

**Zařazení a název:** Lekce 8: Analýza dat a tvorba předpovědí

**Autor (resp. autoři):** Maria Králová, Daniel Němec, Robert Jahoda

**Datum vygenerování dokumentu:** Pá 3. listopad 2017, 12:45:38 CET

**Upozornění:** Obsahem tohoto dokumentu je část učebního textu e-learningového kurzu Akademické psaní, která byla vybrána a automaticky převedena z interaktivní osnovy z IS MU za účelem usnadnění tisku a možnosti studia v elektronických čtečkách typu Kindle, iPad apod. Přestože je naší snahou, aby se tento dokument po obsahové stránce v maximální možné míře shodoval se zdrojovou osnovou, nelze vyloučit, že při převodu mohlo dojít ke ztrátě či „zašumění“ některých publikovaných informací, nevykreslení některých obrázků či speciálních symbolů, rozhození formátování a podobným neduhům v důsledku automatického převodu. V případě nejasností je nutné správnost zobrazovaných informací ověřit přímo v příslušné interaktivní osnově v IS MU. V podobném duchu je žádoucí nahlédnout do osnovy v IS MU v případě citovaných zdrojů – úplný seznam literatury není součástí tohoto dokumentu. Tento dokument byl vytvořen **výhradně** pro studenty kurzu Akademické psaní a bez souhlasu autora **není povoleno jej šířit** třetím osobám.

# 1 Empirický výzkum III: analýza dat a tvorba předpovědí

Neodpovídají-li fakta vaší teorii,  
je třeba se jich co nejrychleji zbavit.

*Maierův zákon*

Jakmile máme k dispozici potřebná data, je třeba je nějak vyhodnotit, tj. zjistit, co vlastně znamenají. Nejjednodušší analýzou dat je pohled na jejich graf. Vhodně zvolený graf vývoje nějaké veličiny či vztahu několika veličin dává velmi dobrý vhled do toho, co se děje. Obvykle však potřebujeme přesnější informace, než jaké nám může grafická analýza poskytnout. Pokud máme k dispozici větší množství různých datových souborů, je třeba je zpracovat hromadně, protože dívat se na všechny možné kombinace dat očima není možné. Z grafů je obtížné vyčíst takové údaje, jako zda se liší střední hodnoty dvou veličin, zda se liší jejich rozptyly, zda mají jednotlivé veličiny stejné statistické rozdělení, zda mezi veličinami existuje nějaký vztah a pokud ano, tak jaký. Samotný graf nám také neumožní předpovědět budoucí vývoj sledované veličiny. Ke všem těmto účelům můžeme použít více či méně sofistikované statistické techniky.

## 1 Statistika

*Tuto část napsala Maria Králová.*

Statistika je studium sběru, organizace a interpretace dat pomocí matematických a výpočetních nástrojů. Na úrovni **popisné statistiky** jde o zřehlednění velkého množství dat. Činíme tak pomocí tabulek, grafů, histogramů, různých číselných charakteristik datových souborů jako jsou průměr, medián, směrodatná odchylka apod. Můžeme také analyzovat a popisovat vztahy mezi proměnnými, např. jak souvisí výše příjmu s dosaženým vzděláním, zabývat se rozložením datového souboru apod. Závěry získané tímto způsobem nelze zobecňovat, platí pouze pro konkrétní datový soubor, z něhož byly výstupy pořízeny. Přesto i takovéto slabé závěry jsou užitečné. Umožňují rychlý vhled do problému, urychlují a zjednodušují čtení a hlavně inspirují k formulaci hypotéz, konstrukci teorií a vedou ke kladení smysluplných otázek. Vedlejším, ale žádoucím efektem je i estetická stránka prezentace dat.

Mnohem mocnějšími nástroji disponuje **matematická statistika**. Pro ekonomickou analýzu jsou důležité především dvě její části. Jedna se týká vzorkování a zobecňování informací ze vzorku na celou populaci. Pokud např. chceme před volbami získat odhad volebních preferencí, nemůžeme dotazovat všechny voliče. Dotazujeme pouze reprezentativní výběr voličů a na základě výběru usuzujeme, jak by volby dopadly pro celý soubor všech voličů. Postupujeme tedy induktivně – z části (z výběrového souboru) usuzujeme na celek (základní soubor). Matematická statistika dává návod, jak takové usuzování provést korektně: jak vytvořit výběrový soubor, aby byl reprezentativní, jak má být tento soubor velký, jak provést zobecnění na základní soubor a jak odhadnout velikost chyby našeho zobecnění.

Druhá část matematické statistiky, která je velmi důležitá pro empirický výzkum v ekonomii, se týká testování hypotéz. Řekněme, že jsme provedli ekonomický experiment, který zkoumal, kolik lidí přispívají na veřejné statky ve dvou situacích: když nemohou a když mohou trestat černé pasažéry, kteří z veřejných rozpočtů čerpají, aniž by na ně sami přispěli. Nyní máme k dispozici data o přispěvcích každého subjektu experimentu při každém nastavení a potřebujeme posoudit, zda se liší. Protože jsou oba datové soubory zatíženy náhodnou chybou, jejich průměrné hodnoty se liší. Otázkou je, zda se liší dost na to, abychom si mohli být dostatečně jistí, že se liší skutečně, ne jen v důsledku náhody. Matematická statistika poskytuje nástroje ke zkoumání takových hypotéz o rozdílnosti středních hodnot, rozptylu, statistického rozdělení apod. Podobně umožňuje také testovat existenci vztahu mezi veličinami. Můžeme např. testovat, jak spolu (pokud vůbec) souvisí náklady na rekvalifikaci, počet hodin rekvalifikace a další proměnné se mzdou rekvalifikovaného zaměstnance.

Základy teorie pravděpodobnosti, popisné i matematické statistiky se na ESF vyučují v povinném dvousemestrálním předmětu Statistika 1 a Statistika 2 a plánuje se zavedení nepovinného Statistika 3. Pokud se chcete o statistice dozvědět víc, můžete začít např. těmito učebnicemi:

- BUDÍKOVÁ, M. - KRÁLOVÁ, M. - MAREŠ, B. *Průvodce základními statistickými*

metodami. Praha: Grada, 2010. ISBN 9788024732435.

- HENDL, J. *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. Praha: Portál, 2006. ISBN8071788201.
- HANOUSEK, J. - CHARAMZA, P. *Moderní metody zpracování dat: matematická statistika pro každého*. Praha: Grada, 1992. ISBN8085623315.
- FREEDMAN, D. - PISSANI, R. - PURVES, R. *Statistics*. 4th ed. San Francisco: Norton and Company, 2007. ISBN 9780393929720.
- BABBIE, E. - HALLEY, F. - ZAINO, J. *Adventures in Social Research with SPSS Student Version: Data Analysis Using SPSS 14.0 and 15.0 for Windows*. Thousand Oaks: Pine Forge Press, 2007.

Vlastní statistické zpracování dat obvykle vyžaduje použití nějakého softwaru. Použít můžete buď obecné výpočetní systémy, jako je Matlab nebo volně šiřitelný jazyk *R*, nebo specializované komerční softwary, např. *Statistica*, *SPSS* či *SAS System*. Ke *Statistice* a *SPSS* vlastní MU multilicenci. Přehled mnoha dalších dostupných programových prostředí pro statistickou analýzu a vizualizaci dat najdete na webové stránce Free Statistical Software, <http://en.freestatistics.info/en/stat.php>.

## 2 Ekonometrie

*Tuto část napsal Daniel Němec.*

Ekonometrie je speciální disciplína, která spojuje statistické techniky s matematickou ekonomikou. Zjednodušeně řečeno, ekonometrie umožňuje odhadnout parametry rovnic, které popisují vztahy mezi jednotlivými ekonomickými proměnnými. Kromě toho, že je taková kvantifikace užitečná sama o sobě, umožňuje nám testovat ekonomické hypotézy a tvořit předpovědi o budoucím vývoji sledovaných veličin. Své využití ekonometrie najde v mikroekonomii, makroekonomii, financích, marketingu, veřejné ekonomii a vlastně ve všech odvětvích ekonomie, kde se provádí empirický výzkum.

Řekněme, že nás zajímá, zda existuje vztah mezi tempem růstu peněžní zásoby a mírou inflace v dané zemi. Pokud bychom si vybrali zemi s vysokou mírou inflace, viděli bychom vztah mezi tempem růstu peněžní zásoby a inflací snadno v grafu. V nízkoinflačních ekonomikách však tento vliv nemusí být z grafu patrný, protože míru inflace ovlivňují i další veličiny, jako je tempo růstu HDP, inovace na finančním trhu apod. Pokud je tempo růstu peněžní zásoby malé, mohou tyto ostatní vlivy způsobit, že vliv tempa růstu peněžní zásoby nebude v grafech patrný. Abychom jej mohli zjistit, musíme odstranit ostatní vlivy (kontrolovat je). Bylo by to snadné, pokud bychom mohli provést experiment; to však v našem případě není možné. Musíme tedy data kontrolovat pomocí nástrojů ekonometrie.

Postup je relativně přímočarý. Z ekonomické teorie odvodíme rovnici, která bude mít na jedné straně míru inflace jako vysvětlovanou veličinu a na druhé straně výraz, který bude záviset na předpokládaných vysvětlujících veličinách, jako je tempo

růstu peněžní zásoby, tempo růstu HDP, jejich zpožděné hodnoty apod. a na neznámých parametrech. Základní rovnice může mít např. tvar, ve kterém se míra inflace může rovnat A krát tempu peněžního růstu plus B krát tempu růstu HDP. Ekonometrické techniky odhadu nám umožní odhadnout neznámé parametry A a B, ale také posoudit míru nejistoty spojenou s odhadem těchto parametrů a celého modelu a zjistit, zda je vztah mezi inflací a tempem růstu peněz skutečný, nebo jen náhodný. Můžeme také spočítat mezní míru vlivu jednotlivých veličin, např. o kolik procent vzroste míra inflace, když tempo peněžního růstu vzroste *ceteris paribus* o jeden procentní bod. Ekonometrické techniky nám také umožní otestovat, zda je zvolený funkční tvar rovnice správný, nebo je třeba jej modifikovat. Když známe správný funkční tvar a odhady parametrů, můžete předpovídat i budoucí míru inflace. Pokud např. víme, že příští rok HDP poroste tempem 3 %, můžeme spočítat, jaká bude míra inflace, když peněžní zásoba vzroste o 4 %. Kromě tzv. bodové předpovědi získáme i intervaly spolehlivosti, tedy pásmo, ve kterém by skutečná inflace měla ležet se zvolenou pravděpodobností. Jednoduše řečeno, ekonometrie dává ekonomické teorii kvantitativní rozměr.

Na ESF se můžete základy ekonometrie naučit v předmětech Základy ekonometrie a Ekonometrie. Můžete se také naučit základy alternativního přístupu k ekonometrii v předmětu Bayesiánská analýza. Pokud se chcete o ekonometrii dozvědět víc, můžete začít např. těmito učebnicemi:

- KOOP, G. *Introduction to Econometrics*. New York: Wiley, 2008. ISBN 9780470032701.
- VERBEEK, M. *A Guide to Modern Econometrics*. 3rd ed. New York: Wiley, 2008. ISBN 9780470517697.
- KENNEDY, P.: *A Guide to Econometrics*. 5nd ed. Oxford: Blackwell Publ., 2003. ISBN 0262112809.
- HEIJ, Ch. - DE BOER, P. - FRANCES, P. H. - KLOEK, T. - VAN DIJK, H. K. *Econometric Methods with Applications in Business and Economics*. Oxford: Oxford University Press, 2004. ISBN 0199268010.

Pro odhad a testování ekonometrických modelů se kromě obecného matematického a statistického software, jako je Matlab, Octave, SPSS, R apod., hodí skvělý volně šiřitelný program Gretl.

### **3 Modelování a mikrosimulační modely v ekonomii**

*Tuto část napsal Robert Jahoda.*

Tvůrci zákonů a politici často stojí před otázkou, jestli jimi navržená opatření budou v realitě skutečně fungovat tak, jak bylo zamýšleno nebo jaký že vlastně bude dopad opatření, které navrhli. Nejsnáze a nejpřesněji se na tyto otázky odpovídá způsobem, že dané „politiky“ zavedeme do praxe, necháme je působit a po nějaké době jejich dopad vyhodnotíme. Pro vyhodnocování po skončení sledovaného období se nejenom v ekonomii ujal název *ex post*. V případě hodnocení více variant

nějakého opatření však metoda ex post není moc vhodná, používá se spíše při hodnocení experimentů v přírodních oborech. Jestliže máte dvě skupiny myší a jedné skupině podáte chemickou substanci a druhé – kontrolní - skupině podáte „placebo“, můžete po určité době zkoumat, jaký je výsledek tohoto experimentu. Tento pokus můžete několikrát opakovat, kdy můžete obměňovat složení nebo množství chemické substance. Pokud odhlédneme od případného černého svědomí (co ty laboratorní myšky?), je tato metoda docela levná, znovu replikovatelná a výsledky většinou vidíte poměrně rychle. Zkuste si ale představit, že obdobným způsobem bude vláda navrhovat a vyhodnocovat dopad u zákonů o sociální politice, kdy nastaví životní minimum občanů na určité (např. příliš nízké) úrovni a po dvou letech se do hodnotící zprávy napíše, že polovina příjemců dávky umřela hlady a druhá polovina je ve vězení za krádeže a loupeže a že by proto bylo vhodné životní minimum nastavit na jinou úroveň (mimořádně, kam se nám ztratila kontrolní skupina?). Asi si lze jen těžko představit, že by vláda touto metodou pokus-omyl řídila svou politiku (i když někteří si mohou myslet, že přesně tímto způsobem dnešní vlády své politiky provádějí). Zkrátka, provádět veřejné politiky a spoléhat se pouze na ex-post evaluace se jeví jako nedostatečné, již při návrhu určité politiky bychom měli znát, jestli přijímané opatření povede ke kýženému cíli. A k tomu nám mohou pomoci ex ante analýzy dopadů založené na simulačních metodách.

Představme si, že se vláda rozhodne zvýšit daň z přidané hodnoty (dále jen „DPH“) za účelem posílení daňových příjmů veřejných rozpočtů. Ze dvou sazeb ve výši 10 % a 20 % má dojít ke sjednocení do jedné sazby ve výši 17,5 %. Povede tato změna opravdu k posílení daňových příjmů veřejných rozpočtů? Určitě ale existují i jiné otázky, které bychom si v souvislosti se zamýšlenou reformou měli pokládat. Politiky nejspíš budou zajímat následující otázky: Bude mít reforma vliv na inflaci a případně, u kterých spotřebních skupin dojde k nejvyššímu cenovému nárůstu? Dojde k přerozdělení čistého důchodu (příjmu) ve společnosti? Jaká je incidence (dopad, rozprostření) daňového břemene ve společnosti, které sociální skupiny ponесou dopady reformy? Soukromý sektor si zase může klást na následující otázky: Jak se změní reálná spotřeba domácností, jak se změní jejich spotřební zvyklosti, které sektory ekonomiky budou postiženy více a které méně? Určitě jsme nevyčerpali všechny možné otázky, ale i tak je zřejmé, že před samotným spuštěním reformy by bylo vhodné provést ex ante analýzu, která by nám pomohla najít odpověď na některé z výše uvedených otázek.

Odpovědi na předcházející otázky nám mohou dát metody založené na modelování a simulaci. Zpravidla při nich dochází k zjednodušení celé zkoumané oblasti a ke snaze postihnout pouze její nejzávažnější skutečnosti. Je tomu tak z důvodu, že zkoumaná problematika může být natolik komplexní, že zahrnout do modelu všechny její aspekty nebývá možné. Druhým důvodem pro zjednodušení zkoumané problematiky je hledisko času, kdy snaha o komplexnost simulačního modelu bude vykoupena časovou náročností jeho přípravy a vyhodnocení. Jestliže se výzkumník v průběhu modelování rozhodne pro modifikaci simulované reformy, mů-

že příliš podrobný model vyústit až v nutnost tvorby úplně nového modelu, který bude schopen simulované alternativy reformy zhodnotit.

V případě výše uvedeného příkladu můžeme simulaci změny sazeb DPH provést na zjednodušeném modelu domácnosti, u které máme dán objem a strukturu její spotřeby (víme, kolik spotřebovává potravin, vody, pohonných hmot, atd.). Abychom mohli provést simulace a posléze vyhodnotit možné dopady změny sazeb DPH, je vhodné si definovat některé předpoklady, které nám analýzu zjednoduší:

1) Dojde k úplnému přesunu (změny) daně do cen výrobků a služeb. Ve skutečnosti nemusí dojít k 100% přesunu daně do cen, protože část daně mohou nést obchodníci. U různého zboží dojde k odlišnému přesunu daně do cen, pokud snížíme základní sazbu daně z 20 % na 17,5 %, některé komodity mohou zlevnit, ale jiné nemusí. Předpoklad úplného přesunu daní do cen nám umožní lépe simulovat změnu spotřeby jednotlivých komodit a umožní nám vyhodnotit, mezi které skupiny zboží a služeb je rozprostřena změna daňové povinnosti. Pokud chceme konstruovat komplexnější model, nemusíme omezení přijímat v tak tvrdé formě, případně ho můžeme přijmout ve více variantách.

2) Objem nebo struktura spotřeby domácnosti se po změně sazeb DPH nemění. V nejjednodušší podobě budeme předpokládat, že domácnost spotřebovává stále stejný koš zboží a služeb, mírně slabší verze pak předpokládá, že se nemění hodnota spotřebního koše a jeho struktura. Ve skutečnosti je tento předpoklad samozřejmě nerealistický, domácnost na změnu cen zboží bude reagovat změnou struktury svého spotřebního koše (u normálních statků bude substituovat dražší zboží levnějším). Pokud nejsme schopni s jistotou říct, jak se změna sazeb daně promítne do cen zboží, nemá moc smyslu se snažit kvantifikovat, jaká bude změna ve spotřebním koši domácnosti. Nejsnadnější je prostě předpokládat, že ke změně spotřebního chování nedojde.

3) S předcházejícím předpokladem souvisí ještě jedna věc. Kromě cen zboží a služeb má na strukturu spotřebního koše svůj vliv i výše disponibilního důchodu domácnosti. Pokud se ale mění výše sazeb DPH, např. za účelem zvýšení veřejných příjmů, mění se u sledované domácnosti výše jejího příjmu? Při naší simulaci budeme předpokládat, že ke změně důchodu u domácnosti nedošlo, v realitě ale vyšší vládní příjmy mohou vyústit v nižší ceny veřejně poskytovaných služeb a statků, což má stejný efekt, jako kdyby domácnost měla vyšší příjem (domácnost je zcela nebo částečně kompenzována za vyšší zdanění). Spočítat nebo odhadnout míru kompenzace domácnosti je však velmi obtížné a proto budeme předpokládat, že k žádné kompenzaci nedošlo a důchod domácnosti se nezměnil.

Pokud přijmeme výše uvedené předpoklady, je samotné provedení simulace zvýšení sazeb DPH na naši modelovou domácnost již poměrně snadné a zvládnete to v jakémkoliv tabulkovém procesoru. Snadné bude i vyhodnocení dopadů reformy sazeb daně, musíte však mít neustále na paměti, že:

1) Simulaci jste provedli za podmínky platnosti výše uvedených předpokladů. Pokud by reforma byla skutečně provedena, téměř jistě by se ukázalo, že tyto předpoklady v realitě zcela neplatí. Co s tím můžete udělat? Můžete si vytvořit více variant předpokladů, na základě kterých poté provedete set simulací. Tím dostanete variabilní výsledky reformy, přičemž pokud jste jednotlivé varianty reformy stanovili reálně a s citem, měl by skutečný dopad reformy nejspíš ležet mezi maximální a minimální hodnotou simulovaných variant.

2) Vyhodnocení dopadů reforem bude poměrně chudé, máte výsledky pouze za jednu modelovou domácnost. Co můžete dělat s tímto problémem? Nejjednodušší bude vytvořit si více modelových domácností (jednotlivec, důchodce, úplná rodina, samoživitelka, student, rentiér, . . .) přičemž počet modelových domácností je jenom na vás. Může se vám ale stát, že až případný čtenář bude číst výsledky simulace u šesté modelové domácnosti z dvaceti, ztratí pozornost a ve výsledcích se ztratí.

Jednoznačně můžeme říct, že téměř vždy je účelné, aby se modelové domácnosti co nejvíce blížily skutečné skladbě domácností, která se nachází v dané společnosti. V našem příkladu reformy sazeb DPH byste proto mohli vyjít ze Statistiky rodinných účtů, kdy Český statistický úřad uvádí strukturu spotřeby pro různé skupiny domácností. V tomto případě statistický úřad v pravidelných intervalech vyšetřuje cca 3000 domácností (šetří se právě struktura jejich spotřebních vydání), toto šetření slouží jako podklad pro měření změn cenové hladiny. Jednou ze skupin, pro kterou ČSÚ uvádí strukturu spotřeby, jsou domácnosti v decilových skupinách. ČSÚ zde na základě příjmové charakteristiky domácností konstruuje deset modelových domácností, kde každá reprezentuje průměrnou domácnost své decilové skupiny, přičemž platí, že v každé decilové skupině je stejný počet domácností a že domácnost ve vyšší decilové skupině „je bohatší“ než domácnost ve skupině nižší. Těchto deset modelových domácností pak odráží stav a strukturu spotřeby české společnosti.

Pokud provedete simulaci pro modelové domácnosti podle decilových skupin, získáváte více informací o dopadu zamýšlené reformy na českou společnost. Z uvedených deseti domácností budete pravděpodobně schopni odhadnout, jak se změní výnos DPH v případě provedení reformy a na které sociální skupiny změna daní dopadne. Protože máte strukturu spotřeby společnosti (průměrná domácnost), jste schopni spočítat, jaký bude mít reforma vliv na cenovou hladinu (změna indexu CPI) a dokonce jste tuto informaci schopni zjistit pro různé příjmové skupiny domácností. Jako ekonom tak můžete dát politikovi do ruky informace, na základě kterých se rozhodne o podobě reformy. Neměli byste ale očekávat, že se vašimi radami bude řídit.

V tomto okamžiku si možná kladete otázku, proč své výpočty dopadu reformy omezovat na deset modelových domácností (dle příjmového decilu), když ČSÚ má podrobné informace o spotřebě všech (cca 3000) domácností. Proč neprovést simulaci reformy pro všechny domácnosti a pak způsob prezentace dopadů měnit podle

toho, jaké zadání od politika obdržím? Ano, tato metoda se taktéž dá použít a nazývá se mikrosimulační modelování.

Ukázali jsme si metody modelování a mikrosimulačního modelování v ekonomii na jednom konkrétním příkladě z oblasti veřejného sektoru. Uvedené metody byste přitom mohli použít při analýze dopadů většiny politik z veřejného sektoru. A nemusíme se držet pouze simulování reforem ve veřejném sektoru, i jiné oblasti ekonomiky využívají popsané metody. Typickým příkladem jsou dnes populární zátěžové testy bank, kdy se posuzuje míra expozice portfolií komerčních bank vůči různým rizikům. Můžeme si klást například následující otázky. Jaký by byl například dopad, kdyby se v letech 2012-2013 vrátila recese a třetina poskytnutých úvěrů by nebyla splácena podle dohodnutých podmínek? Jaký by byl dopad, pokud řecký stát nebude schopen dostát svým závazkům a banky budou muset odepsat 90 % z nominální hodnoty řeckých vládních dluhopisů? Velké použití simulačních metod najdeme v oblasti pojištnictví. Žádná pojišťovna není schopna se sama vypořádat se situací, kdy by jinak náhodně se vyskytující události nastaly ve větší míře v jeden okamžik. Pojišťovna je pro tento případ (např. velké povodně) „zajištěna“ u některé ze světových zajišťoven.

V předcházejícím textu jste se dozvěděli pár informací o modelování a mikrosimulačních metodách. Zkuste nyní sami vymyslet příklad, kdy byste chtěli znát odpověď na otázku, jaký bude dopad nějaké reformy, před tím, než tato reforma bude spuštěna. Zamyslete se, na které otázky byste rádi znali odpověď, jaké předpoklady byste přijímali a na kterých modelových domácnostech byste reformu hodnotili. Nechystá se například změna daně z příjmů? Nezvýšíme daně z cigaret? Nechtějí politici navrhnout daň „z cukru a tuku“?