

Ekonomický výskum a tvorba odborných prác

Prednáška III - **Modely a dáta**

Matej Lorko

matej.lorko@euba.sk

www.lorko.sk

Odporúčaná literatúra:

- Dudenhefer, P. (2009). A guide to writing in Economics. EcoTeach Center and Department of Economics, Duke University.
- Neugeboren, R. H., & Jacobson, M. (2005). Writing Economics. Harvard University.
- Johnson, J. B., Reynolds, H. T., & Mycoff, J. D. (2015). Political science research methods. Cq Press.
- Friedman, S., Friedman, D., & Sunder, S. (1994). Experimental methods: A primer for economists. Cambridge University Press.

Ekonomické modely

- Ekonomovia tvoria modely tak, že zredukujú fenomén, ktorý ich zaujíma na jeho základné prvky a potom tieto prvky poskladajú spôsobom, ktorý sa v relevantných ohľadoch podobá danému fenoménu. Inými slovami: rozoberte to, zistite ako to funguje, potom to dajte znova dokopy a otestujte, či to stále funguje.
- Ekonomické modely špecifikujú vzťahy medzi dvoma druhmi premenných: exogénne premenné a endogénne premenné
- Exogénne premenné sú vstupmi do modelu, faktormi, ktoré ovplyvňujú to, čo sa stane, ale samy vznikajú „mimo“ modelu. Sú to dané, fixné hodnoty, o ktorých sa predpokladá, že sa v priebehu obdobia analýzy nezmenia. Endogénne premenné sú výstupmi modelu, sú teda určené „vnútri“. Zvyčajne sa na vyjadrenie vzťahu medzi exogénnymi a endogénnymi premennými používa matematická funkcia.
- Aplikácia základných modelov umožňuje predvídať reálnu ekonomiku. Vytváraním a používaním modelov sú ekonomovia schopní zamerať sa na jednoduché, niekedy jemné vzťahy v dátach a nájsť v nich príčinné súvislosti. Nájdenie vzoru v dátach umožňuje povedať niečo o tom, ako ekonomika funguje.
- Zhoda medzi modelom a realitou nie je nikdy dokonalá. Ale ak je model dostatočne dobrým zjednodušením reality, môžeme vďaka nemu lepšie predpovedať budúcnosť (ak sa budúce dáta s modelom zhodujú) a tiež lepšie porozumieť minulosti (ak sa zhodujú minulé dáta).
- Ako v každej vede, aj v ekonómii nie je možné teórie potvrdiť, ale iba vyvrátiť.
- Ak však model po dostatočný čas predpovedá dobre, jeho relevancia sa zvyšuje. Ak sú naše predpovede nesprávne, môžeme hľadať iné dáta, alebo model prepracovať. Prepracovaný model možno použiť na generovanie nových predpovedí, ktoré potom možno konfrontovať s novými dátami, ktoré môžu opäť priniesť nepotvrdenie predpovede a vynútiť si ďalšiu revíziu modelu.

Modely a dáta

- Pri prezentácii vašej hypotézy musíte prediskutovať súbor dát, ktorý použijete, a vo väčšine prípadov tiež váš model (zvyčajne nejaký druh regresie). Mali by ste identifikovať zdroj dát, a pomocou tabuľky, grafu alebo jednoduchých štatistík ich zosumarizovať.
- Mali by ste vysvetliť, ako súvisia dáta s vašou hypotézou, a zaznamenať všetky problémy, ktoré by mohli spôsobiť. Ak máte iba malý súbor pozorovaní alebo musíte použiť proxy pre dáta, ktoré nemôžete priamo pozorovať, mali by ste to explicitne vyjadriť.
- Vo vašej práci možno nebude možné dosiahnuť presvedčivé empirické výsledky. Možno máte neúplné dáta, alebo vaše regresné koeficienty nebudú štatisticky významné, alebo ste nemohli odkontrolovať všetky významné nezávislé premené. Je lepšie tieto nedostatky uznať, ako robiť príliš široké a nepodložené tvrdenia.

Meranie

- Pred testovaním hypotéz musíme pochopiť niektoré problémy spojené s meraním javov, ktoré sme sa rozhodli preskúmať. Tiež je potrebné porozumieť spôsobu ako vyvodiť premenné, ktoré reprezentujú sledované javy - ako systematicky zaznamenávať pozorovania pomocou číslíc alebo skóre.
- To, ako vedci merajú sledované javy, môže mať významný vplyv na ich zistenia; rozdiely v meraní môžu viesť k úplne odlišným záverom.
- Je užitočné myslieť na definíciu premenných ako na poslednú fázu procesu presnej definície javu. Často začíname abstraktným pojmom (napríklad demokraciou), potom sa ho pokúsime zmysluplne definovať a nakoniec sa rozhodneme, ako ho budeme merať.
- Na konci tohto procesu by sme mali dostať definíciu, ktorá má zmysel, je blízka výkladu pojmu a zároveň jednoznačne určuje spôsob merania.
- Aby boli merania premenných schopné javy vysvetliť, musia tiež poskytnúť výskumníkovi dostatok informácií na to, aby mohol robiť porovnania a štatistické testy. Preto sa kvalita meraní posudzuje z hľadiska ich jednoznačnosti a presnosti.

Presnosť merania

- Pokiaľ ide o presnosť meraní, existujú dve hlavné riziká. Merania môžu byť nepresné, pretože sú nespoľahlivé a / alebo nie sú validné.
- Spoľahlivosť predstavuje konzistentnosť merania v opakovaných testoch alebo pokusoch. Spoľahlivý spôsob merania je taký, ktorý pri každom použití prináša rovnaký výsledok. Nespoľahlivé je teda meranie, ktoré prináša nekonzistentné výsledky - niekedy vyššie, niekedy nižšie.
- Spoľahlivosť meraní v ekonómii sa dá určiť viacerými spôsobmi.
 - Metóda opakovaného testu - použitie rovnakého „testu“ na rovnaké pozorovania po určitom čase a potom porovnanie výsledkov dvoch meraní.
 - Metóda určenia spoľahlivosti merania prostredníctvom alternatívneho merania - meranie toho istého atribútu viackrát, ale použitím dvoch rôznych spôsobov merania.
 - Metóda určenia spoľahlivosti merania prostredníctvom “split-halves” - tiež ide o použitie dvoch spôsobov merania rovnakého javu, ale tentokrát súčasne. Výsledky týchto dvoch meraní sa potom porovnávajú. Táto metóda zabráni problému, ktoré vznikajú, ak sa jav môže medzi meraniami meniť.
- Validné meranie je také, ktoré meria to, čo sa má merať. Na rozdiel od spoľahlivosti, ktorá závisí od toho, či opakované použitie tých istých alebo ekvivalentných mier vedie k rovnakému výsledku, sa validita týka súladu medzi meraním a javom.

Vzorky

- Predpokladajme, že chceme posúdiť úroveň podpory pre niektorú navrhovanú vládnu politiku na Slovensku. Pretože nie je možné porozprávať sa so všetkými občanmi, musíme si z nich vybrať určitú vzorku.
- Vzorka je akákoľvek podmnožina nejakej populácie. Veľkosť vzorky a spôsob výberu jej členov určuje kvalitu (tj. presnosť a spoľahlivosť) záverov o celej populácii.
- Rozhodnutie výskumníka, či bude zbierať údaje za populáciu alebo za vzorku, je zvyčajne dané praktickými dôvodmi. Výhodou vzorky je často úspora času a peňazí. Nevýhodou je, že informácie založené na vzorke sú zvyčajne menej presné alebo chybovejšie ako informácie zhromaždené od celej populácie.
- Po zhromaždení vzorky je nutné zmerať jej vlastnosti alebo charakteristiky. Najzaujímavejšími atribútmi v empirickom výskume sú číselné alebo kvantitatívne ukazovatele, ako sú percentá alebo priemery. Tieto štatistiky sa potom používajú na aproximáciu zodpovedajúcich hodnôt pre celú populáciu.
- S cieľom zmierniť odchýlku vzorky od populácie by v ideálnom prípade mal mať každý prvok v celkovej populácii konkrétnu pravdepodobnosť zahrnutia do vzorky. Tieto znalosti umožňujú výskumníkovi vypočítať, ako presne vzorka odráža populáciu, z ktorej je vybraná.

Čo sa dá naučiť zo vzoriek?

- Vzorky poskytujú iba odhady alebo aproximácie atribútov populácie. Tieto odhady môžu byť niekedy úplne správne, ale väčšinou sa budú líšiť od skutočnej hodnoty atribútu v populácii.
- Keď uvádzame štatistiku vzorky, vždy predpokladáme, že bude existovať odchýlka, teda rozdiel medzi reportovanými a skutočnými hodnotami.
- Odkiaľ sa berie strata presnosti? Odpoveďou je náhoda, vyskytujúca sa pri výbere členov vzorky. Ak si hodíte desaťkrát mincou, pravdepodobne vám nepadne presne päť hláv, aj keď je minca spravodlivá a pravdepodobnosť hláv je presne 50 percent. Náhodnosť je vrodenu vlastnosť prírody, minimálne na škále, ktorú dokážeme pozorovať.
- Rovnako ako v prípade hodu mincou, nie je pravdepodobné, že náhodná vzorka napríklad desiatich (alebo dokonca oveľa väčšieho počtu) členov presne vystihne celkovú populáciu. Ak ale postupujeme správnym spôsobom a boli splnené určité predpoklady (napríklad vzorka je realizovaná náhodným výberom z dostatočne veľkej populácie), štatistika vzorky sa približuje skutočnej hodnote atribútu v populácii.

Typy dát a techniky ich zberu

- Výskumník zhromažďuje dáta o správaní buď pozorovaním samotného správania (priame pozorovanie), alebo prostredníctvom nejakej fyzickej stopy správania (nepriame pozorovanie).
- Dáta zhromaždené prostredníctvom pozorovania z prvej ruky sú príkladom primárnych dát, to znamená dát zaznamenaných a použitých samotným výskumníkom.
- Dáta z rozhovorov alebo z písomného záznamu môžu byť primárne alebo sekundárne, teda zozbierané niekým iným.
- Študenti často nájdu vhodné sekundárne dáta zadarmo vo verejne prístupných archívoch. Ak ale chcete použiť primárne dáta (priame alebo nepriame), musíte si nájsť spôsob, ako pozorovanie realizovať. Pozorovanie môže byť štruktúrované alebo neštruktúrované.
 - Pri štruktúrovanom pozorovaní sa systematicky zaznamenáva výskyt konkrétneho správania. Pred začatím zhromažďovania dát sa výskumník musí rozhodnúť, aké správanie je pre neho relevantné - čo presne chce pozorovať.
 - Pri neštruktúrovanom pozorovaní sa za relevantné na začiatku považuje akékoľvek správanie - zaznamená sa všetko a až neskôr sa rozlišuje medzi tým, čo je dôležité a čo nie.

Prieskumy a dotazníky

- Ako zabezpečiť validitu a spoľahlivosť dát z prieskumov a dotazníkov?
 - Požadované informácie musia byť pre respondenta dostupné (tj. nesmú byť zabudnuté alebo nepochopené).
 - Respondent musí vedieť, čo je pre výskumníka relevantná a primeraná odpoveď.
 - Respondent musí byť motivovaný informácie výskumníkovi poskytnúť.
 - Respondent musí vedieť ako tieto informácie poskytnúť.
 - Výskumník musí odpovede respondenta zaznamenať presne.
 - Odpovede musia odrážať postoje respondenta, nie výskumníka.
 - Iní ľudia musia chápať otázky a odpovede rovnakým spôsobom ako respondent a výskumník.

Dotazníkový prieskum

- V dotazníkovom prieskume skupina jednotlivcov vyplňa viac alebo menej štandardizované dotazníky. Dotazníky môžu mať rôznu podobu na preskúmanie rôznych hypotéz, nezahŕňajú však spontánne rozhovory.
- Aj keď prieskumy môžu byť pomerne rýchle a lacné, je potrebné, aby výskumník myslel na:
 - Percento dokončenia dotazníka - Ak je nízke, môže byť schopnosť výskumníkov urobiť štatistické závery pre študovanú populáciu obmedzená. Tí, ktorí dotazník vyplnia do konca, sa tiež môžu systematicky líšiť od tých, ktorí ho nedokončia, čo vedie k ďalším odchýlkam. Zväčšenie veľkosti prieskumnej vzorky s cieľom kompenzovať nízku mieru odozvy často iba zvýši náklady a pritom problém vôbec nevyrieši.
 - Zhoda medzi vzorkou a populáciou - teda to, ako dobre členovia vzorky reprezentujú populáciu, je vždy hlavným problémom. Odchýlky môžu vznikáť už v momente počiatočného výberu respondentov, alebo aj neskôr, prostredníctvom neúplných odpovedí.
 - Dĺžka dotazníka - ak prieskum kladie nadmerné množstvo otázok alebo zaberá respondentom príliš veľa času, môžu stratiť záujem alebo začať odpovedať bez veľkého premýšľania.

Kvalita odpovedí

- Kvalita odpovede = miera, do akej odpovede poskytujú presné a úplné informácie. Je to kľúč k vyvodu valídnych záverov.
- Kvalita odpovede závisí od niekoľkých faktorov, medzi ktoré patrí motivácia respondentov, ich schopnosť porozumieť pokynom a riadiť sa nimi, a čo je najdôležitejšie, kvalita kladených otázok.
- Upútanie respondentov - je dôležité dobre začať - tým, že predstavíte seba, svoju organizáciu, svoj cieľ, oceníte ich čas, poukážete na vašu nestannosť a vaše povedomie o dôležitosti anonymity, a tiež prejavíte ochotu podeliť sa o svoje zistenia.
- Pretože hlavným cieľom prieskumu je presne zmerať postoje respondentov a ich správanie prostredníctvom otázok, musíme venovať čas diskusii o dobrých a zlých otázkach.
- Dobré otázky vedú k presným odpovediam; zlé otázky tvoria nevhodné stimuly a vedú k nespoľahlivým alebo nepresným odpovediam. Pri tvorbe otázok by výskumník mal použiť objektívne a jasné slová. V opačnom prípade vám hrozí, že zozbierate neúplné dotazníky a nezmyselné dáta. Základné pravidlo je toto: respondenti musia byť schopní porozumieť otázkam a dokázať na nich odpovedať.
- Niektoré typy otázok situáciu pre respondentov zbytočne sťažujú. Patria sem nejednoznačné a nábehové otázky a tiež otázky s viacerými časťami.

Uzavreté otázky

- Hlavnou výhodou uzavretej otázky je, že je ľahké a rýchle na ňu odpovedať. Ďalšou výhodou je, že odpovede sa dajú ľahko porovnávať, pretože všetky odpovede spadajú do fixného počtu vopred určených kategórií. Tieto výhody umožňujú rýchlu štatistickú analýzu dát.
- Naopak, pri otvorených otázkach musí výskumník prečítať každú odpoveď, potom určiť, ktoré odpovede sú ekvivalentné, a následne rozhodnúť, koľko kategórií alebo rôznych typov odpovedí nakódovať a každej odpovedi priradiť kód, aby bolo možné dáta analyzovať.
- Ďalšou výhodou uzavretých otázok oproti otvoreným je, že respondenti sú zvyčajne ochotní odpovedať na osobné alebo citlivé témy (napríklad príjem, vek, alebo politické názory) výberom kategórie, ale už menej ochotní uviesť presnú cifru.
- Kritici uzavretých otázok kladú dôraz na to, že nútia respondenta, aby si vybral kategóriu odpovede, ktorá nemusí presne zodpovedať jeho postoju. Preto má odpoveď o trochu menší význam.
- Uzavreté otázky sú tiež často formulované tak, aby si respondent musel vybrať medzi alternatívami alebo určiť, ktorá z nich je preferovaná. To môže mať za následok príliš zjednodušený a skreslený obraz. V niektorých situáciách môže byť vhodnejšia uzavretá otázka, ktorá umožní respondentom zvoliť viac ako jednu odpoveď.

Otvorené otázky

- Neštruktúrované otázky s voľným textom umožňujú respondentom uviesť presne to, čo chcú. Nie sú nútení voliť medzi fixnými odpoveďami, ktoré možno nie sú relevantné.
- Nevýhodou otvorenej otázky je, že respondenti môžu odpovedať príliš veľa alebo príliš málo. Niektorí môžu odpovedať príliš zdĺhavo a komplikovane. Na druhej strane, niektorí respondenti so slabšími vyjadrovacími schopnosťami nemusia odpovedať, čo môže spôsobiť skreslenia v dátach.
- Štruktúrované odpovede môžu byť navyše ťažko kódovateľné, interpretácie odpovedí sa môžu líšiť (čo ovplyvňuje spoľahlivosť dát) a spracovanie odpovedí môže byť časovo náročné a nákladné.

Poradie otázok

- Prvých niekoľko otázok v dotazníku je zvyčajne dobré využiť na prelomenie ľadov. Ide o všeobecné otázky, na ktoré sa dá odpovedať ľahko.
- Komplexné a konkrétne otázky môžu spôsobiť, že respondenti dotazník nevyplnia do konca, pretože si myslia, že to bude príliš náročné.
- Otázky týkajúce sa osobných alebo citlivých tém sa zvyčajne nechávajú na koniec.

Popis dát

- Najlepším spôsobom, ako sa naučiť písať o dátach, je prečítať si dátové sekcie v iných článkoch a venovať sa druhu informácií, ktoré obsahujú. Vaša sekcia s dátami by mala obsahovať minimálne toto:
- Identifikujte zdroj dát. Zahrňte vetu, ktorá explicitne hovorí, odkiaľ vaše dáta pochádzajú.
- Popíšte dáta. Mali by ste napríklad uviesť počet pozorovaní, počet vzoriek, časové obdobie, počas ktorého sa dáta zhromažďovali, spôsob zberu dát atď.
- Uvedte silné a slabé stránky dát. Aké sú vaše dáta v porovnaní s inými dátami použitými v literatúre? Poskytuje váš výskum viac pozorovaní alebo novších pozorovaní než iné výskumy? Boli údaje zhromaždené spoľahlivejším spôsobom? Prečo sú tieto dáta pre vašu štúdiu obzvlášť vhodné (alebo nie)? Všimnite si všetky vlastnosti dát, ktoré môžu ovplyvniť vaše výsledky. Boli určité skupiny populácie nadmerne alebo nedostatočne zastúpené? Existuje skreslenie spôsobené prirodzeným úbytkom (attrition) alebo samo-výberom (selection bias)? Zmenil sa počas zhromažďovania dát spôsob merania?
- Vysvetlite všetky vykonané výpočty alebo úpravy. Dáta vám niekedy niečo neposkytujú priamo; možno ste museli pridať / odčítať / vynásobiť / rozdeliť dva zadané údaje, aby ste získali tretí. Popíšte, ako ste vybrali svoju vzorku. Museli ste napríklad vylúčiť určité druhy pozorovaní?

Deskriptívne štatistiky

- Dátové sekcie v článkoch často obsahujú tabuľku deskriptívnych štatistík - teda štatistík ktoré určujú relevantnosť vzorky. Zvyčajne zahŕňajú priemer (napr. priemerný príjem, priemerný vek, priemerné roky školskej dochádzky atď.) a štandardnú odchýlku. V prípade kategorických údajov (napríklad rasy) však neuvádzate priemer; namiesto toho použijete percento pozorovaní v každej skupine.
- Očakávaná hodnota - Priemerná hodnota atribútu vzorky, založená na opakovanom výbere vzoriek z populácie.
- Štandardné chyby - smerodajná odchýlka alebo miera variability / rozptylu vo vzorke. Čím väčšia je vzorka, tým menšia je štandardná chyba.
- Distribúcie vo vzorke - Teoretické (nepozorované) rozdelenie atribútu, ktoré umožňuje výpočet intervalov spoľahlivosti a testy hypotéz.
- POZNÁMKA: Priemer a štandardná odchýlka fungujú dobre pre normálne rozdelenie (v tvare zvonovej krivky). Ak natrafíme na inú distribúciu, môže byť užitočnejšie použiť na opis centrálnej tendencie (očakávanej hodnoty) medián alebo modus.

Grafické znázornenie dát

- Dobre zostavený graf dokáže odpovedať na niekoľko otázok naraz:
- Centrálna tendencia: Kde leží stred distribúcie?
- Rozptyl alebo variácia: Ako veľmi sú pozorovania rozťahnuté alebo koncentrované?
- Tvar distribúcie: Má distribúcia iba jediný vrchol (jedna koncentrácia pozorovaní v relatívne úzkom rozmedzí hodnôt), alebo má vrcholov viac?
- Chvosty: Približne aké percento pozorovaní leží na koncoch (chvostoch) distribúcie?
- Symetria alebo asymetria (nazývaná tiež nesúmernosť): Majú pozorovania tendenciu hromadiť sa na jednej strane distribúcie, zatiaľ čo na druhej ich je relatívne málo? Alebo má každá strana distribúcie zhruba rovnaký počet pozorovaní?
- Odľahlé hodnoty (outliers): Existujú hodnoty, ktoré sa v porovnaní s väčšinou javia ako veľmi veľké alebo veľmi malé?
- Porovnanie: Ako sa dve distribúcie líšia z hľadiska tvaru, natiahnutia a centrálnej tendencie?
- Vzťahy: Je možné, že hodnoty jednej premennej súvisia s hodnotami inej?

Výber správneho grafu

TABLE 11-10 Typical Presentation and Exploratory Graphs

Type of Graph	What Is Displayed	Most Appropriate Level of Measurement	Number of Cases	Comments
Bar chart	Relative frequencies (percentages, proportions)	Categorical (nominal, ordinal)	3-10 categories	Common presentation graphic
Dot chart	Frequencies, distribution shape, outliers	Quantitative (interval, ratio)	<i>Less than 50 cases</i>	Displays actual data values
Histogram	Distribution shape	Quantitative	$N > 50$ cases	Essential exploratory graph for interval or ratio variables with a large number of cases
Boxplot	Distribution shape, summary statistics, outliers	Quantitative	$N > 50$ cases	Can display several distributions; actual data points, an essential exploratory tool
Time series plot	Trends	Quantitative (percentages, rates)	$10 < N < 100$	Common in presentation and exploratory graphics