

JAN HENDL A KOLEKTIV

**ZÁKLADY MATEMATIKY,
LOGIKY A STATISTIKY
PRO SOCIOLOGII
A OSTATNÍ SPOLEČENSKÉ
VĚDY V PŘÍKLADECH**

KAROLINUM



Základy matematiky, logiky a statistiky pro sociologii a ostatní společenské vědy v příkladech

Jan Hendl a kolektiv

Recenzovali:

RNDr. Jindra Reissigová, PhD.

RNDr. Marek Malý, CSc.

Autoři:

Jan Hendl

Martin Moldan

Tereza Ranošová

Jakub Siegl

Martin Štrobl

Jan Žáček

Vydala Univerzita Karlova

Nakladatelství Karolinum

Praha 2021

Sazba Jan Hendl

Druhé, rozšířené vydání

© Univerzita Karlova, 2021

© Jan Hendl a kolektiv, 2021

ISBN 978-80-246-4869-9

ISBN 978-80-246-4326-7 (online : pdf)



Univerzita Karlova
Nakladatelství Karolinum

www.karolinum.cz
ebooks@karolinum.cz

Obsah

I Matematika a logika

1	Úvod do matematiky	13
2	Lineární rovnice	32
3	Vektory	45
4	Maticy	56
5	Logika	86
6	Množiny, relace	119
7	Booleova algebra	140
8	Millova pravidla, QCA	154
9	Analýza sociálních sítí	169
10	Kombinatorika	195
11	Pravděpodobnost	203
12	Markovovy řetězce	224
13	Racionální volba	239
14	Hry se ztrátou nebo ziskem	245
15	Derivace, integrál, modelování	256

II Statistika

1	Úvod do statistiky	287
2	Experimentální vs. observační design	295
3	Popisná statistika	304
4	Náhodné veličiny	321
5	Diskrétní pravděpodobnostní rozdělení	338
6	Normální rozdělení	345
7	Zákon velkých čísel, centrální limitní věta	355
8	Spojité rozdělení a rozdělení statistik	364
9	Teorie statistického usuzování	372
10	Statistická významnost, velikost účinku	384
11	Jednovýběrový z -test pro průměr	394

12	Jednovýběrový t -test pro průměr	399
13	Dvouvýběrový t -test (2 nezávislé výběry)	406
14	Párový t -test (2 závislé výběry)	416
15	Test dobré shody a kategoriální proměnná	422
16	Závislost mezi kategoriálními proměnnými	430
17	Jednovýběrový z -test pro relativní četnost	438
18	Dvouvýběrový z -test pro relativní četnosti	445
19	Neparametrické metody	452
20	Korelační koeficient	466
21	Lineární regresní model	475
22	Doporučení k využívání statistiky	492

III Přílohy

1	Model argumentace podle Toulmina	505
2	Základní vzorce statistiky	509
3	Matematické vzorce, grafy funkcí a tabulky	512
4	Řešení příkladů	525
5	Použitá a doporučená literatura	554

Předmluva

Porozumění základům matematiky a statistiky tvoří často předpoklad pro to, aby mohli studující navázat ve svých oborech na práci velkých předchůdců, aby se udrželi „na ramenou obrů“ a někdy i viděli dál než oni.

Znalostní požadavky v prvních semestrech studia sociálních věd, pokud jde o matematiku a statistiku, se liší od látky, jejíž zvládnutí se předpokládá při studiu přírodních a technických věd. Většinou jde o matematické pojmy a techniky, které lze využít při osvojování probíraných statistických metod (Beneš, Drulák 2020; Gill 2006; Moore, Siegel 2014; Novotná, Špaček, Šťovíčková Jantulová 2019).

Studující se matematice spíše vyhýbají. Je to způsobeno mimo jiné okolností, že střední úroveň vzdělávání je dostatečně nepřipravila, v horším případě se v nich utvrdil pocit jisté úzkosti z matematiky (Jones, Goldring 2014). Negativní pohled studujících na matematiku a vlastní matematické schopnosti dokládají empirické studie (např. Murton, Lehtinen 2003). Zdůvodnění pro tento stav lze přijmout, upřednostňujeme však jiný postoj. Perspektivu má poučený a tvořivý přístup k matematice, logice i statistice.

Matematika je potřeba při zvládnání kvantitativních výzkumných metod. Především objasňování vícerozměrných statistických metod se neobejde bez znalostí základů vektorového a maticového počtu. Jak v kvantitativním, tak v kvantitativním výzkumu využíváme logické principy. S jejich pomocí jsou dokonce vyvinuty metody kvalitativní komparativní analýzy (QCA), která dnes přitahuje velkou pozornost. Nejjednodušší modely časových řad jsou dobrým příkladem mnoha aplikačních možností matematického modelování a využití principů vektorového a maticového počtu. Základy grafů a sociálních sítí je nutné pochopit, protože se pomocí nich řeší problémy v souvislosti s otázkami, které vyvstávají například při využívání internetu (Bonacich, Lu 2014).

Z těchto skutečností autorská skupina vycházela, když se rozhodovala, jaké partie z matematiky do textu zařadit a jak je podat. Také zohlednila, že se jedná o doplněk k přednáškám o základech matematiky a statistiky pro studující bakalářského studia sociologie a ostatních sociálních věd. Přednášející, případně jeho/její/jejich pomocníci a pomocnice, rozhodují, o čem se bude mluvit a co se bude procvičovat.

První část knihy je věnována matematickým základům. Nejprve jsou jednoduše představeny vybrané matematické koncepty, které jsou poté rozšířeny v dalších kapitolách o speciálnější témata, které matematicky zaměřené sociální vědy potřebují. Text se zabývá např. QCA metodou, aplikací jednoduchých náhodných procesů, modelů epidemií, analýzou sociálních sítí, teorií rozhodování a teorií her.

Druhá část knihy se zaměřuje na představení užitečných konceptů statistiky nezbytných pro elementární práci s daty a jejich použití pro vyvozování zobecňujících závěrů. Je probrána popisná statistika, náhodná proměnná, pravděpodobnostní rozdělení a pojmy pojící se s těmito tématy. Větší prostor je věnován porozumění a demonstraci testování statistických hypotéz. Text seznamuje nejprve s jednoduchými metodami (mezi které patří např. z -testy nebo t -testy) a poté s metodami složitějšími, jako jsou např. test dobré shody, analýza kontingenčních tabulek nebo neparametrické metody. Také se osvětlují principy a konstrukce intervalů spolehlivosti. Druhá část knihy je uzavřena úvodem do regresní analýzy.

Každá kapitola obsahuje základní a zjednodušeně pojatou teorii a několik řešených modelových příkladů typických pro dané téma. Modelové příklady jsou základním prvkem této publikace. Na konci každé kapitoly jsou navíc uvedeny další podklady k procvičení společně s výsledky. Text uzavíráme statistickými tabulkami a užitečnými vzorečky. Diagnostický test předpokládaných vstupních znalostí naleznou čtenáři v příkladech k procvičení v první kapitole.

Při aplikaci matematických a statistických metod hrají důležitou roli počítače a příslušný software. V tomto směru doporučujeme tabulkové programy *Excel* od Microsoftu, *Calc* z OpenOffice, komerční statistický systém *SPSS*, volné statistické systémy *PSPP*, *JASP* (oba tyto programy se snaží nahradit systém *SPSS*, *JASP* zpřístupňuje navíc strukturální modelování SEM a síťovou analýzu) a internetový statistický program *SISA* a volný programovací jazyk *R* nebo volný matematický systém *Octave*. Také je cenná knížka Mareše, Rabušice a Soukupa *Analýza sociálně vědních dat [nejen] v SPSS* (vyd. Masarykova univerzita).

Tato kniha by nevznikla bez pomoci mnoha lidí. Dík patří všem, kteří přispěli k její přípravě: studentkám a studentům z FSV UK, zvláště T. Balouškové (podklady pro kapitoly 2, 3, 10 a 11 z matematiky), L. Rozumkové (podklady pro kapitoly 2, 3, 10 a 11 z matematiky) nebo K. Hanzlíkovi (podklady pro kapitoly 10, 11 a 12 ze statistiky), statistikovi a sociologovi P. Soukupovi z FSV UK, logikům M. Pelišovi a P. Arazimovi a oběma recenzentům J. Reissigové a M. Malému z AV ČR.

Koordinátor autorského kolektivu upřímně děkuje všem spoluautorům, zejména mgr. J. Žáčkovi, za jejich příspěvek k této knize. Také je vděčný za přátelskou asistenci PhDr. J. Koukalovi, Ing. K. Novotnému, doc. PhDr. Richardu Papíkovi a prof. MUDr. Š. Svačinovi.

Za chyby v textu přejímá zodpovědnost autorský kolektiv. Za případná pochybení se omlouváme. Budeme každému čtenáři vděční za jakékoliv připomínky a upozornění na nalezené nedostatky a možná vylepšení.

Značení:

\mathbb{N} ... množina přirozených čísel

\mathbb{Z} ... množina celých čísel

\mathbb{Q} ... množina racionálních čísel

\mathbb{R} ... množina reálných čísel

Část I

Matematika a logika

Kapitola 1

Úvod do matematiky

Tradičně byly teoretické vztahy v sociálních vědách formulovány především pomocí kvalitativních pojmů (nemáme na mysli ekonomii). Matematické metody se nepovažovaly za dostatečně vhodné pro rozvíjení teoretického porozumění. Matematika hraje velkou roli v přírodních vědách a biologii, od kterých se sociální vědy liší v mnoha ohledech. Z příslušných úvah proto plyne, že nemůže jít o vulgární přenos přírodovědeckého myšlení přímo do sociálních věd (Lazarsfeld¹ 1954). V posledních desetiletích se matematika a statistika staly důležitými prostředky modelování i v sociálních vědách. V literatuře se objevují speciální matematické metody a postupy. Těm by měli studující porozumět, aby mohli kriticky posoudit popisované skutečnosti, případně zvolený přístup přetvořit.

Příkladem rozšířené aplikace matematiky v sociálních vědách je síťová analýza. Ta se stala domovem výzkumníků a výzkumníků rozličných zájmů, které sahají od konstrukce osobní identity k epidemiologii, od etnografie po analýzu struktur v internetu.

¹Lazarsfeld, Paul, Felix (1901-1976), rakousko-americký sociolog. Věnoval se zejména statistickým metodám a výzkumu veřejného mínění.

Demokracie a matematika

„Pokud jsem viděl dále než ostatní, pak to bylo proto, že jsem stál na ramenou obrů“, řekl fyzik Isaac Newton (1643-1727) v souvislosti s hodnocením významu matematické analýzy pro jeho objevy. Podobně se vyjadřoval Albert Einstein (1879-1955), největší fyzik 20. století. Mezi Newtonem a Einsteinem byla časová mezera stovky let, ale pohled na matematiku měli podobný.

Uvedená myšlenka byla vyslovována mnohokrát autory reformy výuky matematiky na amerických základních a středních školách v 90. letech minulého století. Eseje v knížce „Na ramenou obrů“ (Steen 1990) poukazyvaly k novým pohledům na didaktiku matematiky pro mládež přicházejícího století. „Na ramenou obrů“ pro pedagogy symbolizovalo snahu pomáhat studentům při hledání konceptů matematiky tak, aby docházelo k lepšímu pochopení světa a řešení jeho problémů. Tamejším pedagogům záleželo také na propojení matematiky a demokracie. Matematika hraje určitě významnou roli nejenom ve vědě, ale také při rozvíjení konceptu svobodného občana v demokracii.

Například politici a jejich příznivci se při volbách ohánějí čísly bez omezení a zodpovědnosti. Novináři informují o rozporných číslech s nedostatkem reflexe a odborného vhledu o jejich přesnosti a konzistenci. Pro fungování demokracie je tedy žádoucí jistá míra znalosti „matematiky voleb“ a statistiky, matematická gramotnost občana. Matematická gramotnost však není v tomto pojetí pouze soubor určitých matematických dovedností, nýbrž také zahrnuje uvažování o povaze a kvalitě interakce mezi jedincem s těmito dovednostmi a sociální situací nebo prostředím. To implikuje, že taková gramotnost nemá jednou provždy stejný význam, protože sociální situace a prostředí se neustále proměňují v čase.

Pro pedagogy jsou v této souvislosti aktuální myšlenky filosofa a pedagoga Johna Deweye¹. Ten byl velkým příznivcem moderní demo-

¹Dewey, John (1859-1952), americký filozof, pedagog, psycholog a reformátor vzdělávání. Je považován za jednoho ze zakladatelů filozofie pragmatismu.

kracie. Ve svém díle upomínal, že změna převažuje kontinuity. Proto je podle něho nutné, aby došlo k přeformulování gramotnosti, která musí držet krok s nikdy nekončícím procesem změn. Je to zapotřebí, protože také lidská přirozenost a společnost se neustále mění. Pro Deweyho se věc stává složitější, když se zeptáme, co gramotnost vlastně znamená pro demokratickou společnost a její ideály. Rozlišuje „inertní“ a „osvobozující“ gramotnost. První význam znamená úroveň verbálních a numerických dovedností, aby bylo možné porozumět instrukcím, provádět rutinní procedury a vykonávat úkoly standardním způsobem. Ze sociální perspektivy jde o gramotnost, kde převládají tradice a zvyky, kde je vše na svém místě a inovace jsou omezo­vány. „Osvobozující“ gramotnost předpokládá standardy, které podporují rozvoj dovedností při hledání informací a sílu kritiky, reflexi a rozhodnutí ke změně. Dewey má na mysli „osvícení mas“, které vede ke skutečně „vitální“ demokracii.

Matematika se mění. Mění se také možnosti aplikací (nástupem počítačů). Matematické vědy dnes zahrnují kromě čisté matematiky také statistiku, finanční matematiku, části počítačové informatiky, operačního výzkumu a bioinformatiky (například aplikace statistiky v genetice). Tyto oblasti matematiky třebaže mají mnoho společného se základy matematiky, mají zvláštní charakter, metodologii, standardy a cíle.

Z aplikované matematiky je nutné vyzdvihnout zvláště statistiku, nebo spíše vědu o datech. Bohužel běžná školní matematika jí věnuje málo pozornosti. Na školách se nevytváří most mezi školní matematikou a interpretacemi dat, mezi abstraktními výpočty a světem statistiky v medicíně, zemědělství, psychologii, ekonomii, sociologii a politice.

V demokracii vede numerická a matematická ngramotnost zaručeně ke špatným rozhodnutím. Numerická a matematická gramotnost je však pouze jedním z předpokladů změny. Numerická a matematická gramotnost a sociální vědy mohou pomoci občanům dospět k rozhodnutím, která jsou zakotvena v aktuálních datech, matematických modelech, logice a ne v setrvačnosti a v neprůhledném uvažování.

Teorie, modely a matematika

Návrh teorie je hlavním cílem každé vědy. Obvykle se vyžaduje, aby teorie byla souborem obecných a logicky propojených úsporných tvrzení, které obsahují dobře definované pojmy s cílem formulovat vysvětlení specifikovaných fenoménů ve světě. Teorie mají podporovat přesnou a jasnou komunikaci, rigorózní testování, přesná měření a široké aplikace. Teorie mají mít deduktivní povahu v tom smyslu, že obsahují tvrzení, z kterých lze odvodit deduktivně další tvrzení. Teorie v sociálních vědách však většinou nemají uvedené vlastnosti. Jsou například induktivní ve smyslu, že nová tvrzení lze z nich odvodit pouze induktivně a intuitivně. (Ritzer, Ryan 2011)

Klasické teorie v sociálních vědách jsou silné v předmětném obsahu, ale až na výjimky nepoužívají nějaký formální jazyk, který by umožnil dedukce testovatelných predikcí o pojednávaných jevech. Aby se tato situace zlepšila, někteří výzkumníci či výzkumnice se zabývají vytvářením matematických modelů, přičemž určují jejich oborové předpoklady v matematických termínech, takže odvozené důsledky mohou být empiricky testovány pomocí srovnání s empirickými daty. Model určitých vztahů navazuje na jednu nebo více teorií, ale také nemusí.

V sociologii se matematické modely týkají sociálních struktur nebo sociálních procesů. Sociální sítě popisují vybrané sociální struktury. Ty se modelově reprezentují pomocí matic, ve kterých řádky a sloupce identifikují sociální jednotky a odpovídající prvky takových matic označují druh vztahu pro uvažovaný pár sociálních jednotek. Zájem o sociální struktury v tomto smyslu vedl k širokému užití matematických metod. Pomocí matematických modelů se také zachytily různé sociální procesy (sociální ovlivnění a sociální mobilita). Využívají se matematické prostředky jako teorie pravděpodobnosti, diferenciální rovnice, lineární algebra a stochastické procesy. Také se navrhuje počítačové simulace (např. agent-based modeling, ABM), pokud není možné matematicky nalézt

analytické řešení problému. Takto se zkoumají například otázky sociální emergence, sociální kooperace a vznik sociálního řádu.

Obecně jsou modely kognitivním nástrojem pro uspořádání lidských zkušeností a pozorování. Zkušenosti a pozorování se však liší mezi jedinci a mohou být organizovány mnoha různými způsoby. I když se pozorování sdílejí, mohou být různě prezentovány. Každý jedinec vlastně představuje jedinečný model skutečnosti. Tyto přirozené modely zkoumá etnografie.

Formální nebo matematický model vyjadřuje určitý výsek zkušenosti a pozorování v rámci vztahů vytvářejících formální systém definovaný logikou, matematikou nebo statistikou. Přitom nutně dochází k velké redukci informací.

“Sociální život je velmi komplexní. Je dán nekonečnou množinou informací. Neznamená to, že tuto komplexitu musíme celou zachytit. Většina užitečných modelů se soustředí na jeden základní proces. Reálné situace vznikají simultánním působením mnoha procesů.“ (Land, Farao 2006)

Matematika je lidská aktivita konstruování axiomatických definic abstraktních vztahů mezi nespécifikovanými anebo libovolným prvky, přičemž se tyto vztahy zkoumají deduktivním způsobem a užitím principů logiky. Každý takový vztah, který se v tomto kontextu objeví, definuje třídu matematických objektů, např. „Markovovy řetězce“ nebo „vektorové prostory“. Jestliže je T axiomatická teorie, která definuje třídu M matematických objektů, pak entita v M je T -modelem. Takové modely hrají ústřední roli ve vědě.

V sociálních vědách se využívají matematické modely v mnoha oblastech. Pro většinu výzkumníků a výzkumnic však tento přístup spadá v jedno s teorií statistiky a T -modely většinou představují takové entity jako lineární regrese nebo teorie statistických testů. Spojení mezi matematikou a sociálními vědami jde za toto použití aplikované statistiky.

Uvažujeme dva typy rozdílností matematických modelů v sociologii (Capocchi, V. et al 2010). Známější typ rozlišuje:

- a) statistické modely (klasifikační modely, modely vztahu proměnných),
- b) simulační modely.

Druhý typ klasifikace vymezil Edling (2002):

- a) modely procesů,
- b) modely struktury,
- c) modely racionální volby, agent base modely a modely umělých společností.

Oba tyto typy klasifikace modelů se překrývají, reprezentují však různé metodologické problémy a jinak přispívají k poznání.

Vyjmenujme dále čtyři důvody pro matematizaci sociálních věd.

- Reprezentace teorií. Matematika se používá pro formulaci teorií. Tím se stává struktura teorie více průhlednou. V mnoha případech stačí sestrojít několik rovnic pro vyjádření matematické části teorie.
- Explorace teorie. Jakmile je teorie popsána matematicky, je možné použít různé výpočty a odvodit kvalitativní a kvantitativní důsledky teorie. To pomáhá k jejímu lepšímu porozumění.
- Testování teorie. Predikce matematicky formulované teorie mohou být využity k jejímu testování, jestliže je konfrontujeme s daty o empirickém světě v Popperovském¹ smyslu (srov. Havlík 2015).
- Heuristiky. Pohled na matematicky formulovanou teorii může odhalit analogie k jiným fenoménům. To inspiruje k dalšímu výzkumu a vede k lepšímu porozumění daných fenoménů.

¹Popper, Karl (1902-1994), rakousko-britský filosof. Jeho vědecká činnost je zastoupena i na poli logiky, fyziky, biologie, sociologie a politologie. Svůj filozofický systém sám označil jako kritický racionalismus.

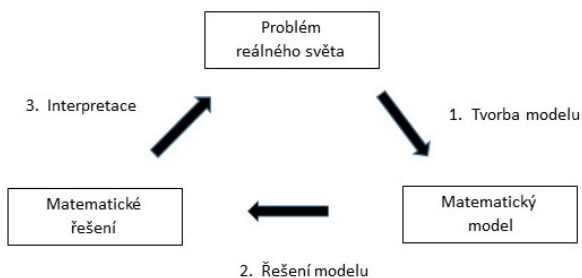
Funkce

Ústředním pojmem v modelování je vztah (korespondence). Se vztahem se setkáváme v běžném životě velmi často. Například:

- a) Každé dospělé osobě odpovídá roční příjem.
- b) Každé položce v obchodě odpovídá nějaká cena.
- c) Každému studentu odpovídá jeho průměrné ohodnocení.
- d) Výrobě x jednotek určitého výrobku ve firmě odpovídají nějaké celkové náklady.
- e) Délce strany čtverce odpovídá jeho plocha.

Ve vědě hledáme vztahy mezi různými fenomény. Pokud známe dobře vztah, můžeme dělat predikce. Například analýza nákladů by měla vést k nalezení matematického vztahu mezi počtem výrobků a celkovými náklady. V medicíně bychom rádi lépe poznali vztah mezi obezitou a srdečním onemocněním, v psychologii bychom rádi předpověděli úroveň výkonu pomocí počtu opakování daného úkolu v tréninku.

Matematické modelování v procesu matematizace znamená zkoumání a tvorbu teorie o reálném světě. Usiluje se přitom o přesné určení vztahů a jejich matematické vyjádření (Capecci et al. 2010). Proces modelování můžeme rozdělit do třech kroků (na obrázku).



První krok: Vytvoření matematického modelu (tedy matematizace problému), který vyřešením poskytuje informace o reálném světě.

Druhý krok: Vyřešení matematického modelu (hledání jeho parametrů).

Třetí krok: Interpretace řešení matematického modelu v pojmech původního problému v reálném světě.

Je nutné se přesvědčit, do jaké míry vytvořený model odpovídá skutečnosti.

Pro úplnost poznamenejme, že klasické schéma modelování není vždy použitelné. U zvláště komplexních vztahů jsme odkázáni na počítačové simulace, např. pomocí agent-based modelování: ABM (viz Edling 2002; Wilensky, Rand 2015). V mnoha oblastech ABM konkuruje modelování založené na rovnicích a proměnných. ABM považuje za jádro modelu množinu jedinců (nebo organizačních jednotek) a jejich interakce. Může modelovat heterogenní populace. Modely pomocí (diferenciálních) rovnic předpokládají určitou homogenitu, což je nevýhodou v oblastech, kde heterogenita hraje velkou roli. ABM se uplatňuje například v epidemiologii nebo v sociologii.

Matematika nám poskytuje mnoho abstraktních modelů. Zobrazení umožňuje jejich aplikaci (Barnett et al. 2015, Gill 2006). Při modelování hrají roli pojmy jako konstanta, proměnná, funkce a zobrazení. Každá teorie v empirických vědách je množinou tvrzení, které obsahují koncepty. Koncepty jsou představy, které nám pomáhají porozumět světu. Ke konceptům přiřazujeme konstanty a proměnné, abychom je mohli využít v teoriích. Proměnné a konstanty mohou být cokoliv, co považujeme za důležité pro teorii. Formálněji vyjádřeno, konstanta je koncept, která má jedinou fixní hodnotu. Proměnná je koncept, který může nabývat více hodnot z dané množiny hodnot. Koncepty a jejich vztahy jsou jádrem empirické vědy. Myšlení v pojmech konstant a proměnných je prvním krokem při vytváření teorií.