnou pravděpodobnost dokončení studia, za každou vysokou školu, bez ohledu na charakteristiky studentů, musíme postupovat následovně. Určíme průměrného studenta s určitými charakteristikami X (například s určitým výsledkem z maturitní zkoušky, z určitého typu školy s určitým vzděláním rodičů) a tyto charakteristiky vložíme do modelu 2 a díky odhadnutým koeficientům získáme pravděpodobnost dokončení studia na každé škole (do proměnné S vložíme vždy binární proměnnou za dotyčnou školu a použijeme příslušný koeficient). Díky tomuto výpočtu dostaneme pravděpodobnost dokončení za každou školu pro studenta s danými charakteristikami, které jsme po odhadu použili jako hodnoty pro X .

Nyní k samotné prezentaci výsledků. Uživatel by měl mít vedle sebe minimálně dva výsledky pro jednu školu/program. Hrubou pravděpodobnost dokončení, vypočítanou v kapitole 3.2.1.1 a potom pravděpodobnost dokončení, která bude odhadnuta pro určitého studenta s průměrnými charakteristikami za všechny školy. Pokud takto porovná více škol pro daný program, měl by dostat představu o tom, do jaké míry je hrubá pravděpodobnost dokončení dána obtížností školy a do jaké pouze selekcí (ne)kvalitních studentů do dané školy.

Prezentace pro dvě školy by konkrétně mohla vypadat podle tabulky 1.

Tabulka 1: Pravděpodobnosti dokončení a přijetí do bakalářského studia ve studijním programu chemie a chemické inženýrství (průměr (například) za posledních pět let)

Vysoká škola	Hrubá pravděpodobnost	Pravděpodobnost pro průměrného studenta chemie v ČR*	Pravděpodobnost přijetí do programu
Univerzita Karlova	63%	43%	35%
VŠCHT	32%	60%	58%

* Popis charakteristik průměrného studenta chemie v ČR. Uvedené hodnoty jsou odhady, které vycházejí z minulých hodnot a nezaručují budoucí vývoj

Fiktivní čísla, která jsou v této tabulce ukazují, že na UK je hrubá pravděpodobnost dokončení studia výrazně vyšší než na VŠCHT. Vyšší pravděpodobnost dokončení studia v případě UK může být dána například vyšší selekcí studentů při přijímacích pohovorech než na VŠCHT. To je též reflektováno v nižší pravděpodobnosti přijetí v případě UK (ne vždy však tomu tak musí být, některé školy jednoduše lákají k přihlášení v průměru kvalitnější uchazeče a ti méně kvalitní chodí automaticky jinam). Průměrný student přijatý na UK je proto kvalitnější než ten na VŠCHT už při vstupu do programu. To je též vidět ve druhém sloupci, kde uvádíme pravděpodobnost dokončení pro virtuálně stejného studenta. Pokud má student průměrné charakteristiky, pravděpodobnost dokončení je již vyšší na VŠCHT než na UK. Tento údaj se tedy blíží k vyjádření obtížnosti daného programu.

Pokud bychom tento systém chtěli ještě zdokonalit, mohli bychom nechat uživatele navolit si přímo vlastní charakteristiky do systému a ten by mu mohl následně ukázat odhad dokončení studia pro jeho kombinaci charakteristik na jednotlivých vysokých školách, které daný program nabízejí. Tento systém by byl ovšem mnohem technicky náročnější, ale přesto proveditelný.

5 Návrhy na propojení SIMS

Rozšíření SIMS by mělo popisovat především to, odkud na vysoké školy jednotlivci přicházejí a samozřejmě jak se následně uplatňují na trhu práce. Klíčový identifikační údaj by mělo být rodné číslo studentů, i přes všechny problémy spojené s ochranou osobních údajů.

Pokud bychom měli začít podle časové posloupnosti, tak mělo by dojít k propojení s databází studentů středních škol a to především maturitních programů. Zjistilo by se z jakých středních škol odcházejí studenti na konkrétní vysoké školy a jak jsou následně úspěšní pokud jsou přijati. Tím by se značně pomohlo i dětem a rodičům, kteří se rozhodují o konkrétní střední škole. Dále je však nutné zdůraznit, že stále více vysokoškoláků bude přicházet na školu z praxe s cílem doplnit si již dosažené vzdělání.

Uplatňování absolventů na trhu práce je však důležitá informace v rozhodování jednotlivce o investování času a prostředků do vysokoškolského vzdělání. V současné době již existují statistiky o nezaměstnanosti absolventů jednotlivých škol a fakult. To ovšem není plnohodnotný údaj o uplatnění absolventů.

Nejzásadnější informací je totiž cena, za kterou absolvent prodává na trhu práce znalosti získané na vysoké škole. Přiřadit mzdu a její vývoj v čase ke každému absolventovi by měl být hlavní cíl rozšíření SIMS o externí zdroje. Naplnění tohoto cíle v praktické rovině bude samozřejmě velmi náročné díky omezeným datovým zdrojům, které by měly identifikátor –tj. rodné číslo – a mzdy za jednotlivce.

Význam mezd jako indikátoru úspěšnosti absolventů je však natolik podstatné, že by se k tomuto cíli mělo dlouhodobě směřovat. Pro uchazeče o studiu může být vodítkem o tom jaký program studovat, na jaké škole a jak dlouho. Lze předpokládat, že se zavedením školního důležitost mezd absolventů značně stoupne.

V následujících podkapitolách se věnujeme jednotlivým indikátorům a statistikám, které by mohly být prezentovány.

5.1 Předchozí studium

5.1.1 Propojení s databází Uchazeč

V současné době by bylo možné připojit k SIMS databázi Uchazeč, která obsahuje rodné číslo, podle kterého by bylo možné propojit jednotlivce mezi těmito databázemi. Databáze uchazeč obsahuje informace o podaných přihláškách na vysokou školu a částečně o předchozím studiu na střední školu. Úplný přehled o údajích v databázi

Uchazeč je příloze P3. Hlavní obohacení SIMS by v současné době bylo v tom, že předchozí střední škola zatím není povinně zaznamenávaná a databáze Uchazeč ji spolehlivě obsahuje. Dále by bylo zajímavé sledovat počet přihlášek a úspěšných přijetí uchazečů ve vztahu k jejich úspěšnosti studia. Díky této databázi by se též dali spolehlivě rozpoznat uchazeči, kteří vstupují na VŠ hned po maturitě nebo s nějakou přestávkou. Následně by se dal pozorovat vztah mezi touto charakteristikou a úspěšností dokončení. Celkově lze říci, že propojení s touto databází by bylo v současné době nejjednodušší.

5.1.2 Výsledky maturit a jiných testů

Obecné metody měření kvality VŠ vyžadují měření kvality nastupujících studentů, a to proto, aby bylo možné určit přidanou hodnotu studia na dané VŠ. Tedy odlišit lepší výsledky studentů způsobené výběrovostí školy a její schopností naučit. I když SIMS nesleduje a ani nebude zřejmě sledovat všechny nezbytné údaje pro podrobnou analýzu, je minimálně užitečné sledovat známky přicházejících studentů a to jak již běžné známky dnes sledované (bohužel těžko srovnatelné pro rozdílnou obtížnost středních škol), tak potenciálně výsledky státních maturit.

Lepší výsledky studentů ze střední školy přijatých na danou VŠ (oboru, programu) by indikovaly, že jde o výběrovější a náročnější VŠ (obor, program). Uchazeč o studium může této informace využít k lepšímu vyhodnocení, zda na daný obor má, či zda jde o obor pro něj dostatečně náročný. Zprostředkovaně lze z tohoto indikátoru i odvodit, o které obory je největší zájem. Srovnání indikátoru pro stejné obory a různé VŠ odráží zájem studentů, tedy jejich individuální informaci (očekávání) o kvalitě spolužáků na vysoké škole.

5.1.3 Výsledky z předchozího studia na vysoké škole

Tato informace by byla samozřejmě relevantní pouze tehdy pokud, student studuje již další studijní program. Například by mohlo jít o výsledky z bakalářského studia, pokud je student na studiu magisterském.

5.2 Úspěšnost absolventa na trhu práce – mzdy

5.2.1 Možné datové zdroje - diskuse

Hlavní problém s připojením mzdy jako další informace o absolventovi dané vysoké školy je v získání dostupného zdroje dat a jeho technické připojení k databázi SIMS. Jeden z možných a nejpravděpodobněji získatelných zdrojů jsou data z České správy sociálního zabezpečení. Zde je sledována prakticky celá populace jednotlivců v České republice a jejich úspěšnost na trhu práce. Teoreticky vzato data z tohoto zdroje by postačovala k získání informací o úspěšnosti absolventů na trhu práce. Nicméně k tomu, aby to bylo realizovatelné, musela být ČSSZ na to vnitřně připravena a mít vlastní data v takové formě, která by umožňovala propojení s jinou externí databází pomocí identifikátoru rodného čísla. Databáze ČSSZ v tomto stavu nyní podle dostupných informací není, nicméně probíhá studie proveditelnosti, která byla zadána společnosti Trexima a která by mohla naznačit jakým způsobem by bylo možné přetvořit současné informační zdroje do podoby pro externí využití. Další administrativní zdroj informací o mzdách absolventů, jsou daňové záznamy z finančních úřadů. Přestože tento zdroj je teoreticky možný, zatím pro to nejsou připravené podmínky.

Poslední, relativně dostupný, ale poměrně nákladný zdroj informací o mzdách je vlastní dotazníkový průzkum o mzdách. Tento způsob má však několik nevýhod a důležitých předpokladů, za kterých by to bylo možné realizovat. Za prvé je nutné mít kontakt na všechny absolventy, případně na jejich reprezentativní výběr. Za druhé je tento způsob poměrně nákladný na lidské a finanční zdroje.

Celkově lze říci, že momentálně není možné říci z jakého zdroje bude možné v budoucnu čerpat informace o mzdách na trhu práce. Proto od tohoto bodu popisujeme jejich využití pouze obecně a nezaobíráme se technickými detaily.

Indikátory mezd je možné specifikovat v různých dimenzích (například časové a geografické) s různou podrobností z hlediska charakteristik vystudovaných programů. Následující kapitoly poukazují na základní členění, které podle našeho názoru může být potenciálně nejzajímavější. Interpretace těchto statistik však podléhá jistým omezením,

které budeme vždy zdůrazňovat. Nejdříve prezentujeme navrhované statistiky mezd pro čerstvé absolventy, čímž máme na mysli mzdy do jednoho roku od absolvování daného programu, přitom je též podstatné, aby se zdůrazňovaly rozdíly mezi muži a ženami.

5.2.2 Průměrná mzda čerstvých absolventů podle programů a škol

Průměrné, nástupní, mzdy čerstvých absolventů by byl zřejmě nejžádanější indikátor. V nejpodrobnějším členění může spočítán na úrovni škol a programů, za předpokladu dostatečného počtu pozorování (absolventů) pro danou školu a program na ní.

Vypovídací schopnost této statistiky je značně závislá na časovém období, za které byla spočítána. Pokud například ekonomiku postihne recese, v období nástupu absolventa na trh práce, minulé časové řady nemusí nutně odpovídat nastalé situaci na trhu práce. Dále absolventi mohou nastupovat do zaměstnání v různých regionech, v kterých je odlišná poptávka a nabídka pro dané absolventy. Tomu se věnujeme v další kapitole.

5.2.3 Průměrná mzda čerstvých absolventů podle regionů

V České republice existují značné regionální rozdíly ve mzdách na trhu práce. Proto může být výrazně podstatné, kde se daný absolvent uplatňuje. Pokud se například potenciální student rozhoduje zda studovat medicínu, může pro něj být podstatné zda „musí“ pracovat v Praze, aby dostal dostatečné ohodnocení nebo v jakémkoli regionu České republiky. Je dost pravděpodobné, že regionálního členění mezd nebude možné dosáhnout podrobné členění na programy v rámci vysokých škol. Zde proto doporučujeme členění čistě na programy (bez ohledu na školy), případně pouze na jednotlivé školy respektive fakulty.

5.2.4 Vývoj mezd absolventů během kariéry

Statistiky mezd čerstvých absolventů samozřejmě nedávají zdaleka ucelený obraz o uplatnění absolventa na trhu práce. To se týká především případného srovnávání mezi jednotlivými programy. Je například zřejmé, že relativní mzdy čerstvých absolventů medicíny oproti doktorům s desetiletou praxí mohou být jiné od relativních mezd čerstvých absolventů ekonomických programů. Jinými slovy zkušenosti u různých programů a profesí jsou na trhu práce oceňováni jinak a v jiné míře. Z tohoto pohledu je

nutné prezentovat průměrné mzdy v začátcích a pozdějších letech kariéry pro jednotlivé programy zvlášť, případně zdůrazňovat vazbu mezi růstem, volatilitou a úrovní mezd v čase.

5.2.5 Mzdy přijatých uchazečů

Stále větší část uchazečů o studium na vysoké školy přichází ze zaměstnání. Jejich motivace mohou být různé. Většinou si přejí zvýšit kvalifikaci, aby jim usnadnila další kariérní postup. Z tohoto pohledu bude jistě zajímavé pozorovat, jaké mzdy měli tito uchazeči před studiem na vysoké škole a po absolvování. V tomto případě bude nutné postupovat jako v případě přiřazení mezd k absolventům. Pomocí rodného čísla připojit mzdy z databáze ČSSZ. Pomocí jistých předpokladů bude následně možné odhadovat dopad určité vysoké školy na pozdější mzdu absolventu.

5.3 Zaměstnanost absolventů

Kromě mezd absolventů je též možné sledovat v jakém zaměstnání a u jakého zaměstnavatele absolventi pracují. Zdrojová data by mohla být opět Česká správa sociálního zabezpečení, přestože není zcela jisté, jak detailní informace by bylo možné získat. Další hypotetické zdroje byly popsány v předchozí kapitole.

Indikátory nebo statistiky by mohli ukazovat, v jakém typu pracovních pozic a v jakém sektoru absolventi pracují (soukromý, státní, neziskový), případně by se mohlo sledovat přímo kategorie zaměstnání (KZAM). Podle toho by též mohlo být zjištěno do jak kvalifikačně náročných profesí absolventi odcházejí.

5.4 Nezaměstnanost absolventů

Již v současnosti existují veřejné informace o nezaměstnanosti absolventů jednotlivých fakult a škol. Tyto informace se nacházejí například na internetové stránce www.infoabsolvent.cz. Je však nutné dodat, že v současnosti nejsou tyto informace prezentovány v časové a regionální struktuře.

Absolventi různých programů a škol mají totiž jiné možnosti uplatnění na trhu práce v různých regionech. U některých programů trvá přirozené déle nalézt první zaměstnání.

Prezentace by tudíž měla dávat důraz na dlouhodobou nezaměstnanost, která je skutečný problém pro absolventy.

5.5 Index celkové kvality uplatnění absolventů na trhu práce

Podat komplexní informaci o kvalitě uplatnění na trhu práce není jednoduché. Nicméně je možné vytvořit souhrnný index uplatnění na trhu práce, který bude zahrnovat informace o mzdách, zaměstnanosti a kvalitě zaměstnání. Takový index může být jednoduše prezentovaný, ale může mít samozřejmě řadu problému při konstrukci a interpretaci.

5.6 Pokročilé indikátory z propojené databáze SIMS

5.6.1 Regresní analýza vysvětlující vliv školy na mzdy

V této části se dostáváme do části, která spíše směřuje k velmi hypotetické úvaze o tom, co by se dalo dělat s daty, o jejichž podobě momentálně příliš nevíme. Proto se zaměříme především na obecnou úvahu o možných výstupech a jejich užitečnosti.

Pokročilá analýza s daty, které by obsahovaly jak mzdy, tak školu, z které absolvent pochází, by se měla především zaměřit na hledání kauzálního vztahu dopadu absolvované školy na mzdy. Kvalita této analýzy bude do značné míry záviset na velikosti a rozsahu dat, které budou k dispozici. Základní databáze by však měla obsahovat informace o mzdách po dosažení titulu.

V tomto případě by šla dělat jednoduchá regresní analýza, která bude vysvětlovat nástupní mzdy v závislosti na absolvované škole nebo programu. Tento typ analýzy by měl tu výhodu, že je schopen odfiltrovat kontrolovatelné socioekonomické vlivy, včetně regionálních rozdílů.

Pokud bychom pokračovali dále a získali bychom data panelového typu, to znamená více pozorování mezd u každého jednotlivce, mohli bychom zkoumat vliv školy na celkový mzdový profil absolventa.

Význam této analýzy je především v tom, že dokáže oddělit vliv dané školy od jiných pozorovatelných charakteristik. Díky tomu bychom například mohli říci do jaké míry

jsou pozorovatelné mzdové rozdíly mezi absolventy různých škol dány rodinným zázemím nebo absolvováním dané školy. Podobně je možné testovat hypotézu, zda mzdy jsou především determinovány regionem, kde si absolvent najde práci, nebo je mzdový rozdíl dán rozdílem v kvalitě školy. Podobně by bylo možné zahrnout do této analýzy vliv typu absolvované střední školy a zjistit do jaké míry typ střední školy determinuje pozdější mzdy, bez ohledu na pozdější absolvovanou vysokou školu.

Samozřejmě každá taková analýza má své problémy a limity. Například je možné, že absolventi většiny škol se uplatňují pouze v tom regionu, kde sídlí ta daná škola. Řešení problému tohoto typu je v takzvaném neparametrickém odhadu pomocí metody párování. Tento problém by však pravděpodobně neexistoval v takové míře v analýze, kdy bychom chtěli oddělit vliv ostatní socioekonomických charakteristik například rodičů od vlivu absolvované školy.

Pokud by se podařilo vytvořit databázi s panelovými daty, tj. více pozorování mezd za jednotlivce v čase, je možné provést i sofistikovanější analýzu, která by mohla očistit výsledky o tzv. nepozorovatelnou heterogenitu (viz příloha tohoto textu).

5.6.2 Regresní analýza vlivu školy na zaměstnanost a nezaměstnanost

Podobně jako v případě mezd, lze podobným způsobem analyzovat problém zaměstnanosti a nezaměstnanosti absolventů jednotlivých škol. Pomocí probit nebo logit modelu by bylo možné zjistit, zda absolvování dané školy vede k získání vysoce kvalifikované práce, případně k nižší pravděpodobnosti, že se absolvent stane nezaměstnaný. V tomto případě by daná testovaná hypotéza měla znít, jak daná škola přispěje k tomu, že absolvent má kvalifikovanou pracovní pozici, na rozdíl od průměrného absolventa jiné školy. Jako i v předchozím případě bychom dosáhli toho, že odhadnutá pravděpodobnost bude očištěna o ostatní kontrolované charakteristiky. Podobně by mohla být provedena analýza toho, jak často absolventi určité školy nebo programu mění zaměstnání. Tato analýza by mohla sloužit k porovnání škol v jejich úspěšnosti v připravování studentů na získání prvního povolání na trhu práce.

Analýza nezaměstnanosti by v současné době byla asi nejdostupnější, jelikož existují dostupná data o nezaměstnanosti absolventů z databáze úřadů práce. Zde by mohla být

vytvořena analytická studie jak dlouho zůstávají absolventi různých škol v nezaměstnanosti před vstupem do prvního zaměstnání. Pokud by byly k dispozici panelová data o absolventech a jejich působení v nezaměstnanosti, lze použít ekonometrickou analýzu hazardní míry, která je blíže popsána v příloze tohoto textu.

6 Informace sbírané v zahraniční a indikátory z nich generované

V této části diskutujeme různé zahraniční přístupy – jak ke sběru dat, tak k výstupům z nich generovaným. Je třeba zdůraznit, že následující přehled není všeobsažný. Pro jednoduchost jsme se zaměřili na země, které jsou pro ČR v určitém smyslu relevantní. Zmíníme studie mezinárodní – věnující se primárně přehledům mezi různými zeměmi (OECD nebo celého světa). Dále budeme diskutovat případ Německa, Slovenska, Velké Británie, Nizozemska a Finska. Na závěr se krátce zmíníme o tom, jak s dostupnými daty nakládají akademické studie – jaké otázky řeší, jak je zodpovídají a co k tomu potřebují.

6.1 Německo

Německý informační systém o terciárním vzdělání je trochu více vyvinutý než ten v České republice. Na stránkách Německého statistického úřadu (Statistische Bundesamt Deutschland - SBD, www.destatis.de) lze najít časové řady různých indikátorů o tamním školství, většinou rozdělených podle zemí a programů. Základní indikátory jsou počty studentů, počty nově přijatých studentů, a počty absolventů v daném roce. V principu jsou tyto informace velmi podobné těm, co publikuje ÚIV, akorát jsou ve výrazně přehlednější formě a s poměrně dlouhými časovými řadami dostupné na jednom místě, zatímco ÚIV nabízí tabulky v excellu pouze vždy za jeden rok a časovou řadu si uživatel musí u většiny indikátorů vytvořit sám. Na rozdíl od ÚIV Německý statistický úřad neukazuje údaje za konkrétní školy, ale pouze za několik hlavních oborů: jazyky, sport, sociální vědy a právo, matematika a přírodní vědy, veterinářství, medicína, veterinární medicína, umění, technické obory. Zajímavé jsou též statistiky uplatnění žen v celé akademické sféře, včetně studentů na různých stupních studia.

Německým statistický úřad odkazuje i na další zdroje informací o terciárním vzdělávání. Asi nejobsáhlejší je internetový portál Informationssystem Studienwahl und Arbeitsmarkt (ISA - <http://www.uni-due.de/isa/>) - Informační systém pro výběr studia a pracovní trh. Tento informační systém podává informace o následujících oborech:

- Ekonomie/právo (ekonomické vědy, právo, advokacie)
- Technické obory (Strojírenství, elektrotechnika, stavitelství)
- Matematika a přírodní vědy (matematika, informatika, fyzika, chemie, biologie)
- Medicína a farmacie (všeobecné lékařství, farmacie, zubní lékařství)
- Sociální vědy (sociální vědy, pedagogika, psychologie)
- Humanitní vědy (Historie, anglistika/ameristika, germanistika)

Dle takto rozdělených oborů se uvádí řada statistik a indikátorů, některé jsou obdobné jako u v případě SBD a české ÚIV: počty studentů, počty absolventů a podobně. Nicméně navíc prezentují i data za délku studia a to jak teoretickou, tak skutečnou a věk studenta při ukončení studia. Další indikátor je úspěšnost dokončení studia, která je počítána jako poměr začínajících studentů a absolventů daného oboru.

Další indikátory se týkají úspěšnosti absolventů na pracovním trhu. Zde uvádějí podíl aktivních na trhu práce podle oborů a věkové skupiny (<30, 30-40, 40-50, >50), dále podíl aktivních na trhu práce s vystudovaným určitým oborem, ale pracujících v jiném, příbuzném oboru, historický počet nezaměstnaných s daným oborem vzdělání (plus podíl věkové skupiny pod věk 35) a nakonec množství volných pracovních míst v daném oboru.

6.2 Finsko

Na úvod bychom rádi připomněli studii „[Education and Science in Finland](#)“, zpracovanou finským ministerstvem školství. Tato souhrnná zpráva poskytuje základní pohled napříč finským vzdělávacím programem. V sekci věnované vysokému školství poskytuje základní statistické informace z pohledu participace na terciárním vzdělávání, distribuce studentů napříč obory, ale i distribuce absolventů na trhu práce. Zároveň poskytuje

informace i o připravované reformě vysokého školství plánované ke schválení na jaře 2009.

Studii s detailnějším zaměřením na terciární vzdělávací systém je výroční zpráva „[Universities 2006](#)“, opět vypracovaná ministerstvem školství. Poskytuje základní souhrnné statistiky jednak na úrovni jednotlivých univerzit či vzdělávacích oborů, ale zároveň porovnává i časový vývoj vybraných ukazovatelů (například financování, participace studentů v jednotlivých stupních terciárního vzdělávání) během období posledních deseti let.

Dodatečný pohled na statistiky finského vysokého školství je možné získat i od finského statistického úřadu, který má na starosti sběr dat o studentech všech finských univerzit. Poskytnutí údajů o jednotlivých studentech ze strany univerzit je povinné, bohužel však stránka statistického úřadu neinformuje o tom, jak jsou jednotlivé informace o studentech sbírané. Sebrané údaje jsou statistickým úřadem použity ke sledování statistik, jako je například přerušení studia, počet uchazečů o studium, průchod studentů studiem a jejich následný přechod do dalšího vzdělávání anebo vstup na trh práce.

Kromě výše uvedeného statistický úřad poskytuje sebrané údaje o studentech ve formě souhrnných dat ministerstvu školství, které je spravuje prostřednictvím databáze „[KOTA](#)“. Databáze KOTA poskytuje možnost vyhledávání údajů v těchto kategoriích: uchazeči o studium a přijatí ke studiu, noví, zahraniční a pokračující studenti, udělené tituly, místo dalšího působení absolventů, medián délky trvání studia, učitelé a další pracovníci, výroční účty, vědecké publikace, studentské výměnné pobyty, a v dalších kategoriích, které už nemají přímou souvislost se zadanou úlohou tohoto projektu. Všechny údaje jsou k dispozici napříč jednotlivými univerzitami a roky (databáze obsahuje údaje od roku 1981), jakož i podle pohlaví studentů.

6.3 Nizozemí

V souvislosti s touto zemí bychom rádi připomenuli zprávu „[Key Figures](#)“, která tematicky pokrývá oblast celého vzdělávacího systému. Obsahově se jedná o obširnou studii zaměřenou nejen na charakteristiky a statistické údaje holandského vzdělávacího

systemu, ale se dívá na holandský vzdělávací systém i z pohledu mezinárodní perspektivy. Konkrétně, daná studie porovnává kvalitu i kvantitu výstupů jednotlivých úrovní vzdělávacího systému s ostatními krajinami včetně České republiky. Za připomenutí opět stojí i fakt, že předmětná studie obsahuje i pohled na stranu učitelů a profesorů, a opět se věnuje porovnání jejich postavení v mezinárodním měřítku.

Z hlediska vysokoškolského vzdělávání a v rámci kontextu zadaného úkolu se jeví zajímavě část zaměřená na analýzu míry úspěšnosti dokončení studia, jakož i informace o předpokládané délce studia pro ten který obor a jejich změnu v čase. Této údaje jsou dostupné pro každý jednotlivý stupeň terciárního vzdělávání jakož i pro různé druhy univerzit z pohledu jejich zaměření. Bohužel, podrobnější statistiky poskytující informace o použitých proměnných jakož i o způsobu výpočtu už nejsou v předmětné studii dostupné.

Další studií, tentokrát věnující se výhradně pohledu na holandské vysoké školství je aktuální studie „[Facts and Figures: Universities in the Netherland](#)“. Tato studie se zaměřuje hlavně na porovnání základních statistik charakterizujících vysoké školy a jejich vzájemný porovnání. Zajímavým přínosem oproti studii „[Key Figures](#)“ jsou informace upozorňující na mezinárodní porovnání kvality vybraných holandských univerzit.

6.4 Slovensko

Slovensko coby země je velmi blízké České republice a proto je co se týče vysokoškolského vzdělávacího systému přirozenou volbou při porovnávání kvantity i kvality statistických údajů o vysokých školách.

Ministerstvo školství Slovenské republiky, coby zastřešující instituce vysokoškolského studia na Slovensku, každoročně vypracovává [výroční zprávu o stavu vysokého školství](#), kterou předkládá vládě Slovenské republiky. Tato zpráva obsahuje tradiční informace o počtu studentů, o jejich přílivu a odlivu ve formě nově přijatých studentů a absolventů, a také informace o podpoře rozvoje a financování vysokých škol.

Při MŠ SR funguje [Ústav informací a prognóz školství \(ÚIPŠ\)](#), který zabezpečuje komplexní zpracování informací o vzdělávání na Slovensku. Co se týče charakteru, jakož i náplně činnosti, ÚIPŠ je možné přirovnat k [Ústavu pro informace ve vzdělávání v České republice](#). V oblasti analýz vysokého školství na Slovensku stojí za povšimnutí každoroční [“Odhad hodnôt a trendov rozhodujúcich ukazovateľov činnosti vysokých škôl”](#). Kromě tradičních vstupně-výstupních časových řad počtu nově přijatých studentů a počtu absolventů předmětná zpráva analyzuje i model průchodu studujícími vzdělávacím systémem. Zjednodušeně řečeno, diskutovaný model využívá časové řady počtu zapsaných studentů, počtu absolventů a počtu neúspěšných studentů na výpočet dvou ukazatelů použitých při výpočtu dlouhodobé prognózy: poměr počtu absolventů k odpovídajícímu počtu nově přijatých studentů a poměr počtu neúspěšných studentů k počtu studujících v předchozím roce. Tyto dvě informace jsou následně využity při tvorbě dlouhodobé prognózy počtu absolventů a počtu studentů, kteří studium nedokončí. V tomto bodě je však nutné připomenout, že na výpočet prognózy se používá pevná hodnota těchto poměrů vypočítaná z dlouhodobého trendu jejich vývoje.

Analýza zabývající se neméně důležitou otázkou [“Uplatnenia absolventov vysokých škôl v praxi”](#) byla opět vypracována Ústavem informací a prognóz školství, oddělením analýz a prognóz vysokých škol. Daná studie analyzuje problematiku uplatnění absolventů na trhu práce na základě disponibilních statistických údajů ze školské statistiky a ze statistiky o zaměstnanosti a nezaměstnanosti dostupných prostřednictvím statistických výsledků o nezaměstnanosti absolventů škol a mladistvých z dílny [Ustredia prace socialnych veci a rodiny](#). Kromě pohledu na domácí trh práce se studie zaměřuje i na srovnání s ostatními zeměmi Evropské unie a to zejména s odvoláním se na dva výzkumy hodnotící přínos vysokoškolského studia, a to [CHEERS \(zkratka pro “Careers after Higher Education: an European Research Study”\)](#) a [REFLEX \(zkratka pro “The Flexible Professional in the Knowledge Society: New Demands on Higher Education in Europe”\)](#), v kterých na rozdíl od českých absolventů ti slovenský neparticipovali.

Na závěr této statě bychom se rádi věnovali způsobu, jakož i obsažnosti sběru dat o

individuálních studentech. Nakolik databáze projektu SIMS obsahuje rovněž individuální data o jednotlivých studentech, je možné porovnat kvalitu, kvantitu jakož i vytyčené cíle slovenského projektu Centrálního registru studentů vysokých škol s českým projektem SIMS.

Na základě zákona Slovenské republiky č. 131/2002 Z. s. o vysokých školách má každá vysoká škola povinnost vést údaje o svých studentech. Tyto údaje jsou potom sdružené v centrálním registru studentů veřejných vysokých škol, studentů vojenských vysokých škol, kteří nevykonávají vojenskou službu, studentů policejních vysokých škol, kteří nejsou v služebním poměru, a studentů zdravotnických vysokých škol. Do registru studentů vysoké školy se zaznamenává jméno, příjmení, tituly, rodné číslo, datum narození u cizinců, místo narození, rodinný stav a místo trvalého pobytu studenta, pohlaví, bydliště v Slovenské republice, národnost, státní občanství, číslo občanského průkazu nebo číslo pasu u cizinců. Kromě těchto základních informací registr studentů vysokých škol obsahuje dále údaje o zápisu do studia, předcházejícím vzdělání, studijním programu, zápisu do vyššího ročníku nebo další části studia, přidělení ubytování na koleji, přiznání sociálního stipendia, absolvování státní zkoušky a udělení akademického titulu, přerušení a ukončení studia.

Při porovnání slovenské databáze s databází SIMS vidíme, že databáze SIMS je obsažnější, jelikož obsahuje všechny údaje zjišťované v slovenském centrálním registru studentů vysokých škol plus některé dodatečné informace zakrňující financování studia anebo místo výuky či formu studia. Je třeba však dodat, že na základě informací z Ministerstva školství Slovenské republiky je projekt Centrálního registru studentů vysokých škol stále v procesu vývoje.

6.5 Velká Británie

Jako příklad země poskytující širokou škálu statistických údajů o vysokoškolském vzdělání patří Velká Británie. Hlavním zdrojem těchto dat je "[Learning and Skills Council](#)", veřejně prospěšná organizace s hlavním cílem budování dynamického systému dalšího vzdělávání.

V kontextu zadané úlohy má LSC za úkol publikaci široké škály informací, ať už za účelem usměrňování nebo poradenství, či za účelem informování ve formě analýz. Úlohu sběru a správy dat má na starosti [“The Data Service”](#), instituce s výlučnou zodpovědností v tomto směru. Přestože se jedná o dvě různé instituce, v současné době jsou data dostupná přes portál LSC.

V případě Velké Británie, na rozdíl od ostatních zkoumaných zemí Evropské unie, je zřejmé, že je sběr dat vykonáván na úrovni jednotlivce, a to prostřednictvím [„Individualised Learner Record \(ILR\) a Individualised Staff Record \(ISR\)“](#).

ILR je tedy souhrn dat o studujících jakož i o jejich individuálním postupu studiem. Následně, podle oficiálních zdrojů, je tato obsáhlá databáze používána pro monitoring jednotlivých poskytovatelů vysokoškolského vzdělávání ve vztahu k plnění jejich vzdělávacích plánů, kvality vzdělávacího procesu i způsobu jeho financování. Ze specifikací projektu ILR vyplývá, že subjekt povinný nahlašovat údaje o studentech je povinný tak učinit pětkrát do roka.

Co se týče struktury samotné databáze, individuální záznam studenta je rozdělený do čtyř navazujících kategorií. Nejvyšší kategorii tvoří základní informace o studentovi zahrnující osobní informace shodné s informacemi obsaženými v databázi SIMS. Navíc tato část databáze obsahuje i informace o potenciálních zdravotních problémech, které mohou mít vliv na proces vzdělávání, jakož i informace o zaměstnání před a po započetí studia.

Na tuto část databáze navazuje souhrn dat popisující cíle studijního programu konkrétního studenta. V této kategorii se nacházejí informace o druhu a výši finanční podpory pro daného studenta a daný rok studia, typ programu a jeho forma (denní studium, víkendové studium, blokové studium), dále se rozlišuje studium tradiční formou s pravidelným kontaktem s vyučujícím například od studia formou internetového vzdělávání. Je důležité připomenout, že všechny tyto charakteristiky mohou mít významnou roli při vyhodnocování efektivity a kvality daného studijního programu. Dále se tato část databáze skládá z očekávaných záznamů typu začátek studijního programu,

předpokládaná délka studia, počet kreditů v rámci celého studia, skutečná délka studia, počet skutečně dosažených kreditů, výsledné hodnocení a také informace o speciálních projektech, kterých se student účastnil.

Na konkrétní typ studijního programu navazuje následně část databáze zahrnující informace o spolufinancování ze strany Evropského sociálního fondu (ESF). Tyto informace jsou uváděny jenom v případě, pokud je daný studijní program spolufinancován ze strany ESF. V daném případě tato část databáze obsahuje kromě jiného následující informace: začátek a konec spolufinancování, předcházející zaměstnání studujícího, popřípadě délka nezaměstnanosti, hospodářský sektor, ve kterém studující působil před začátkem studia a předcházející dosažené vzdělání.

Opět v návaznosti na konkrétní druh studijního programu obsahuje tato část informace o nejvyšší dosažené kvalifikaci studujícího, aktuálním roce vzdělávání, informace o potenciálním přerušení studia a jeho důvodech. Důležitými údaji jsou i informace o potenciálním vzdělávání na jiné instituci jako části studijního programu a jeho proporcionální délce vzhledem k délce studia, což opět umožňuje vyvodit jisté závěry ve vztahu k dosaženým výsledkům studenta. Dalším zajímavým údajem, který se v databázi SIMS nenachází, je informace o formě ubytování studenta v tom kterém roce studia.

Z pohledu zadané úlohy je v rámci porovnání databáze ILR ve Velké Británii s údaji dostupnými prostřednictvím databáze SIMS nutné poznamenat, že obě databáze se soustředí na zkoumání především kvantitativních údajů o studentech. Je možné říci, že ILR je ve své formě obsáhlejší, a proto umožňuje analýzy v některých zajímavých oblastech zdůrazněných v předcházejících odstavcích. Přesto, ani v databázi ILR se nenachází bližší osobní charakteristiky jednotlivých studentů, které by umožnily analýzu zaměřující se především na porovnání vhodnosti jednotlivých typů uchazečů pro daný typ školy, jakož i na porovnání úspěšnosti dokončení resp. délky trvání studia vzhledem k rozdílným kvalitám studujících. V tomto ohledu bychom viděli potenciální možnost ve spojení více databází, které by obsahovaly individuální údaje tohoto charakteru (úplná verze popisující detailně jednotlivé položky je dostupná na <http://www.theia.org.uk>).

Stejně jako v případě ostatních států i ve Velké Británii vychází [statistické ročenky](#) na téma vysokého školství. Avšak podle našich informací v porovnání například se Slovenskem v nezanedbatelné míře sledují úspěšnost dosažení stanovených cílů, ať už ze strany studentů, nebo ze strany jednotlivých institucí. Zároveň samostatnou sekci tvoří i sledování efektivity výuky v té které instituci se zaměřením na podporu a vedení studentů, což je sledováno prostřednictvím meziročních změn v počtu studentů pokračujících ve studiu.

6.6 Mezinárodní studie

Studie OECD a Světové banky týkající se vysokého školství je možné rozdělit na studie zaměřující se na jednotlivé země a na studie poskytující souhrnný pohled a vzájemné porovnání mezi zeměmi. Mezi poslední jmenované studie zaměřující se na celou oblast školství, nejen na vysoké školy, patří například studie "[Education at a Glance 2007](#)" umožňující porovnat množství indikátorů výkonnosti školského systému v různých zemích. Jako příklad ve vztahu k zadané úloze projektu SIMS, jakož i tvorby modelu odhadujícího pravděpodobnost úspěšnosti dokončení studia se jeví zejména indikátory opisující vliv socioekonomického statusu rodičů na participaci studentů v terciárním vzdělávacím systému. Zajímavou kategorií tvoří však i část zkoumající jako školy přistupují k monitorování jejich vlastní výkonnosti.

Na druhé straně jako doplněk studií výše uvedeného typu slouží reporty OECD a Světové banky zaměřující se na analýzu vysokého školství v dané zemi. Jako příklad OECD studie zaměřenou na Českou republiku uvádíme „[Thematic Review of Tertiary Education, Czech Republic](#)“. Zároveň je však nutné dodat, že studie se stejným zaměřením jsou přirozeně dostupné i pro další státy, či už z Evropy anebo dalších částí světa. Ze strany Světové banky se nám nepodařilo najít studii zaměřující se výhradně na Českou republiku v oblasti terciárního vzdělávání, a tak jako příklad uvádíme studii se zaměřením na blízkou Polskou republiku: „[Tertiary Education in Poland](#)“ vypracovanou Světovou bankou a Evropskou Investiční Bankou.

V rámci shrnutí předchozích dvou odstavců je nutné zdůraznit, že společným rysem výše uvedených studií obou typů je jejich zaměření na porovnání se zeměmi západní Evropy. V případě zájmu o detailnější informace a statistiky se musí zájemce obrátit na příslušné ministerstvo školství, případně pověřené instituce zaměřující se specificky na oblast analýz a statistického vyhodnocování školství.

Tematicky odlišnou kategorií tvoří studie poskytující pohled na vnímání významu vysokoškolského vzdělání z pohledu jeho využití v praxi. Do této kategorie je možné zahrnout studii [CHEERS – “Careers after Higher Education: an European Research Study”](#) snažící se odpovědět na otázku, jak vidí absolventi vztah mezi svým vzděláním a praxí 4 roky po ukončení studia. Této studii se zúčastnili absolventi 12 krajín, převážně z rozvinutých ekonomik. Zajímavým pro nás je, že účastníci zjišťování byli i absolventi z České republiky, což umožňuje poukázat na rozdílné vnímání zkoumaného vztahu v porovnání se západoevropskými krajinami.

Studii s podobným zaměřením je také [REFLEX – “The Flexible Professional in the Knowledge Society: New Demands on Higher Education in Europe”](#) odpovídající především na tuto otázku: Jakým způsobem instituce terciárního vzdělávání pomáhají absolventům při zaujímání významné pozice z pohledu vzdělanostní ekonomiky? K jakým střetům dochází, když se každá ze skupin (absolventi, vysokoškolské instituce a zaměstnavatelé) snaží dosáhnout svých vlastních cílů, a jakým způsobem je možné těmto střetům zájmů předejít, případně je vyřešit? Opět se této studii zúčastnili i absolventi České republiky, jakož i absolventi z dalších 14 převážně západoevropských zemí plus Japonska.

7 Využití indikátorů pro potřeby uživatelů

Navržené indikátory splní svůj účel, pokud budou dostatečně srozumitelně jednoduše prezentovány veřejnosti. Tato práce si neklade za cíl konkrétní návrh, jak by měl informační systém o terciárním vzdělávání vypadat, nicméně bychom rádi zdůraznili určité obecné pravidla prezentace pro různé typy veřejnosti.

7.1 Laická veřejnost

- Informační systém by měl být prezentován pomocí internetového rozhraní, které bude maximálně jednoduché pro používání a získání důležitých informací.
- Vše by mělo směřovat k tomu, aby uchazeč měl možnost udělat informované rozhodnutí o přihlášce a studiu na vysoké škole. Ideálně by proto měl mít možnost porovnat více škol s daným oborem a udělat si obrázek jak studium na jednotlivých školách přibližně probíhá.
- Přestože v této práci prezentujeme velké množství možných indikátorů, jejich výsledný prezentovaný počet může být výrazně nižší. Případně by měly být rozděleny do několika souhrnných kategorií
- Uživatel by si měl být vědom, co si pod daným indikátorem má představit. To znamená, že každý název indikátoru by měl být „sebe-vysvětlující“ a doprovázen podrobnějším popisem v poznámce.
- Pro pochopení významnosti a vypovídací hodnotě jednotlivých indikátorů, je nutné zdůraznit jejich interpretační omezení a to především v tom, že se jedná o indikátory napočítané nebo odhadnuté z dat za minulé období. Z toho vyplývá hlavní omezení, které je nutné brát pro jejich interpretaci.
- Srovnání škol, které nabízejí stejný obor, by nemělo vyústit ve vytváření žebříčků domnělé „kvality“. Naše navrhované indikátory pouze statistickým způsobem popisují danou školu, nebo obor na dané škole. Nic z nich nutně a přímo nevypovídá o kvalitě dané školy. Vytváření žebříčků kvality by mělo být přenecháno jiným institucím, které využívají jiný typ informací a kritérií.
- V tomto kontextu by měl být celý výstup prezentován médiím. Na komunikaci s médii by měl být dán velký důraz, jelikož veřejnost bude věrohodnost výstupu hodnotit právě optikou médií.

- Podobně jako média, i samotné vysoké školy by měly být informovány o plánovaných výstupech a jejich prezentaci. Rozhodně by se však nemělo podléhat možným zájmům, aby se jisté indikátory neprezentovaly, nebo se prezentovaly pouze v určité podobě.

7.2 Odborná veřejnost

Současný společenskovední výzkum stále více využívá individuální data. Rozšířená databáze SIMS bude velmi vhodný zdroj těchto dat a stát by měl mít zájem na tom, aby vznikalo co nejvíce vědeckých prací a tím se rozšiřovalo poznání o terciárním vzdělávání v České republice a jeho dopad na úspěšnost absolventů na trhu práce. Databáze SIMS by proto měla být v takové technickém stavu, aby na ní mohl probíhat empirický výzkum. To mimo jiné znamená anonymizovat náhodný výběr z populace studentů a změnit rodné číslo na jiný identifikátor. Tento systém anonymizace dat je v zahraničí běžný a odpovídá i současnému trendu v Evropské unii.

7.3 Ministerstvo školství, mládeže a tělovýchovy

MŠMT může využívat informační systém k monitoringu stavu jednotlivých vysokých škol a monitoringu toku finančních prostředků v rámci terciárního vzdělávání. Z tohoto pohledu by mělo MŠMT mít možnost získávat informace, které potřebuje a zadávat výpočet potřebných indikátorů.

7.4 Akreditační komise MŠMT

Akreditační komise se zabývá především kvalitou studijních programů a řada indikátorů, které navrhujeme, se týká průchodu studentů studiem, tudíž mohou sloužit minimálně jako doplňková informace k tomu, co daná škola předkládá Akreditační komisi. Stejně jako MŠMT, může se Akreditační komise vyžádat speciální indikátory, které ji mohou zajímat.

7.5 Zahraničí instituce

Se zvyšující se mobilitou studentů a propojenosti pracovního trhu v Evropské unii, bude narůstat i poptávka po mezinárodním statistickém srovnávání jednotlivých vysokých

škola. Proto by SIMS měl být připraven poskytovat indikátory, které budou odpovídat centralizovanému informačnímu systému na úrovni EU a OECD.

8 Závěr

V této studii diskutujeme možnosti, jak usnadnit rozhodování jednotlivců o studiu na vysoké škole. Popisujeme jak získat dodatečné statistické informace z existující databáze Sdružených informací matrik studentů. Všechny indikátory vycházejí buď ze současných informací o studentech zaznamenaných v matrice studentů, případně z jejího rozšíření o další charakteristiky studentů, nebo z možností propojení s externími datovými zdroji.

Většina indikátorů má formu deskriptivních statistik (počty studentů podle řady kritérií, průměrné délky studií, úspěšné versus neúspěšné studia). Navíc diskutujeme i využívání pokročilých statistických metod a navrhuje využití regresní analýzy pro doplnění základních indikátorů. Výsledky této regresní analýzy by měly především vnést vzhled do kauzálních dopadů jednotlivých škol na úspěšnost studentů ve studiu, případně na jejich působení na trhu práce. V příloze tohoto textu diskutujeme technické náležitosti této analýzy a ukazujeme užitečnost i limity na jednotlivých příkladech možného využití.

Výčet možných indikátorů, které uvádíme, však zdaleka není konečný a bezesporu si neklademe za cíl jejich úplný výčet. Pouze se snažíme upozornit na některé, dle našeho názoru, podstatné a zároveň ukázat jejich užitečnost a případně interpretační omezení.

V této studii též upozorňujeme na mnoho výhod rozšíření a propojení matrik studentů s externími databázemi, a to zejména s mikroekonomickými daty z trhu práce a případně s informacemi o předchozím studiu, které jsou dnes dostupné například v databázi Uchazeč.

Rozšíření by mělo také směřovat k získání více socioekonomických dat od samotných studentů, které se mohou získávat dotazníkovým nebo jiným šetřením. Námi uváděný seznam charakteristik však nemusí být logicky naplněn, vzhledem k předpokládané obtížnosti získávání informací o rodinných příslušnících.

Velká pozornost by se však měla upínat k propojení současné SIMS s daty o mzdách nebo o (ne)zaměstnanosti na trhu práce. Navrhovaná analýza úspěšnosti absolventů na trhu práce by výrazně obohatila informovanost uchazečů o jejich perspektivách, pokud si zvolí danou vysokou školu a studijní program. Na druhou stranu též zdůrazňujeme, že toto propojení má řadu technických problémů a v blízkém časovém horizontu se nemusí podařit a to bez ohledu na připravenost matriky studentů. Přesto si myslíme, že o tomto propojení je nutné do budoucna uvažovat a směřovat k tomu.

Též diskutujeme poskytované informace o terciárním vzdělávání v zahraničí a docházíme k závěru, že v současné době, neexistuje ideální informační systém, který by mělo význam plně následovat. Německý informační systém přináší jisté indikátory, které jsou do značné míry společné s těmi, které diskutujeme, v současné době jsou pouze na úrovni široce definovaného oboru studia a ne na úrovni programu na vysoké škole.

Závěrem je nutné dodat, že pokud mají navrhované indikátory najít úspěšně cílovou skupinu – uchazeče o studium – je nutné je správně a srozumitelně prezentovat. Zde je především nutné brát na zřetel omezení v interpretaci napočítaných nebo odhadnutých indikátorů. Je nutné vytvořit v rámci prezentace jejich přesný popis a vysvětlit, že jsou logicky zatíženy nejistotou, vyplývající z využití minulých dat. Minulý vývoj samozřejmě nemůžeme automaticky extrapolovat do budoucna.

Zvýšená pozornost by se měla dát komunikaci indikátorů veřejnosti, aby indikátory byly brány jako věrohodné a zároveň aby se nevytvářela přehnaná očekávání. Též by měly být do procesu zveřejňování zapojeny i vysoké školy. Navržený informační systém má však především sloužit studentům, MŠMT a případně společenským vědcům, nikoliv jiným parciálním zájmům.

9 Reference a internetové zdroje

Články a publikace

- Barefoot B. (2004) *Higher education's revolving door: confronting the problem of student drop out in US colleges and universities*, *Open Learning*, 19(1):9-18, Získaný 10. Listopadu, 2009 z <http://web.ebscohost.com/ehost/pdf?vid=2&hid=7&sid=8b07ccf2-e639-4c12-a571-a87d7667cd0a%40sessionmgr11>
- Cingano, F. & Cipollone P. (2003). *Determinants of university drop-out probability in Italy*. (Bank of Italy, Research Department). Získaný 25. Května, 2009, z http://www.econ.upf.edu/~cingano/papers/dropout_EALE.pdf
- File J., Weko T., Hauptman A., Kristensen B. & Herlitschka S. (2006). *Thematic Review of Tertiary Education: Czech Republic, Country Note*, Organization for Economic Co-operation and Development (OECD). Získaný 15. Května, 2009, z <http://www.oecd.org/dataoecd/8/32/37730231.pdf>
- Learning and Skills Council (Rada pro učení a dovednosti) (2001). *Summary Statistics for Further Education Institutions, England 2000 – 2001*. Získaný 15. Května, 2009, z <http://readingroom.lsc.gov.uk/lsc/National/SummaryStatistics200001.pdf>
- Learning and Skills Council (Rada pro učení a dovednosti) (2009) *Individualised Learner Record (ILR)*. Dostupný na http://www.theia.org.uk/NR/rdonlyres/6814BC6A-4E90-4AC4-9980-5432C1BF6EA9/0/natILRSpecification2008_09v427Feb2009.pdf
- Maddala, G.S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press.
- Ministerstvo Školstva Slovenskej Republiky (2008). *Výročná správa o stave vysokého školstva za rok 2008*. Získaný 21. Května, 2009, z http://www.minedu.sk/data/USERDATA/VysokeSkolstvo/VyrSprav/sprava_stav_VS_2008.zip
- Ministry of Education, Culture and Science, Netherland (Ministerstvo školství, kultury a vědy, Nizozemí) (2007). *Key Figures: Education, Culture and Science, 2003 – 2007*. Získaný 15. Května, 2009, z http://www.minocw.nl/documenten/Key_Figures_PDF.pdf
- Ministry of Education, Finland (Ministerstvo školství, Finsko) (2007). *Universities 2006: Annual Report*. Získaný 16. Května, 2009, z <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2007/liitteet/opm19.pdf>

- Ministry of Education, Finland (Ministerstvo školství, Finsko) (2008). *Education and Science in Finland*. Získaný 16. Května, 2009, z <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2008/liitteet/opm25.pdf?lang=en>
- Montmarquette, C., Mahseredjian S. & Houle. R. (2001). *The determinants of university dropouts: a bivariate probability model with sample selection*. *Economics of Education Review*, Vol. 20, No.5. (Oct., 2001), pp. 475-484.
- OECD (2007). *Education at a Glance: OECD Indicators – 2007 Edition*, OECD, Paris. Získaný 15. Května, 2009, z <http://www.oecd.org/dataoecd/36/4/40701218.pdf>
- Reason R. (2003) *Student Variables that Predict Retention: Recent Research and New Developments*, *NASPA Journal*, 40(4): 172-191, Dostupný na <http://www.uccs.edu/~faculty/bgaddis/retention/Student%20Variables%20that%20Predict%20Retention.pdf>
- Steen, J.V. (2009). *Facts and Figures: Universities in the Netherlands*. Rathenau Institute. Získaný 16. Května, 2009, z <http://www.rathenau.nl/downloadfile.asp?ID=1511>
- Stinebrickner R., Stinebrickner T. (2008) *The Effect of Credit Constraints on the College Drop-Out Decision: A Direct Approach Using a New Panel Study*, *American Economic Review*, 98(5): 2163-2184, Získaný 10. Listopadu, 2009 z <http://www.atypon-link.com/AEAP/doi/pdfplus/10.1257/aer.98.5.2163>
- Suhre C..M., Jansen E.. and Harskamp E. (2007) *Impact of degree program satisfaction on the persistence of college Students*, *Higher Education*, 54:207–226, Získaný 10. Listopadu, 2009 z <http://www.springerlink.com/content/3136081372717380/fulltext.pdf>
- The World Bank (Světová banka) (2004). *Tertiary Education in Poland*, The World Bank, Poland. Získaný 15. Května, 2009, z <http://www.oecd.org/dataoecd/36/4/40701218.pdf>
- Ústav informácií a prognóz školstva, SR (2006). *Vývoj ukazovateľov slovenského vysokého školstva v rokoch 1990 – 2005 a výhľad do roku 2020*. Získaný 21. Května, 2009, z http://www.uips.sk/sub/uips.sk/images/MK/Studie/Trendy_90_05_UIP.pdf
- Ústav informácií a prognóz školstva, SR (2007). *Uplatnenie absolventov vysokých škôl v praxi*. Získaný 22. Května, 2009, z http://www.uips.sk/sub/uips.sk/images/MK/Studie/Uplatnenie_pre_tlac.pdf

Internetové zdroje

Idnes.cz: Žebříček VŠ: kde najdete dobré vědecké zázemí -

http://zpravy.idnes.cz/zebricek-vs-kde-najdete-dobre-vedecke-zazemi-fzr-/studium.asp?c=A071114_113408_vysokeskoly_bar

Ihned.cz: Nový žebříček: Nejlepší vysoké školy v Česku - <http://ihned.cz/c1-34587670>

Individualised Learner Record (ILR) -

<http://www.lsc.gov.uk/providers/Data/Datacollection/ILR/>

Informační systém o uplatnění absolventů škol na trhu práce – www.infoabsolvent.cz

Informationssystem Studienwahl und Arbeitsmarkt (ISA) (Informační systém pro výběr studia a pracovní trh.) - <http://www.uni-due.de/isa/>)

Internationales Zentrum für Hochschulforschung (Mezinárodní centrum pro výzkum v oblasti vysokého školství) – Univerzita Kassel. *Careers after graduation – An European Research Study (CHEERS)*. - <http://www.uni-kassel.de/incher/cheers/index.gkh>

Learning and Skills Council <http://www.lsc.gov.uk/>

Lidovky.cz: Žebříčky vysokých škol -

http://www.lidovky.cz/ln_redakce.asp?r=ln_redakce&c=A060206_150019_ln_redakce_hrn

Ministry of Education, Finland (Ministerstvo školství, Finsko) *KOTA online service*. - https://kotaplus.csc.fi/online/Etusi_vu.do?lng=en

Research Centre for Education and the Labour Market (Výzkumné centrum pro vzdělání a trh práce) *The Flexible Professional in the Knowledge Society: New Demands on Higher Education in Europe (REFLEX)*. Dostupný na <http://www.fdewb.unimaas.nl/roa/reflex/>

Statistische Bundesamt Deutschland (SBD) (Německý statistický úřad), www.destatis.de

Štefl, Ondřej (2009): Na které VŠ jdou nejlepší a jak se to změnilo,

<http://blog.aktualne.centrum.cz/blogy/ondrej-steffl.php?itemid=7678>

The Data Service (Datový servis) <http://www.thedataservice.org.uk/>

Ústav informací a prognóz školství (ÚIPS) <http://www.uips.sk/>

Ustredia práce socialnych veci a rodiny <http://www.upsvar.sk/>

10 Přílohy

P.1 Úplný seznam vedený záznamů v SIMS

Sdružené informace matrik studentů

Popis vstupních dat

Záznamy o studentech a studiích předávané mezi vysokými školami a MŠMT ČR jsou definovány v souladu s § 50 odst. 2 zákona č. 111/1998 Sb. jako **student ve spojení se studijním programem, který studuje na dané škole/fakultě** (jednoznačná kombinace položek "Rodné číslo", "Vysoká škola/fakulta", "Studijní program" a "Datum zápisu"), a to v aktuální situaci v okamžiku sběru dat, resp. spadá-li rozsah platnosti záznamu do období, za něž jsou data sbírána. Při prvním ostrém sběru v roce 1999 byly sbírány všechny záznamy vyjadřující "historické" stavy o všech aktuálních studentech (tj. studentech zapsaných do příslušného studijního programu k 1. lednu 1999 a později, včetně studentů, kteří mají studium přerušeno).

Struktura matričních záznamů je určena dle § 88 (matrika studentů), § 18 odst. 3 a 4 (rozpočet vysoké školy a dotační politika), § 21 odst. 1 písm. c) (poskytování informací MŠMT), a § 87 písm. i) (sdružování a využívání informací) zákona č. 111/1998 Sb. Jednotlivé položky jsou definovány v souladu se závaznými standardy Informačního systému veřejné správy (Katalog datových prvků ISVS) , resp. se závaznými číselníky.

Element	Atribut	Typ	Délka	Povinná	Kontrola	Nadřazený element
Student	<u>RodneCislo</u>	S	10	A	věcná kontrola	Sber Vlozeni Nahrada Oprava
<u>Jmeno</u>		S	24	N	syntaktická kontrola	Student
<u>Prijmeni</u>		S	35	A	syntaktická kontrola	Student
<u>RodnePrijmeni</u>		S	35	N	syntaktická kontrola	Student

<u>TitulPred</u>		S	50	N	ne	Student
<u>TitulZa</u>		S	50	N	ne	Student
<u>Adresa</u>				A		Student
<u>Okres</u>		I	4	A	číselník <u>OKRES</u>	Adresa
<u>Obec</u>		I	6	N	číselník <u>OBEC</u> ; kontrola s elementem "Okres"	Adresa
<u>CastObce</u>		I	6	N	číselník <u>COBCE</u> ; kontrola s elementem "Obec"	Adresa
<u>Ulice</u>		S	48	N	ne	Adresa
<u>UliceCislo</u>		S	10	N	ne	Adresa
<u>PSC</u>		I	5	N	číselník <u>ZV_PSCPO</u>	Adresa
<u>Stat</u>		S	3	A	číselník <u>ISO3166</u>	Adresa
<u>RodinyStav</u>		I	1	A	číselník <u>K2101</u>	Student
<u>PredchoziVzdelani</u>		S	1	A	číselník <u>K21K1</u>	Student
<u>StredniSkola</u>		S	9	N	číselník <u>IZO</u>	Student
<u>RokMatZkousky</u>		I	4	N	kontrola s atributem "RodneCislo"	Student
<u>Studia</u>				A		Student
<u>Studium</u>	<u>VSFakulta</u>	S	5	A	číselník <u>REGPRO</u> ; věcná kontrola	Studia
	<u>StudijniProgram</u>	S	5	A	číselník <u>STUDPROG</u> ;	Studia

					věcná kontrola	
	<u>ZapisDoStudia</u>	D	10	A	věcná kontrola	Studia
<u>DelkaStudia</u>		R	2.1	A	kontrola s atributem "StudijniProgram"	Studium
<u>NovePrijaty</u>		S	1	A	"A" nebo "N"; kontrola s atributem "ZapisDoStudia", "StudijniProgram"	Studium
<u>NavazujiciStudProgram</u>		S	1	A	"A" nebo "N"; kontrola s atributem "StudijniProgram"	Studium
<u>PocetRocniku</u>		I	2	N	ne	Studium
<u>AktualniRocnik</u>		I	2	N	kontrola s elementem "PocetRocniku"	Studium
<u>UbytovaniVKoleji</u>		I	1	A	číselník <u>UBYT</u>	Studium
<i>SocialniStipendia</i>				N		Studium
SocialniStipendium	<u>NarokOd</u>	D	10	A	kontrola s atributem "ZapisDoStudia"	SocialniStipendia
<u>NarokDo</u>		D	10	N	kontrola s atributem "Datum"	SocialniStipendium
UkonceniStudia	<u>Datum</u>	D	10	A	kontrola s atributy "ZapisDoStudia", "Zpusob", "UdelenyTitul" a elementem "PlatnostDo"	Studium
	<u>Zpusob</u>	I	2	A	číselník <u>UKONSTU</u> ;	Studium

					kontrola s atributy "Datum" a "UdelenyTitul"	
	<u>UdelenyTitul</u>	S	15	A	číselník <u>TITUL</u> kontrola s atributem "Zpusob"	Studium
<i>StudiumEtapy</i>				A		Studium
StudiumEtapa	<u>PlatnostOd</u>	D	10	A	kontrola s atributy "ZapisDoStudia" a "Datum" věčná kontrola	StudiumEtapy
<u>ObcanstviKvalifikator</u>		I	1	A	číselník <u>AA0005</u>	StudiumEtapa
<u>ObcanstviStat</u>		S	3	N	číselník <u>ISO3166</u> ; kontrola s elementem "ObcanstviKvalifikator"	StudiumEtapa
<u>PobytVCR</u>		S	1	N	"A" nebo "N"; kontrola s elementem "ObcanstviStat" (<i>pouze u studentů, kteří nejsou občany ČR</i>)	StudiumEtapa
StudijniPobyt	<u>Forma</u>	S	1	N/A ¹⁾	číselník <u>FORMASTUPOBYT</u>	StudiumEtapa
	<u>Program</u>	I	2	A	číselník <u>PROGRAMSTUPOBYT</u>	StudiumEtapa
<u>JazykVyuky</u>		S	3	A	číselník <u>Jazyk</u>	StudiumEtapa
<i>StudijniObory</i>				N		StudiumEtapa

<u>Obor (1)</u>		S	8	N	číselník <u>AKVO</u>	StudijniObory
<u>Obor (2)</u>		S	8	N	číselník <u>AKVO</u>	StudijniObory
<u>Obor (3)</u>		S	8	N	číselník <u>AKVO</u>	StudijniObory
<u>Obor (4)</u>		S	8	N	číselník <u>AKVO</u>	StudijniObory
<u>AprobaceOboru</u>				N		StudiumEtapa
<u>Aprobace (1)</u>		S	2	N	číselník <u>R2117</u>	AprobaceOboru
<u>Aprobace (2)</u>		S	2	N	číselník <u>R2117</u>	AprobaceOboru
<u>Aprobace (3)</u>		S	2	N	číselník <u>R2117</u>	AprobaceOboru
<u>MistoVyuky</u>		I	6	A	číselník <u>OBEC</u>	StudiumEtapa
<u>FormaStudia</u>		S	1	A	číselník <u>FORMASTU</u>	StudiumEtapa
<u>Financovani</u>		I	2	A	číselník <u>FINANC</u>	StudiumEtapa
<u>PreruseniStudia</u>		S	1	A	číselník <u>PRERSTU</u>	StudiumEtapa
<u>PlatnostDo</u>		D	10	N	kontrola s atributy "ZapisDoStudia", "Datum" a "PlatnostOd" věčná kontrola	StudiumEtapa

Poznámky, vysvětlivky:

1) – uvedení elementu "StudijniPobyt" je nepovinné, uvedení atributu "Forma" v rámci elementu povinné je.

Typ – String - řetězec znaků; Integer - číslo; Real - desetinné číslo; Date - datum ve formátu "rrrr-mm-dd".

Délka – maximální počet znaků řetězce, data nebo počet cifer čísla; u desetinného čísla znamená "celkový počet cifer". "počet míst za des. tečkou".

- Každé jednotlivé studium je jednoznačně definováno klíčovými atributy "RodneCislo", "VSFakulta", "StudijniProgram" a

"ZapisDoStudia". Dojde-li ke změně hodnot v některém z atributů "RodneCislo" nebo "ZapisDoStudia" (např. opravou rodného kódu na rodné číslo nebo opravou chybně vloženého data zápisu), budou tyto změny zaznamenány do aktuálního záznamu (není nutné vytvářet novou etapu studia či nové studium studenta). Dojde-li ke změně hodnot v některém z atributů "VSFakulta" nebo "StudijniProgram", bude aktuální studium ukončeno přestupem/převodem a vytvoří se nový záznam (sekvence etap) s aktualizovanými hodnotami.

- Dojde-li ke změně hodnot v elementech či attributech sekce "Student" ("Jmeno" až "RokMatZkousky"), nebo sekce "Studium" ("DelkaStudia" až "UkonceniStudia"), budou tyto změny zaznamenány do aktuálního záznamu (není nutné vytvářet novou etapu studia či nové studium studenta).
Při ukončení studia v daném studijním programu (element "UkonceniStudia") bude současně ukončena platnost aktuální etapy vyplněním elementu "PlatnostDo" stejným datem jako v atributu "Datum" elementu "UkonceniStudia".
- Dojde-li ke změně hodnot alespoň v jednom z elementů nebo atributů sekce "StudiumEtapy" ("ObcanstviKvalifikator" až "Prerusenistudia"), bude platnost dosavadní etapy ukončena datem, kdy došlo ke změně. Následně bude vytvořena nová etapa s tímto datem zahájení platnosti, obsahující aktuální - nově platné údaje. Obsah atributů "RodneCislo", "VSFakulta", "StudijniProgram" a "ZapisDoStudia" se nemění.

Popis jednotlivých elementů a atributů:

- **Rodné číslo** - uvádí se podle občanského průkazu, resp. jiného dokladu stejné úrovně průkaznosti, ve formě posloupnosti čísel (bez lomítka). Má-li rodné číslo pouze devět cifer, na desáté pozici se uvede mezera. Rodná čísla je nezbytné uvádět i u cizinců, kteří si je nechají v souladu s příslušným standardem ÚSIS ČR přidělit buď Správou sociálního zabezpečení (narození do 31.12.1968), Krajskou správou ČSÚ v Brně (narození od 1.1.1969) nebo Cizineckou policií.
Nemá-li cizinec, přiděleno rodné číslo, přidělí mu vysoká škola uměle vygenerovaný rodný kód: prvních šest znaků se vygeneruje stejně jako u rodného čísla, na 7.-8. pozici se uvede dvouciferný kód VŠ z číselníku REGPRO, na 9. pozici libovolné písmeno a na 10. pozici libovolná cifra nebo písmeno (písmena lze volit malá i velká, bez diakritiky).
Podrobnější informace o přidělování rodných čísel resp. rodných kódů studentům-cizincům jsou shrnuty v samostatném dokumentu.
- **Jméno, resp. více jmen**, - uvádí se podle občanského průkazu, resp. jiného dokladu stejné úrovně průkaznosti, ve formě: první písmena velká, ostatní malá.

Skládá-li se u cizinců jméno z několika jednotlivých jmen, která jsou si rovnocenná a jejichž pořadí není pevně stanoveno (např. "Ali Abhud" nebo "Abhud Ali"), zapíše se tolik jmen, kolik jich lze umístit do prostoru 24 znaků bez krácení, a to ta jména, která nebyla použita jako příjmení cizince. Přitom se zachovává pořadí jmen podle předloženého cestovního pasu (cestovní pas je pro ČR identifikací osoby) a jednotlivá jména jsou oddělena vždy jednou mezerou. Jestliže je jméno cizince jediné slovo delší než 24 znaků, zkrátí se na délku 24 znaků vynecháním 25. znaku a dalších znaků. Při přepisu jména cizince z cizího jazyka nepoužívajícího latinku do češtiny se používá transliterace.

- **Příjmení** - uvádí se podle občanského průkazu, resp. jiného dokladu stejné úrovně průkaznosti, ve formě: první písmena velká, ostatní malá.

Skládá-li se u cizinců příjmení z několika jednotlivých jmen, která jsou si rovnocenná a jejichž pořadí není pevně stanoveno (např. "Ali Mohammed" nebo "Mohammed Ali"), zapíše se tolik jmen, kolik jich lze umístit do prostoru 35 znaků bez krácení. Přitom se zachovává pořadí jmen podle předloženého cestovního pasu (cestovní pas je pro ČR identifikací osoby) a jednotlivá jména jsou oddělena vždy jednou mezerou. Jestliže je jméno cizince jediné slovo delší než 35 znaků, zkrátí se na délku 35 znaků vynecháním 36. znaku a dalších znaků. Při přepisu jména cizince z cizího jazyka nepoužívajícího latinku do češtiny se používá transliterace.

- **Rodné příjmení** - uvádí se podle občanského průkazu, resp. jiného dokladu stejné úrovně průkaznosti, ve formě: první písmena velká, ostatní malá.
U cizinců je postup stejný jako u příjmení.
- **Titul před jménem** - uvádí se v souladu s příslušným standardem ÚSIS ČR.
Uvede se zkratka stavovského označení, akademického titulu nebo ostatních titulů absolventa vysoké školy. Neexistuje-li tato informace, položka se ponechá nevyplněna. U zahraničních titulů se uvede zkratka uvedená v zahraničních dokladech o ukončení studia.
- **Titul za jménem** - uvádí se v souladu s příslušným standardem ÚSIS ČR.
Uvede se zkratka vědecké hodnosti nebo akademického titulu. Neexistuje-li tato informace, položka se ponechá nevyplněna. U zahraničních titulů se uvede zkratka uvedená v zahraničních dokladech o ukončení studia.

Element "Adresa" obsahuje adresu trvalého pobytu studenta. V případě cizinců s povoleným trvalým pobytem na území ČR může touto adresou být jeho trvalé bydliště v České republice. Položky identifikující trvalé bydliště studenta jsou ve struktuře především za účelem demografických analýz a verifikace sociálních příspěvků (ubytovací stipendium).

- **Okres** - uvádí se číselný kód podle číselníku **OKRES**.

- **Obec** - uvádí se číselný kód podle číselníku **OBEC**. Neexistuje-li tato informace, položka se ponechá nevyplněna.
 - **Část obce** - uvádí se číselný kód podle číselníku **COBCE**. Neexistuje-li tato informace, položka se ponechá nevyplněna.
 - **Ulice** - vyplní se textem názvu ulice. Neexistuje-li tato informace, položka se ponechá nevyplněna.
 - **Číslo** - vyplní se podle povahy adresy číslem domovním, popisným, orientačním, resp. evidenčním. Neexistuje-li tato informace, položka se ponechá nevyplněna.
 - **PSČ** - uvádí se podle číselníku České pošty **ZV_PSCPO**. Neexistuje-li tato informace, položka se ponechá nevyplněna.
 - **Stát** - uvádí se číselný kód podle číselníku **ISO3166**.
-
- **Rodinný stav** - uvádí se číselný kód na základě sdělení studenta (povinnost sdělit tento údaj vyplývá z par. 88 odst. 2 zákona) podle číselníku **K2101**.
 - **Stupeň předchozího vzdělání** - uvádí se znakový kód jako aktuální nejvyšší dosažené vzdělání studenta podle číselníku **K21K1**.
 - **Střední škola** - uvádí se číselný kód podle maturitního vysvědčení školy, na které student složil maturitu, podle číselníku **IZO**. Neexistuje-li tato informace, položka se ponechá nevyplněna.
 - **Rok maturitní zkoušky** - uvádí se podle maturitního vysvědčení (ve specifických případech, kdy student nemá maturitu, se položka nevyplňuje).
Předchozí dvě položky jsou ve struktuře především z důvodů vyhodnocení úspěšnosti absolventů, jakožto jednoho z kritérií pro hodnocení a vzájemné srovnávání středních škol.
 - **Studovaná škola/fakulta** - uvádí se číselně-znakový kód podle číselníku **REGPRO** té fakulty vysoké školy, která uskutečňuje příslušný studijní program. Pokud se vysoká škola nedělí na fakulty, uvádí se patřičný 5-ti místný kód s devítkou, např. 72900.
 - **Studijní program** - uvádí se číselně-znakový kód podle číselníku **STUDPROG**.
 - **Zápis do studia** - uvádí se zadáním data prvního zápisu do studia příslušného studijního programu na dané škole/fakultě (par. 61 odst. 1 zákona) ve tvaru rrrr-mm-dd.
 - **Délka studia** uvádí se desetinným číslem podle standardní doby studia příslušného studijního programu dle Statutu vysoké školy (např. 4.0 nebo 5.5).
 - **Student nově přijatý** - uvádí se zadáním "A", jedná-li se o jeho první přijetí na vysokou školu v ČR, a to do úvodní etapy studia (par. 61 odst. 1 zákona). Student může být označen jako nově přijatý v bakalářském a magisterském studijním programu pouze v prvním roce studia (tedy neuplyne-li od data zápisu více jak 365 dní). V ostatních případech se uvádí

"N".

- **Navazující studijní program** - uvádí se "A", jedná-li se o navazující studijní program ve smyslu par. 58 odst. 4, v opačném případě se uvádí "N". "A" se uvádí pouze tehdy, jedná-li se o magisterský studijní program navazující na bakalářský studijní program. U jiných typů studia se nevyplňuje.
Položka identifikuje skutečnost, zda student, který již absolvoval vysokou školu (zpravidla bakalářské studium), studuje či nestuduje v navazujícím magisterském studiu, a tedy se jej netýkají či týkají poplatky za studium.
- **Počet ročníků** - uvádí se číslo vyjadřující maximální počet ročníků, na něž se dělí studium příslušného studijního programu (uvádí se pouze v případech, je-li studium děleno na ročníky; jinak se nevyplňuje).
- **Aktuální ročník** - uvádí se číslo vyjadřující označení ročníku, v němž student aktuálně studuje (uvádí se pouze v případech, je-li studium děleno na ročníky; jinak se nevyplňuje).
- **Ubytování v koleji** - uvádí se číselný kód podle číselníku **UBYT**.
- **Sociální stipendium – Nárok od** - uvádí se datum vzniku nároku na přídavek na dítě ve zvýšené míře na základě oznámení o přiznání přídatku předloženém studentem.
- **Sociální stipendium – Nárok do** - uvádí se datum ukončení nároku na přídavek na dítě ve zvýšené míře.
- **Ukončení studia – datum** - uvádí se zadáním data ukončení studia příslušného studijního programu ve tvaru rrrr-mm-dd (neexistuje-li tato informace, položka se ponechá nevyplněna).
- **Ukončení studia – způsob** - uvádí se číselný kód podle číselníku **UKONSTU** (neexistuje-li tato informace, položka se ponechá nevyplněna).
- **Ukončení studia – titul** - uvádí se znakový řetězec v souladu se Statutem vysoké školy podle číselníku **TITUL** (neexistuje-li tato informace, položka se ponechá nevyplněna).

- **Platnost od** - uvádí se datum zahájení platnosti etapy studia ve formátu rrrr-mm-dd. Obsahuje
 - buď datum změny, která nastala v jednom z elementů nebo atributů sekce "StudiumEtapa";
 - nebo datum zápisu do studia, jedná-li se o první etapu studia dané sekvence.
- **Občanství – kvalifikátor** - uvádí se číselný kód podle číselníku **AA0005**.
- **Občanství – kód státu** - uvádí se číselný kód podle číselníku **ISO3166**.
- **Trvalý pobyt v ČR** uvádí se zadáním "A", má-li student povolen trvalý pobyt v ČR (má smysl pouze u studentů, kteří nejsou občany ČR); v opačném případě se uvádí "N".
- **Studijní pobyt – forma** - uvádí se znakový kód formy krátkodobého studijního pobytu podle číselníku **FORMASTUPOBYT**.

- **Studijní pobyt – program** - uvádí se číselný kód financujícího programu krátkodobého studijního pobytu podle číselníku **PROGRAMSTUPOBYT**.
- **Jazyk výuky** - uvádí se znakový řetězec jazyka výuky studijního programu podle číselníku **JAZYK**.
- **Studijní obor (1)-(4)** - uvádí se číselně-znakový kód podle číselníku **AKVO**, přičemž nezáleží na pořadí oborů (neexistuje-li tato informace, položka se ponechá nevyplněna).
- **Aprobace (1)-(3)** - uvádí se číselný kód podle číselníku **R2117**, přičemž nezáleží na pořadí aprobací (neexistuje-li tato informace, položka se ponechá nevyplněna).
- **Místo výuky** - uvádí se číselný kód obce podle číselníku **OBEC**.
Položka slouží k verifikaci sociálních příspěvků (ubytovací stipendium).
- **Forma studia** - uvádí se (je dána Statutem vysoké školy) znakový kód podle číselníku **FORMASTU**.
- **Financování** - uvádí se číselný kód podle číselníku **FINANC**.
Položka je nezbytným základem pro stanovení způsobu a výše financování studia studenta v daném studijním programu na dané škole/fakultě.
- **Přerušění studia** - uvádí se znakový kód podle číselníku **PRERSTU**.
Položka identifikuje časové úseky, kdy student nestudoval. Slouží rovněž k verifikaci financování.
- **Platnost do** - uvádí se datum ukončení platnosti etapy studia ve formátu rrrr-mm-dd. Obsahuje
 - buď datum změny, která nastala v jednom z elementů nebo atributů sekce "StudiumEtapa";
 - nebo datum ukončení studia, jedná-li se o poslední větu definitivně ukončující danou sekvenci.

Současně se rovná obsahu atributu "PlatnostOd" v bezprostředně následující etapě (existuje-li taková etapa).

P.2 Přehled indikátorů prezentovaných Ústavem pro informace ve vzdělání

(Zdroj www.uiv.cz)

Souhrn VŠ – studenti (fyzické osoby) podle formy a typu studijního programu – podle státního občanství a zřizovatele

Souhrn VŠ – počet VŠ, studií a absolvování podle formy a typu studijního programu – podle státního občanství a zřizovatele

Souhrn VŠ – školy, fakulty, studenti (fyzické osoby), zaměstnanci – podle zřizovatele a vysoké školy

Souhrn VŠ – studenti (fyzické osoby) – podle zřizovatele a města

VŠ – studenti (fyzické osoby) podle formy a typu studijního programu – podle zřizovatele a vysoké školy

VŠ – studenti (fyzické osoby) ženy podle formy a typu studijního programu – podle zřizovatele a vysoké školy

VŠ – studenti (fyzické osoby) celkem podle formy studia a typu studijního programu – podle zřizovatele a skupiny studijních programů

VŠ – studenti (fyzické osoby) ženy podle formy studia a typu studijního programu – podle zřizovatele a skupiny studijních programů

VŠ – studenti (fyzické osoby) ve všech formách a typech studia podle zřizovatele – podle státního občanství

VŠ – studenti (fyzické osoby) v typu studijního programu – podle zřizovatele a roku narození

VŠ – studenti (fyzické osoby) podle trvalého bydliště – podle území a vysoké školy

VŠ – výuka cizího jazyka – studenti (počet studií), učitelé – podle cizího jazyka

VŠ – poprvé zapsaní (fyzické osoby) na VŠ podle formy studia a typu studijního programu – podle zřizovatele a vysoké školy

VŠ – poprvé zapsaní (fyzické osoby) na VŠ celkem podle formy studia a typu studijního programu – podle zřizovatele a skupiny studijních programů

VŠ – poprvé zapsané ženy (fyzické osoby) na VŠ podle formy studia a typu studijního programu – podle zřizovatele a skupiny studijních programů

VŠ – poprvé zapsaní (fyzické osoby) – podle typu studijního programu – podle zřizovatele a roku narození

VŠ veřejné a soukromé – přijímací řízení (uchazeči se státním občanstvím ČR) – podle zřizovatele, vysoké školy

VŠ veřejné a soukromé – přijímací řízení (uchazeči se státním občanstvím ČR) – muži/ženy – podle formy studia a typu studijního programu

VŠ veřejné a soukromé – přijímací řízení (uchazeči se státním občanstvím ČR) – podle formy studia a zřizovatele

VŠ veřejné a soukromé – přijímací řízení (uchazeči se státním občanstvím ČR) – podle formy studia a skupiny studijních programů

VŠ veřejné a soukromé – přijímací řízení (uchazeči se státním občanstvím ČR) – podle zřizovatele a skupiny studijních programů

VŠ veřejné a soukromé – přijímací řízení (uchazeči se státním občanstvím ČR) – podle formy studia a předchozího působiště

Vysoké školy – absolventi (fyzické osoby), z toho ženy podle formy studia a typu studijního programu – podle zřizovatele a VŠ

Vysoké školy – absolventi (fyzické osoby) celkem, z toho ženy podle formy studia a typu studijního programu – podle zřizovatele a skupiny studijních programů

Vysoké školy – absolventi (fyzické osoby) – podle typu studijního programu – podle zřizovatele a roku narození

Vysoké školy – celoživotní vzdělávání – studenti (počet studií) – podle zřizovatele a vysokých škol

Vysoké školy – celoživotní vzdělávání – absolventi (úspěšně ukončená studia) – podle zřizovatele a vysokých škol

Veřejné vysoké školy – zaměstnanci – podle vysokých škol

Státní vysoké školy – školy, nově přijatí, studenti (počet studií), absolventi (počet absolvování) podle formy studia a typu studijního programu – podle státního občanství a zřizovatele

Státní vysoké školy – studenti (počet studií), nově přijatí, absolventi (počet absolvování) podle formy studia a typu studijního programu – podle zřizovatele vysoké školy a fakulty

Státní vysoké školy – studenti (počet studií) celkem podle formy studia a typu studijního programu – podle zřizovatele a skupiny studijních programů

Státní vysoké školy – nově přijatí celkem, podle formy studia a typu studijního programu – podle zřizovatele a skupiny studijních programů

Státní vysoké školy – absolventi (počet absolvování) celkem, podle formy studia a typu studijního programu – podle zřizovatele a skupiny studijních programů

P.3 Základní údaje obsažené v databázi Uchazeč

Databáze Uchazeč obsahuje data o všech přihláškách na vysoké školy odeslaných v daném školním roce:

- Název vysoké školy a fakulty, na kterou se uchazeč hlásí
- Typ studia, na který se uchazeč hlásí (Bc., Mgr., Ph.D.)
- Forma studia (Prezenční, Distanční, Kombinované)
- Studijní program
- Studijní obor
- Datum podání (registrace) přihlášky
- Výsledek přijímacího řízení (a datum rozhodnutí)
- Výsledek zápisu (a datum zápisu)
- Rodné číslo (pohlaví + datum narození: možnost propojení s databází SIMS)
- PSČ bydliště
- státní příslušnost uchazeče
- Působení uchazeče v době podání přihlášky (střední škola, zaměstnání, domácnost, apod.)
- Identifikátor střední školy
- Obor studia na střední škole
- Rok maturity

Zdroj: <http://dsia.uiv.cz/uch2008b.html>

P.4 Modelování pravděpodobnosti úspěšného dokončení studia

Jednou z otázek, kterou je možné se v rámci tohoto projektu zabývat, je i to, jaká je pravděpodobnost dokončení studia jednotlivých oborů a co tuto pravděpodobnost ovlivňuje. Tato otázka je zásadní pro všechny, kteří se celého procesu vysokoškolského vzdělávání účastní, nebo to mají v úmyslu. Absolventy středních škol při výběru vysoké školy zajímá, jaké jsou jejich šance vybranou školu či studijní obor dokončit. Obdobný problém trápí i studenty nižších ročníků, kteří již na daném oboru studují, nebo ty, kteří přecházejí z bakalářského na magisterské studium. Stejně tak i samotné vysoké školy se snaží vyhodnotit, jaké faktory jsou pro studenty kritické, a buď tyto faktory změnit, nebo jim lépe přizpůsobit výběrové řízení mezi uchazeči. A příslušné instituce chtějí mít co nejlepší přehled o charakteru výuky poskytované na vysokých školách.

Studovat pravděpodobnost úspěšného dokončení studia nám umožní metoda lineární regrese. Jedná se o odvětví statistiky, které se zabývá právě odhalováním závislostí mezi různými veličinami. V tomto případě budeme zkoumat, jak to, zda student vybraný obor na vybrané vysoké škole dokončí či nikoli, závisí na jeho osobních charakteristikách a na charakteristikách daného oboru. Za tímto účelem zvolíme probitový model, jehož parametry získáme z dat studentů, kteří již studium ukončili (ať už úspěšně či neúspěšně) - tedy z dat, která máme z databází. Technické detaily modelu jsou uvedeny v technické části této přílohy, i bez těchto detailů se dá ovšem snadno nahlédnout, jaké výstupní informace můžeme od modelu očekávat, jaký typ vstupních dat pro něj budeme potřebovat a s jakými problémy se bude třeba vypořádat.

Výstupní informace

Regresní model poskytuje dva různé typy výstupu, z nichž každý má trochu jiné informační zaměření. Prvním výstupem je odhad parametrů modelu, prováděný na základě údajů o studentech, kteří své studium již ukončili, ať už úspěšně či neúspěšně. Parametry modelu jsou čísla, která nás mohou informovat o vlivu jednotlivých

charakteristik na celkovou pravděpodobnost úspěšného dokončení studia. Každé jednotlivé charakteristice (ať už studenta nebo školy), kterou do modelu zahrneme, přísluší jeden číselný parametr. Čím je hodnota tohoto parametru větší, tím větší je pravděpodobnost úspěchu. Kladné parametry proto ovlivňují tuto pravděpodobnost pozitivně a ty záporné negativně³. Z toho je patrné, že je to právě tento první výstup, který nám umožní pochopit mechanismy, kterými se studium řídí, tzn. odhalit charakteristiky, které mohou být pro studenty kritické, nebo naopak ty, které jim studium usnadňují. Tento výstup je proto určen především vysokým školám a příslušným institucím, které na systém vysokoškolského vzdělávání dohlíží.

Druhým možným výstupem regresního modelu je výpočet pravděpodobnosti dokončení vybraného oboru pro daného studenta. Tento dynamický výstup by mohl být implementován v podobě jednoduchého programu (webové aplikace), do kterého by student zadal své osobní charakteristiky a název vybraného oboru a školy, případně více oborů, mezi kterými si vybírá a chtěl by je srovnat. Program by potom vypočítal pravděpodobnost, s jakou daný student tento obor dokončí, případně by ji porovnal s pravděpodobností dokončení jiného oboru, či nějakého příbuzného oboru na jiné vysoké škole. Tento výstup se dá navíc přizpůsobit dvěma základním situacím – jednak pro uchazeče, kteří se teprve rozmyšlí, pro jakou vysokou školu se rozhodnout, a jednak pro studenty vysokých škol, kteří by rádi vyhodnotili své šance daný obor dokončit.

Typy vstupních dat

Zamyslíme-li se nad celým problémem čistě intuitivně, je zřejmé, že úspěch či neúspěch studia je ovlivněn dvěma základními skupinami faktorů. Prvním jsou charakteristiky daného oboru a školy, druhým jsou charakteristiky každého studenta (jinak řečeno, záleží na tom, nakolik se daný student hodí pro studium daného oboru a naopak, nakolik se daný obor hodí pro tohoto studenta). Vzhledem k charakteru databází, které máme k dispozici, budou mezi našimi daty převládat spíše charakteristiky oborů, neboť

³ Tento vztah není v našem modelu lineární – pro bližší vysvětlení jakou měrou parametr pravděpodobnost ovlivňuje viz. příloha.

informací specifických pro jednotlivé studenty je v databázích relativně málo. Toto považujeme za jeden z největších problémů, a proto ho později rozebereme podrobněji.

Prozatím si uveďme, které z charakteristik uvedených v databázích bychom mohli obecně použít, později si upřesníme, jak konkrétně sestavíme model pro různé otázky, které můžeme v souvislosti s pravděpodobností dokončení studia zkoumat.

Z charakteristik studenta můžeme vybrat především

- věk a pohlaví (lze zjistit z rodného čísla)
- národnost
- region, ze kterého student pochází
 - tato charakteristika by se také dala použít ve vztahu k regionu, ve kterém se nachází zvažovaná vysoká škola, neboť některé studie naznačují, že delší „dojezdnost“ negativně ovlivňuje úspěšnost studia
- předchozí dosažené vzdělání
 - zde by bylo vhodné mít informace o konkrétní střední škole a také o roku maturitní zkoušky (lze předpokládat, že delší proluka mezi maturitou a nástupem na VŠ bude pravděpodobnost úspěchu snižovat) – bohužel, oba tyto údaje jsou nepovinné
 - také by jistě bylo přínosné mít k dispozici údaje o výsledku maturitní zkoušky, bohužel tyto v databázi chybí úplně
- předchozí studium na dané vysoké škole
 - týká se jen studentů, kteří již daný obor jistou dobu studují
 - zde by bylo vhodné mít i kvalitativní data – tedy údaje o prospěchu daného studenta (zatím v databázích není)
- předchozí neúspěšné studium na vysoké škole
- sociální údaje

Z charakteristik vysoké školy/oboru pak vybereme

1. obecné charakteristiky
 - region (město)

- forma studia
 - počty studentů po ročnících
2. charakteristiky vystihující „atraktivitu“ školy – tedy její kvalitu a také schopnost motivovat studenty pro pokračování studia
 - počty přerušení
 - počty studentů z kraje působnosti školy
 - přestupnost na N. Mgr.
 - přestupy mezi školami
 3. charakteristiky vystihující snadnost/obtížnost oboru
 - počet duplicitních studentů
 - relativní délka studia vzhledem k akreditované délce
 - průměrná délka neúspěšného studia
 - poměr počtu neúspěšných a úspěšných studentů

Zde je však třeba poznamenat, že z jistých technických omezení metod lineární regrese nebude většinou možné do modelu zahrnovat jednotlivé charakteristiky oboru (a tím pádem studovat jejich vliv na pravděpodobnost dokončení studia), ale pouze indikátorové proměnné zastupující jednotlivé obory, tak jak bylo popsáno v sekci 4.3.2.

Možné problémy

S implementací modelu tohoto typu je spojeno několik technických problémů, které jsou blíže rozvedeny níže v popisu modelu. Pokusme se však osvětlit zde alespoň některé základní body, ve kterých spatřujeme možná úskalí.

Nepozorované a chybějící charakteristiky

Jak již bylo řečeno výše, za základní problém považujeme nedostatečné údaje o individuálních charakteristikách studenta. Je zřejmé, že osobní dovednosti daného studenta mají na jeho úspěch na vysoké škole zásadní vliv. Dovednosti jsou pochopitelně značně problematicky měřitelné charakteristiky, v našich databázích však naprosto chybí

jakékoli veličiny, které by osobní dovednosti alespoň přibližovaly. Částečně by se schopnosti studentů daly odvodit z typu střední školy, kterou studovali, ale i to je údaj pouze nepovinný. Další možné údaje, jako je například alespoň výsledek maturitní zkoušky, chybí úplně. Proto se domníváme, že by bylo vhodné databáze o tyto údaje doplnit.

Studie zabývající se úspěšností studentů na vysokých školách dále naznačují, že ani údaje o předchozích školních výsledcích nemusí být postačující pro správné vyhodnocení pravděpodobnosti úspěchu na vysoké škole. Jako další významné faktory ovlivňující tuto pravděpodobnost je zmiňováno především rodinné prostředí, tedy vzdělání a povolání rodičů, počet sourozenců a další. Tyto údaje jsou ovšem daleko nad rámec našich databází, je však třeba si uvědomit, že tento problém bude negativně ovlivňovat přesnost našich závěrů. Pokud by bylo možné databázi propojit s nějakými dalšími informačními zdroji, které by podobné údaje doplnily, bylo by možné dosáhnout lepších výsledků.

Kromě dat o studentech by bylo možné zvážit též doplnění informací o vysokých školách na základě různých hodnocení, které provádí státní instituce i soukromé agentury.

Počet pozorování

Dalším problémem pro úspěšnou implementaci modelu by mohlo být nedostatečné množství dat, zejména u menších a mladších oborů, které ještě nejsou sledovány po dostatečně dlouhou dobu a nemají příliš velký počet studentů a absolventů. Všechny statistické metody použité v našem modelu jsou totiž tím přesnější, čím více údajů máme k dispozici, a proto je třeba mít nebezpečí malého množství dat na paměti.

Zkreslení výběru vzorku

Posledním velmi důležitým problémem je také nebezpečí nesprávného odhadu parametrů a tedy nesprávné určení kauzality v modelu dané výběrem pozorovaných jednotlivců, který máme k dispozici. Správně formulovaný model vyžaduje, aby studenti, jejichž údaje zkoumáme, byli náhodně vybráni ze všech těch, kteří by potenciálně mohli

studovat vysokou školu. V našem případě tomu tak není, protože úspěšnost pochopitelně zkoumáme pouze u těch studentů, kteří se vůbec rozhodli pro studium na vysoké škole. Tento problém je poněkud techničtějšího rázu, a proto ho i s možnými návrhy řešení přibližujeme spíše v příloze.

Konkrétní příklady možných modelovaných situací a popis modelů

1. Zkoumání pravděpodobnosti dokončení konkrétního oboru na konkrétní škole

Tato otázka by byla zajímavá pro rozhodujícího se studenta a také pro danou vysokou školu. Lze ji modifikovat pro tři základní případy, a to pro bakalářský obor, pro navazující magisterský obor a pro oba najednou.

a) bakalářský obor – z pohledu absolventa střední školy (který za sebou může mít předchozí neúspěšné studium na jiné VŠ)

Zkoumáme pravděpodobnost úspěšného dokončení bakalářského oboru v závislosti na všech výše zmíněných charakteristikách studenta kromě předchozího úspěšného studia na VŠ (zahrnujeme pouze proměnnou vyjadřující, zda daný student již nějaké studium započal a nedokončil).

Pro odhad parametrů používáme data studentů, kteří na daném bakalářském oboru někdy studium započali, a již je definitivně rozhodnuto, zda ho úspěšně dokončili či školu definitivně opustili.

b) navazující magisterský obor – z pohledu bakaláře

Zkoumáme pravděpodobnost úspěšného dokončení magisterského oboru v závislosti na všech výše zmíněných charakteristikách studenta, včetně specifikace předchozího bakalářského studia (zda proběhlo na stejné VŠ, zda bylo dokončeno v řádném termínu apod.).

Pro odhad parametrů používáme data studentů, kteří na daném magisterském oboru někdy studium započali, a již je definitivně rozhodnuto, zda ho úspěšně dokončili či školu definitivně opustili.

c) bakalářský + magisterský obor – z pohledu absolventa střední školy

Zkoumáme pravděpodobnost úspěšného dokončení bakalářského i navazujícího magisterského oboru v závislosti na všech výše zmíněných charakteristikách studenta.

Pro odhad parametrů používáme data studentů, kteří na daném oboru někdy započali bakalářské studium, a již je definitivně rozhodnuto, zda ho úspěšně dokončili magisterské studium či školu definitivně opustili (ať už v průběhu bakalářského či magisterského studia).⁴

Jakkoli je tato otázka velice zajímavá, jejím velkým problémem jsou nedostatečné údaje. Vzhledem k tomu, že zkoumáme jeden daný obor, nehrají pro vysvětlení rozdílů v úspěšnosti charakteristiky oboru žádnou roli a vše je založeno jen na rozdílných charakteristikách studentů, které jsou, jak již bylo řečeno, v databázích nedostatečně zastoupeny. Proto by tento přístup připadal v úvahu, pokud by se podařilo rozšířit údaje o osobních charakteristikách studentů v databázích. Dalším problémem tohoto přístupu by mohl být též malý počet dat, zejména pro menší obory.

2. Porovnání několika oborů, mezi kterými si uchazeč vybírá

Tato otázka je zajímavá pro rozhodujícího se studenta a také pro vysoké školy a příslušné instituce pro přehled o konkurenci mezi školami.

Co se týče charakteristik studenta, otázku můžeme opět modifikovat pro bakalářské obory i pro navazující magisterské obory. V tomto směru se jedná o obdobný problém jako v bodech 1. a), b) a použijí se stejné charakteristiky studentů. Jelikož však porovnáваме i jednotlivé obory, použijeme i charakteristiky oboru.

V tomto směru musíme otázku dále specifikovat jako

- a) rozhodování mezi různými obory na jedné vysoké škole (například francouzská filologie oproti překladatelství z francouzštiny na FF UK v Praze)

⁴ Tento model je poněkud speciální a vyžaduje o něco komplikovanější strukturu než ostatní. Detaily viz příloha.

- b) rozhodování mezi několika příbuznými obory na různých školách (například francouzština v Praze, Brně či Olomouci)
- c) kombinaci obojího (překladařství z francouzštiny v Praze oproti klasické filologii v Brně)

Pro odhad parametrů použijeme údaje o studentech, o jejichž úspěchu či neúspěchu je již definitivně rozhodnuto a to pro všechny porovnávané obory. Tím také redukuje problém nedostatku dat.

Počet porovnávaných oborů může být různý. Je také zřejmé, že počet kombinací pro takovéto porovnávání je velmi velký.

3. Obecná pravděpodobnost dokončení studia na libovolné vysoké škole

Tato otázka je vlastně rozšířením bodu 2. pro všechny obory. Spíše než pro uchazeče stojícího před konkrétní volbou je zajímavá ze sociologického hlediska.

Lze si samozřejmě pokládat mnoho dalších podobných otázek, kombinovat je, modifikovat je pro studenty různých ročníků a podobně. Pro každou z těchto otázek je pak třeba prozkoumat možné problémy a omezení daná charakterem dat.

Popis modelu

Základem probitového modelu je lineární rovnice (1) zahrnující vysvětlující veličiny a parametry modelu, doplněná o soustavu nerovnic (2) umožňující rozhodnout o hodnotě vysvětlované veličiny.:

$$Y_i^* = \beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \dots + \beta_k X_i^k + \varepsilon_i \quad i = 1, \dots, n \quad (1)$$

$$Y_i = \begin{cases} 1 & \text{pokud } Y_i^* > 0 \\ 0 & \text{pokud } Y_i^* \leq 0 \end{cases} \quad i = 1, \dots, n \quad (2)$$

Veličina Y_i zde vyjadřuje hodnotu naší vysvětlované veličiny, respektive její binární kódování (1 pro dokončené studium, 0 pro nedokončené studium), veličina Y_i^* je tzv. latentní veličina modelu. Tuto veličinu nelze pozorovat, můžeme si pod ní však představovat například užítkovou funkci vyjadřující to, nakolik je pro studenta přínosné (a možné) pokračovat ve studiu daného oboru. Pokud je tento užitek záporný, studium bude ukončeno.

Veličiny X_i^j vyjadřují hodnoty vysvětlujících veličin (charakteristiky studenta a školy) a β_j jsou parametry našeho modelu, které je třeba statistickými metodami odhadnout. Veličina ε_i je stochastická chyba, zahrnující všechny vlivy, které neumíme určit a považujeme je proto za náhodné prvky. Tato náhodná veličina má podle předpokladů modelu normální rozdělení s distribuční funkcí

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

kteřá je pro naše účely velice důležitá, neboť se dá použít pro výpočet pravděpodobnosti, že uchazeč vybraný obor skutečně dokončí. Označíme-li Π pravděpodobnost, že uchazeč studium dokončí, dále ve shodě s výše zmíněným modelem X^j vysvětlujících veličiny (charakteristiky studenta a školy) a β_j parametry našeho modelu, lze ukázat, že

$$\Pi = \Pr(Y_i^* > 0) = F(\beta_0 + \beta_1 X^1 + \beta_2 X^2 + \dots + \beta_k X^k) \quad (3).$$

Z tohoto vzorce je zřejmé, proč je naší snahou odhadnout parametry β_j : umožní nám pro každého nového uchazeče o studium vypočítat, s jakou pravděpodobností studium zvoleného oboru dokončí, a vysvětlit, jak tato pravděpodobnost závisí na charakteristikách studenta a oboru. Odhad parametrů se provádí metodou maximální věrohodnosti, věrohodnostní funkce je definována jako

$$L = \prod_{i=1}^n \Pr(Y_i^* \leq 0)^{1-Y_i} \Pr(Y_i^* > 0)^{Y_i}.$$

Výstup modelu

Probitový model nám poskytuje dva základní druhy výstupu. Prvním výstupem jsou přímo odhadnuté parametry modelu, které určují vliv jednotlivých proměnných na pravděpodobnost úspěšného dokončení studia. Na rozdíl od standardní regrese však v probitovém modelu není tento vztah lineární, proto statistické programy zpracovávající probitový model poskytují také derivace pravděpodobnosti podle vysvětlujících proměnných, které přímo udávají, jakým způsobem a jakou měrou změna ve vysvětlující proměnné ovlivní pravděpodobnost úspěchu. Druhým výstupem je odhad hledané pravděpodobnosti úspěšného dokončení studia pomocí odhadnutých parametrů modelu a vzorce (3). Tento výstup tedy slouží především pro predikci.

Problémy modelu

Kromě možných problémů s nedostatečným počtem pozorování a s charakteristikami, které v databázích chybí, ačkoli jsou dle našeho názoru potřebné, jsou hlavními úskalími našeho modelu problém nepozorovaných charakteristik a problém výběrového zkreslení,

Problém nepozorovaných charakteristik

Jak již bylo řečeno, lze intuitivně očekávat, že jednou z charakteristik, které budou ovlivňovat úspěšnost studenta na VŠ, jsou jeho osobní dovednosti. Tyto dovednosti jsou obtížně měřitelné, a proto nelze očekávat, že bychom je někdy v nějaké databázi mohli mít zachyceny. Proto se snažíme tyto charakteristiky aproximovat jinými veličinami, které dovednostem co nejlépe odpovídají. Takovými veličinami by mohly být například údaje o prospěchu na střední škole.⁵ Je ovšem třeba mít na paměti, že tyto „zástupné“ charakteristiky nikdy nebudou vystihovat nepozorované dovednosti tak přesně, jak bychom si představovali. Tento problém bude působit zkreslení našeho odhadu pro

⁵ Tyto údaje sice v databázích také nemáme, nicméně to je problém úplně jiného druhu, protože takovéto údaje se jistě dají zjistit a doplnit, na rozdíl od dovedností, které jsou prakticky nezměřitelné.

parametry ostatních proměnných v modelu, které je pak pro jednotlivé proměnné třeba vzít v úvahu.

Problém výběrového zkreslení a jeho možná řešení

Problém výběrového zkreslení se týká toho, že parametry modelu neodhadujeme na náhodně vybraném vzorku, ale na datech studentů, kteří se rozhodli pro studium vysoké školy (případně přímo konkrétního oboru) a byly na ni přijati.

Je zřejmé, že mnoho charakteristik, které ovlivňují úspěšnost studia na vysoké škole, také ovlivňují rozhodnutí studenta se na příslušnou vysokou školu přihlásit a rozhodnutí školy příslušného studenta přijmout. Nebo by to tak alespoň mělo být, pokud předpokládáme, že studenti dovedou racionálně ohodnotit své studijní předpoklady a také že vysoké školy pořádají přijímací řízení tak, aby vybrali uchazeče, kteří budou mít největší šanci školu dokončit.

Z toho vyplývá, že pokud máme při studiu úspěšnosti na VŠ k dispozici pouze údaje o studentech, kteří již vysokou školu studují, ztrácí se značná část informace, protože vlastně studujeme pouze skutečně realizované úspěchy a neúspěchy a ve výběru nám chybí potenciální výsledky těch, kteří se sice na tuto školu nepřihlásili nebo nedostali, ale jejichž charakteristiky by jejich úspěšnost ovlivňovaly též.

Samotný fakt, že některá pozorování chybí, by nemusel být tak závažný, kdyby tato pozorování chyběla „náhodně“, tedy kdyby nezávisela na studovaných charakteristikách. Právě v našem případě tomu však tak není, neboť některé charakteristiky ovlivňující úspěšnost studia ovlivňují zároveň rozhodnutí o nástupu na vysokou školu. Tento fakt působí právě tzv. výběrové zkreslení a následné vychýlení našich odhadů. Jinak řečeno, parametry tohoto modelu nebudou odhadnuty správně a budou vychýleny na jednu či druhou stranu.

Představme si například, že jednou ze studovaných charakteristik je socioekonomický status studenta. Dá se předpokládat, že tento faktor negativně ovlivňuje jak možnost nástupu na vysokou školu tak i úspěšnost při studiu, neboť studenti z lépe situovaných rodin mají pro studium lepší podmínky. Ovšem zároveň se dá také předpokládat, že pokud se student ze sociálně znevýhodněné rodiny na vysokou školu přihlásí, může to signalizovat, že má výrazně vyšší osobní dovednosti nebo motivaci, a to naopak zvýší pravděpodobnost, že vysokou školu dokončí. Kdybychom tento fakt nevzali v potaz a studovali pouze dopad socioekonomického faktoru na úspěšnost na VŠ, náš model by mu přiřadil příliš velkou váhu. Podobné to bude u dalších charakteristik.

Pokud bychom tedy ignorovali problém výběrového zkreslení a provedli odhad pomocí výše popsaného jednoduchého probitového modelu, naše výsledky nebudou přesné. Bohužel, zároveň se ukazuje, že řešení problému výběrového zkreslení v našem případě není zrovna triviální. Obdobný problém řeší autoři článku *Determinants of University drop-out probability in Italy* (Cingano, Cipollone, 2003), kde se k základnímu modelu pro úspěšnost studia přidává druhý, popisující zda se student rozhodne či nerozhodne na vysokou školu nastoupit.

Pokud bychom stejný přístup aplikovali na naši situaci, náš model by nyní vypadal takto:

$$Y_i^* = \beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \dots + \beta_k X_i^k + \varepsilon_i \quad , \quad Y_i = \begin{cases} 1 & \text{pokud } Y_i^* > 0 \\ 0 & \text{pokud } Y_i^* \leq 0 \end{cases}$$

$$W_i^* = \gamma_0 + \gamma_1 Z_i^1 + \gamma_2 Z_i^2 + \dots + \gamma_j Z_i^j + \eta_i \quad , \quad W_i = \begin{cases} 1 & \text{pokud } W_i^* > 0 \\ 0 & \text{pokud } W_i^* \leq 0 \end{cases}$$

Veličina W popisuje, zda student na vysokou školu nastoupil ($W = 1$ pokud ano, $W = 0$ pokud ne) a veličina Y popisuje úspěšnost při následném studiu ($Y = 1$ pro úspěch, $Y = 0$ pro neúspěch). Náhodné chyby ε a η zde mají sdružené dvourozměrné normální rozdělení, který vyjadřuje, že přijetí na vysokou školu a úspěch při následném studiu jsou provázané. Charakteristiky X a Z mohou být obecně různé, ale mohou se také částečně shodovat.

Parametry β a γ obou rovnic se pak odhadují pomocí metody maximální věrohodnosti, přičemž věrohodnostní funkce je definována jako

$$L = \prod_{i=1}^n \Pr(W_i^* \leq 0)^{1-W_i} \left[\Pr(Y_i^* \leq 0 | W_i = 1)^{1-Y_i} \Pr(Y_i^* > 0 | W_i = 1)^{Y_i} \Pr(W_i^* > 0) \right]^{W_i},$$

kde se příslušné pravděpodobnosti vyjádří pomocí předpokládaného sdruženého normálního rozdělení.

Zásadním problémem tohoto modelu je však opět to, že v našem případě pro jeho odhad nemáme dostatečná data. Na základě našich databází totiž můžeme opět pozorovat pouze rozhodnutí jen těch studentů, kteří na vysoké škole studují, a proto nám chybí údaje o těch, kteří se rozhodli nepřihlásit, anebo nebyli přijati. Není proto možné do logistické funkce zahrnout výrazy $\Pr(W_i^* \leq 0)^{1-W_i}$, které právě těmto studentům odpovídají. Autoři článku tento problém nemají, neboť sbírají svá data dotazníkovou metodou, a to i pro studenty, kteří ukončili střední školu a na vysokou nenastoupili.

My data týkající se těchto studentů nemáme, a proto nemůžeme tento přístup použít. Návrh řešení pro náš případ lze nalézt v knize Limited-Dependent and qualitative variables in econometrics (Maddala, 1983), kde se používá tentýž systém dvou rovnic jako v článku Cingana a Cipolloneho a normální sdružené rozdělení náhodných chyb v obou rovnicích. Věrohodnostní funkce je definována jako

$$L = \prod_{i=1}^n \left[\frac{\Pr(W_i^* > 0, Y_i^* > 0)}{\Pr(W_i^* > 0)} \right]^{Y_i} \left[1 - \frac{\Pr(W_i^* > 0, Y_i^* > 0)}{\Pr(W_i^* > 0)} \right]^{1-Y_i},$$

kde sdružené pravděpodobnosti $\Pr(W_i^* > 0, Y_i^* > 0)$ se napočítávají z dvourozměrného sdruženého normálního rozdělení a jednoduchá pravděpodobnost $\Pr(W_i^* > 0)$ z normálního rozdělení jednorozměrného.

Sekvenční probitový model

Na závěr bychom také chtěli upřesnit strukturu modelu uvedeného v bodu 1. c). Ukazuje další možný přístup k problematice modelování, založený na rozdělení postupu studiem

na dvě fáze (zde na bakalářské a magisterské studium), z nichž obě mohou skončit úspěchem či neúspěchem, a z nichž druhá fáze vůbec nemusí být realizovaná. Jedná se tedy vlastně o sekvenční probitový model s výběrovým zkreslením v druhé fázi. Podobný přístup je popsán v článku *The Determinants of University Dropouts : A Sequential Decision Model with Selectivity Bias* (Montmarquette, Mahseredjian, Houle, 1993). Tento model neuvažuje výběrové zkreslení pro první fázi. Odhad je proto založen také na dvourozměrné distribuční funkci a model je tak strukturou velmi podobný modelu Cingana a Cipolloneho, ovšem pro jeho odhad bychom i v našem případě mohli mít dostatečná data, neboť máme údaje o studentech, kteří dokončí bakalářské studium a nenastoupí do studia magisterského.

Analýza pomocí hazardní míry

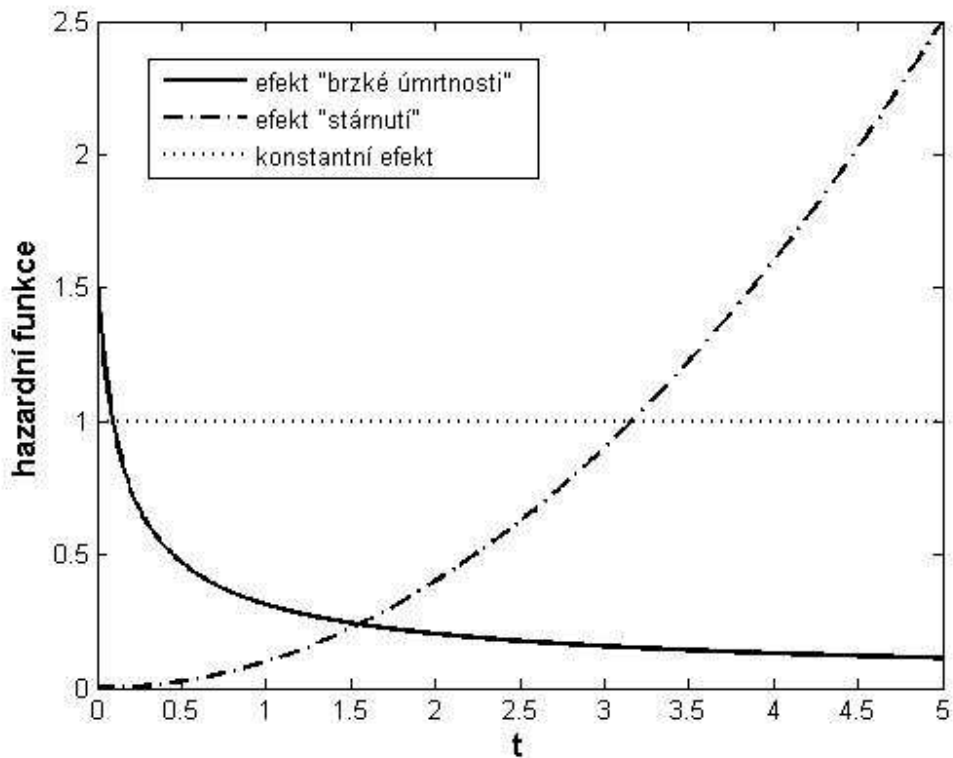
Probitová regresní analýza není jedinou metodou, kterou nám matematická statistika nabízí jako nástroj pro zkoumání problému pravděpodobnosti dokončení studia na vysoké škole. V následující části si přiblížíme metodu modelů hazardní míry, která nám umožňuje poněkud jiný problém na danou problematiku.

Nebudeme se zde již zabývat pravděpodobností úspěšného dokončení studia, ale naopak pravděpodobností, že bude studium předčasně a neúspěšně ukončeno. Přesněji řečeno, budeme se zajímat o to, jak pravděpodobné je v každém časovém okamžiku, že v následujícím časovém okamžiku bude studium daného studenta neúspěšně ukončeno. Tato pravděpodobnost se nazývá hazardní funkce a je definována jako

$$\lambda(t) = \lim_{\delta \rightarrow 0} \frac{P(t \leq T < t + \delta | T \geq t)}{\delta},$$

kde T označuje délku studia a t je časový okamžik, ve kterém hazardní funkci vyhodnocujeme. Je to tedy pravděpodobnost, že studium trvající až do času t skončí v krátkém časovém intervalu následujícím po čase t , normalizovaná délkou tohoto intervalu. Například měříme-li čas v letech, bude $\lambda(2)$ vyjadřovat, jaká je pravděpodobnost, že studium trvající dva roky bude ihned po těchto dvou letech neúspěšně ukončeno.

Hazardní funkce je koncept převzatý z technické oblasti, zejména ze studií o poruchovosti. Tři typické průběhy hazardní funkce jsou znázorněny na následujícím obrázku pro lepší ilustraci.



Hazardní funkce popisující efekt „brzké úmrtnosti“ je typická například pro elektronické součástky na počátku jejich funkčnosti. Popisuje situaci, kdy velmi pravděpodobné, že součástka přestane fungovat zejména krátce po zapojení do provozu, což může být dané například vadou danou při její výrobě a podobně. Pokud se tento efekt neprojeví ihned po zapojení do provozu, součástka je pravděpodobně v pořádku a nadále bude fungovat bez problémů. Na tomto principu je ostatně založen celý systém záručních lhůt různých výrobků.

Naopak, hazardní funkce popisující efekt „stárnutí“ je typická spíše pro součástky na konci doby jejich životnosti, kdy se součástka stává stále opotřebovanější a pravděpodobnost, že přestane fungovat, se s časem zvyšuje.

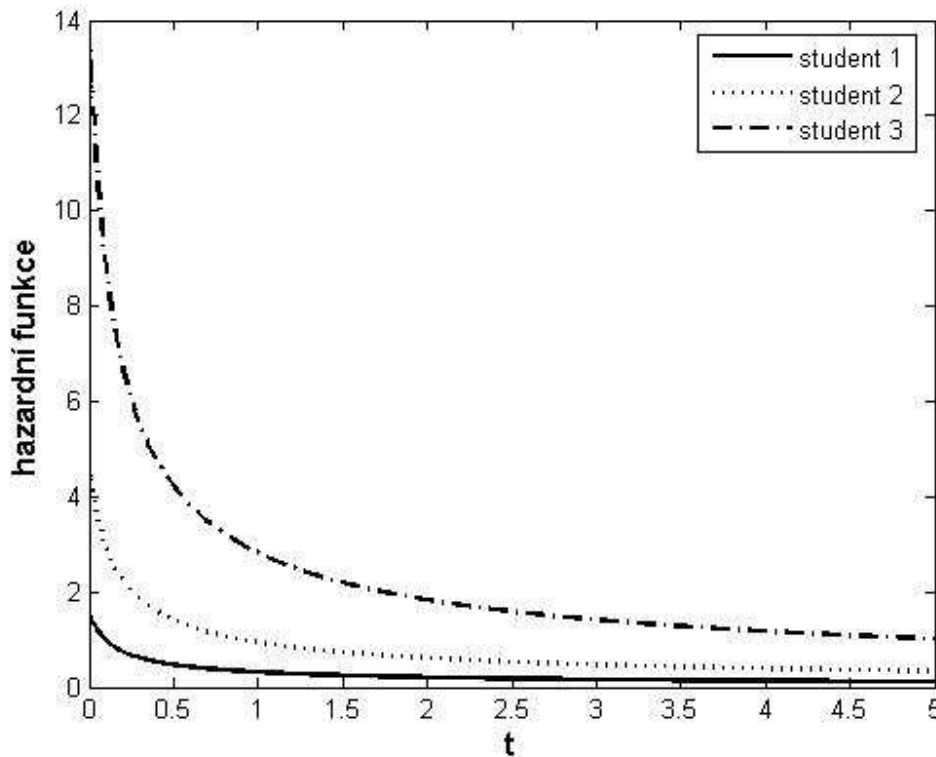
Třetí ze znázorněných možností je konstantí hazardní funkce, kdy pravděpodobnost ukončení životnosti nezávisí vůbec na čase. Pro případ naší analýzy vysokoškolského studia se zdá nejvhodnější hazardní funkce prvního typu, kdy je největší pravděpodobnost neúspěšného předčasného ukončení studia krátce po nástupu na vysokou školu. U vysokoškolských studentů se pochopitelně nejedná o nějakou „poruchovost“, ale o to, že jednak na mnohých školách (zejména technického typu) bývají první ročníky nejtěžší, aby došlo k určitému dodatečnému výběru studentů, a jednak mnozí studenti právě v prvních měsících po nástupu teprve zjišťují, zda je vybraná škola pro ně opravdu ta pravá. Ve vyšších ročnících bývá pravděpodobnost předčasného ukončení nižší, protože studenti jsou již na dané škole lépe adaptováni a navíc již do studia vložili nemalé množství prostředků a úsilí, které nechtějí jen tak „zahodit“. Samozřejmě, toto je pouze naše hypotéza, kterou uvádíme pro lepší přiblížení konceptu hazardní funkce. Ve skutečnosti metoda analýzy pomocí hazardní míry nepotřebuje, abychom předem specifikovali tvar naší hazardní funkce, ale sama ho pomocí semiparametrických metod odhadne na základě dat.

Nyní tedy přistoupíme k vysvětlení toho, jak tato metoda funguje v kontextu naší analýzy vysokoškolského studia. Je zřejmé, že hazardní funkce pro jednotlivé studenty budou vypadat jinak a budou záviset jak na jejich osobních charakteristikách, tak na konkrétním oboru, který studují. Je to proto, že hazardní funkce vlastně odráží pravděpodobnost neúspěchu studenta a ta se pochopitelně pro jednotlivé studenty liší. Proto zavádíme pro každého studenta „individuální“ hazardní funkci

$$\theta_i(t) = \lambda(t) \exp(\beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \dots + \beta_k X_i^k)$$

kde, stejně jako v probitovém modelu, veličiny X_i^j představují charakteristiky studenta a školy a β_j jsou parametry našeho modelu. Individuální hazardní funkce $\theta(t)$ je tedy odvozena ze základní hazardní funkce $\lambda(t)$ transformací, která závisí na individuálních charakteristikách. Na následujícím obrázku je pro lepší představu znázorněno, jak podobná transformace modifikuje základní tvar hazardní funkce pro tři různé studenty. Z obrázku je patrné, že v každém okamžiku je pravděpodobnost neúspěšného ukončení studia nejvyšší pro studenta č. 3 a naopak nejnižší pro studenta č. 1. lze tedy usuzovat, že

charakteristiky prvního studenta jej zdaleka nejlépe předurčují ke studiu na dané vysoké škole.



Pomocí individuální hazardní funkce můžeme vyjádřit pravděpodobnost, že student předčasně neúspěšně ukončí studium a tu potom dosadit do věrohodnostní funkce. Ta je posléze maximalizována pomocí semiparametrických metod, které určí optimální tvar hazardní funkce $\lambda(t)$ a především koeficienty β_i . Tyto koeficienty nás pak informují o vlivu jednotlivých charakteristik na pravděpodobnost předčasného neúspěšného ukončení studia podobným způsobem, jaký ilustruje předchozí obrázek.

Celý proces výpočtu daných pravděpodobností a sestavení věrohodnostní funkce pro odhad je poněkud technicky náročnější, neboť vyžaduje přihlídnutí k diskrétní povaze problému (studium bývá ukončeno zpravidla na konci jednotlivých let či semestrů) a také k tomu, že studium je časově omezeno maximální povolenou dobou. Nezbytné modifikace a další technické detaily lze nalézt ve studii A Hazard Model of the Probability of Medical School Dropout in the United Kingdom (2001).

Naším cílem zde není seznámit čtenáře s veškerými technickými detaily analýzy pomocí hazardní míry, ale spíše poukázat na existenci této metody, která je zajímavou alternativou k probitovému modelu a jako taková se v literatuře běžně uvádí. Oproti probitovému modelu není analýza pomocí hazardní míry závislá na předpokladu rozdělení disturbancí, protože hazardní funkce, která toto rozdělení odráží, není pevně dána, ale je odhadována pomocí semiparametrických metod. Na druhou stranu, tyto metody jsou pochopitelně výpočetně náročnější. Je proto vhodné provést jak analýzu hazardní míry tak odhad probitového modelu a následně porovnat výsledky.

Je také možné poznamenat, že analýza pomocí hazardní míry nám dává poněkud jiný typ informace, než základní probitový model, neboť studuje vlastně dynamický vývoj pravděpodobnosti neúspěchu při studiu. Základní probitový model pouze odpovídá na otázku, jaká je pravděpodobnost, že student vysokou školu úspěšně dokončí. Pochopitelně, parametry obou modelů by měly vypovídat podobným způsobem o tom, jak jednotlivé charakteristiky ovlivňují případný úspěch a neúspěch, je však třeba mít na paměti, že základní myšlenka obou modelů se liší. I proto je vhodné provádět oba typy analýz a následně porovnat získané výsledky.

P.5 Akademické studie

V následující části se budeme věnovat akademickým studiím zaměřujícím se na identifikaci faktorů s vlivem na rozhodování studentů o ukončení studia aneb o změně studijního programu. V souladu se zadanou úlohou projektu SIMS se budeme koncentrovat především na popis dat a následné zdůraznění faktorů, které autoři ve své ekonometrické analýze shledali významnými při vysvětlování jevu předčasného ukončení studia na vysoké škole. Jelikož odhad pravděpodobnosti dokončení studia je jednou ze zkoumaných otázek projektu SIMS, důraz v následující stati bude kladený hlavně na rozbor dostupnosti diskutovaných dat v databázi SIMS. Protože technický popis následujících ekonometrických modelů je mimo rozsah této statě, budeme se mu věnovat blíže v příslušné kapitole věnované modelu odhadu pravděpodobnosti dokončení studia na vysoké škole.

Studie *The determinants of university dropouts: a bivariate probability model with sample selection* (2000) se zabývá identifikací faktorů majících vliv na předčasné ukončení studia na Université de Montréal. Předmětná analýza se skládá ze dvou kroků, a to z pohledu studenta před nástupem na vysokou školu a z pohledu studenta po ukončení prvního semestru. Zkoumané faktory je možné rozdělit do čtyř základních skupin zahrnujících osobní charakteristiky, socioekonomické faktory, enviromentální faktory a lokální faktory.

V rámci osobních charakteristik autoři zkoumali vliv pohlaví, věku, výsledků standardizovaných testů před nástupem na vysokou školu, popřípadě výsledků po ukončení prvního semestru, formy studia (denní, externí) a typu školy (veřejná, soukromá). Ekonometrickou analýzou zjistili, že významnou úlohu vzhledem k míře předčasných ukončení studia má věk studenta. Došli k závěru, že vyšší věk nastupujícího studenta znamená zvýšenou pravděpodobnost, že student školu nedokončí, neboť jeho čas má větší hodnotu, a student je proto méně ochotný experimentovat s daným studijním oborem. Stejný vztah platí i pro externí studenty, kteří jsou vystaveni větší pravděpodobnosti odchodu ze školy než studenti denního studia. Překvapivě nebyl nalezen vztah mezi dosaženým skóre v standardizovaných testech před nástupem na vysokou školu a mírou předčasného odchodu ze školy. Naopak, v případě stejné analýzy u studentů s ukončeným prvním semestrem měly výsledky dosažené v prvním semestru významně pozitivní vliv na pravděpodobnost setrvání na škole. Na závěr, velikost třídy v prvním ročníku se ukázala jako významný ukazatel, mající pravděpodobně vliv na kvalitu výuky a přímý kontakt s přednášejícím. Ve vztahu k dostupnosti zkoumaných faktorů v databázi SIMS, je nutno dodat, že kromě výsledků standardizovaných testů před nástupem na studium, jakož i výsledků z prvního semestru, jsou všechny ostatní diskutované údaje dostupné, ať už prostřednictvím databáze SIMS, nebo prostřednictvím dodatečných zdrojů (průměrný počet studentů v třídě).

Podobnou analýzou se zabývá i studie *Determinants of University drop-out probability in Italy* (2003). Na rozdíl od předcházející studie se autoři kromě určení faktorů majících vliv na pravděpodobnost nedokončení vysokoškolského studia věnují i faktorům

ovlivňujícím pravděpodobnost, že se student do vysokoškolského studia vůbec zapojí. K faktorům, kterým jsme se již věnovali při předcházející studii, tato práce přidává důraz na vliv rodiny na výsledky studenta (délka vzdělávání rodičů) jakož i další socioekonomické vlivy typu nezaměstnanost v regionu anebo vzdálenost bydliště od místa vzdělávání.

Jak bylo možné očekávat, vyšší úroveň vzdělání rodičů má pozitivní dopad na pravděpodobnost dokončení vysokoškolského vzdělání. Nezaměstnanost v regionu má podle autorů vliv na pravděpodobnost participace na vysokoškolském vzdělávání, kde nižší míra nezaměstnanosti znamená vyšší pravděpodobnost účasti studenta na terciárním vzdělávání. Opět, ve vztahu k možnostem databáze SIMS je nutné připomenout, že ne všechny diskutované faktory jsou obsaženy v databázi. Faktory typu nezaměstnanost v regionu, ze kterého student pochází, jakož i vzdálenost místa bydliště od místa studia se sice v databázi projektu SIMS nenachází, avšak vzhledem k minimálně povinnému údaji o okrese, ve kterém má studující trvalé bydliště, je zde možnost, jak databázi SIMS provázat s dodatečnými datovými zdroji. Horší je to už v případě zahrnutí vlivů rodinného prostředí, které nejsou v databázi SIMS zahrnuté, a případné navázání na databáze obsahující informace tohoto typu může být komplikované.