

Ekonomický výskum a tvorba odborných prác

Prednáška IV - Štatistická analýza

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.

Testovanie hypotéz

- Existujú dva typy štatistických hypotéz: nulové hypotézy a výskumné alebo alternatívne hypotézy.
- Nulové hypotézy sú jednoznačné tvrdenia o atribútoch populácie, napríklad že priemer sa rovná určitej hodnote, že dve proporcie sa medzi sebou nelíšia, alebo že numerický indikátor vzťahu medzi dvoma premennými je nulový. V mnohých výskumných článkoch nulová hypotéza tvrdí, že niečo (napríklad stredná hodnota alebo proporcia) sa rovná nule. Preto ju nazývame nulová. Nula predstavuje žiaden efekt, napríklad žiadnen rozdiel.
- Okrem stanovenia nulovej hypotézy výskumníci uvádzajú aj ďalšiu hypotézu, ktorá sa volá výskumná alebo alternatívna hypotéza. Zvyčajne dúfajú, že budú schopní odmietnuť nulovú hypotézu v prospech svojej výskumnej hypotézy.
- Pri testovaní hypotéz, teda pri určovaní platnosti nulovej hypotézy, môžu nastať dva druhy chýb. Prvým typom chyby je odmietnutie pravdivej nulovej hypotézy. Štatici túto chybu nazývajú chybou typu I. Ďalšou možnou chybou je, ak dáta nedokážu odmietnuť nulovú hypotézu, ktorá nie je pravdivá. Tento typ chyby sa nazýva chyba typu II.

Testovanie hypotéz

- Na označenie pravdepodobnosti chyby typu I sa používa tzv. úroveň štatistickej významnosti. Tri najbežnejšie úrovne štatistickej významnosti v ekonómii sú 0,05, 0,01 a 0,001
- Ak na získanie odhadov atribútu populácie použijeme dostatočne veľa vzoriek, budú naše odhady normálne distribuované a zoskupené okolo skutočnej hodnoty atribútu. Distribúcie nám naznačujú pravdepodobnosť, že naše odhady spadajú do určitej vzdialenosti od skutočnej hodnoty atribútu. Táto pravdepodobnosť sa nazýva hladina spoľahlivosti. Pre každú hladinu spoľahlivosti existuje konkrétny interval spoľahlivosti - teda rozsah hodnôt, ktoré majú danú pravdepodobnosť alebo hladinu spoľahlivosti.
- Všeobecná forma intervalu spoľahlivosti je nasledovná: Odhadovaná hodnota parametra \pm štandardná chyba \times kritická hodnota.

TABLE 12-1 Types of Inferential Errors

	In the "Real" World, the Null Hypothesis Is . . .	
Decision is to	True	False
Accept H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Korelácia vs. kauzálnosť

- Fundamentálnym problémom, ktorému budete čeliť pri empirickom výskume v ekonómii je odlíšenie kauzálnosti (príčinnej súvislosti) od korelácie.
- Hovoríme, že dve premenné korelujú, ak sa pohybujú súčasne.
- Takýto vzťah je však kauzálny iba v prípade, že jedna z premenných spôsobuje pohyb druhej.
- Identifikačný problém: ak objavíte, že dve premenné sú korelované, ako zistíte, či jedna spôsobuje druhú?

Korelácia vs. kauzálnosť

- Postup analýzy
 - Krok 1: Zdokumentovanie korelácie, teda, či sa dve premenné pohybujú súčasne.
 - Krok 2: Vyšetrenie, či pohyby jednej premennej spôsobujú pohyby druhej.
 - Pre akúkoľvek koreláciu medzi dvoma premennými A a B môžu existovať 3 vysvetlenia:
 - A ovplyvňuje B
 - B ovplyvňuje A
 - Existuje iný faktor, ktorý ovplyvňuje zároveň A aj B
- Dobre navrhnutý výskum musí demonštrovať platné kauzálne závery. V podstate musí preukázať 3 veci:
 - 1. Kovariácia: Výskum demonštruje, že predpokladaná príčina (X) naozaj koreluje s jej predpokladaným efektom (Y).
 - 2. Časové poradie: Výskum demonštruje, že príčina predchádzala následku: X musí nastať skôr ako Y.
 - 3. Eliminácia možných alternatívnych príčin („rušivých faktorov“): Výskum musí demonštrovať, že boli vylúčené všetky možné spoločné príčiny X a Y.

Zlatý štandard pre posudzovanie kauzálnosti: RCT

- Z korelácie teda ešte priamo nemusí vyplývať kauzálnosť. Inými slovami, to, že sú dve premenné korelované, nám neposkytuje žiadnu schopnosť predikcie, ak nevieme, či je medzi nimi aj kauzálny vzťah. Ako kauzálnosť odhaliť? Najlepším riešením sú RCT.
- RCT (randomized control trial) - experiment designovaný priamo na odhalenie kauzálnosti a to tak, že skupina jednotlivcov je náhodne (hodom mince) rozdelená do intervenčnej skupiny, na ktorej je testovaná určitá zmena alebo intervencia a do kontrolnej skupiny, ktorá intervenciu nedostane.

Zlatý štandard pre posudzovanie kauzálnosti: RCT

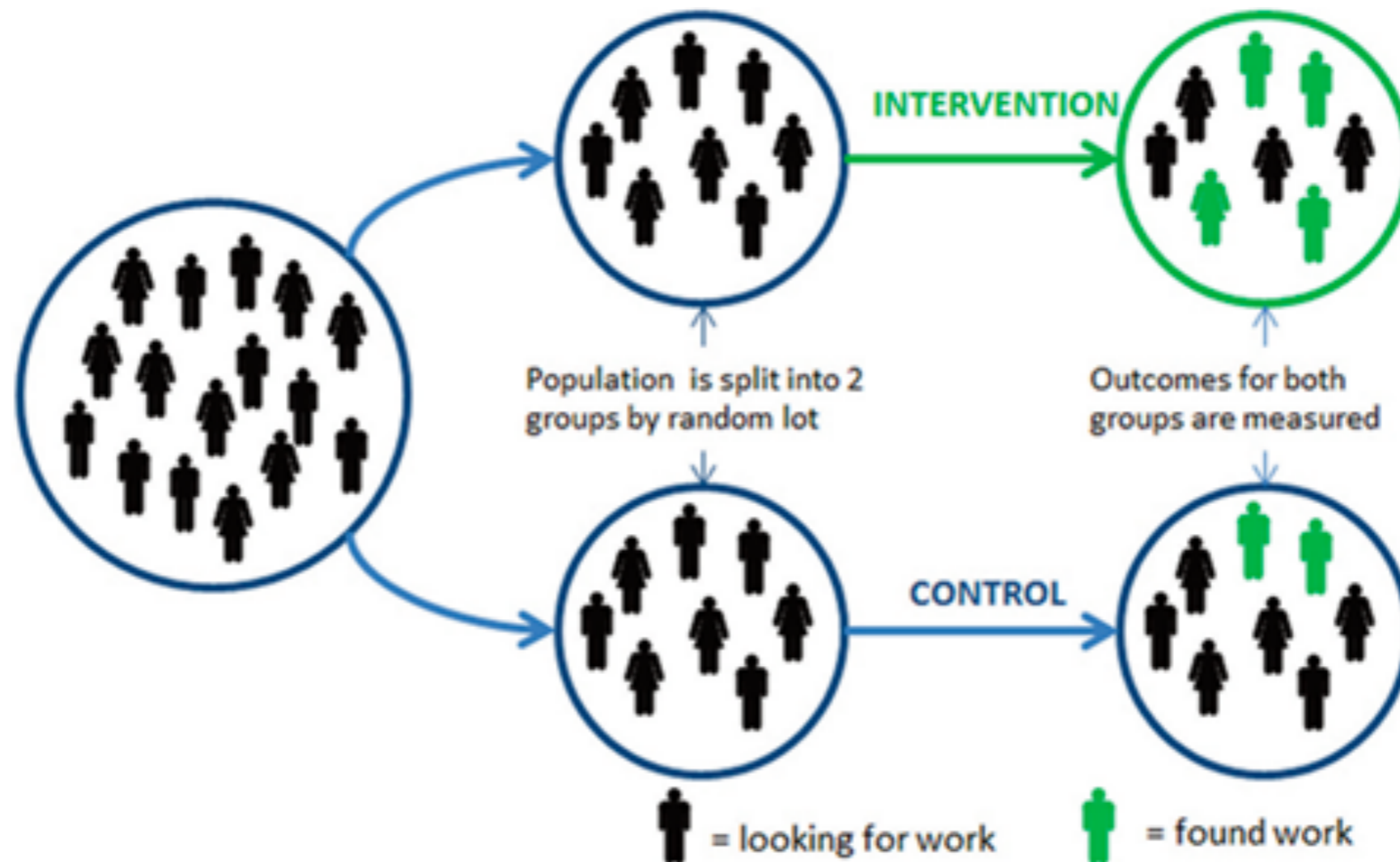


Figure 1. The basic design of a randomised controlled trial (RCT), illustrated with a test of a new 'back to work' programme.

Problém skreslenia (bias)

- Každú analýzu empirickej metodológie by sme mali začať jednoduchou otázkou - líšia sa intervenčná a kontrolná skupina aj z nejakého iného pohľadu, ako iba intervenciou?
- Rozdiely medzi intervenčnou a kontrolnou skupinou, ktoré nemajú súvis so samotnou intervenciou sú fundamentálnym problémom pri určovaní kauzálnosti z korelácie. Tieto rozdiely nazývame “skreslenia” (anglicky bias). Skreslenie je akýkoľvek zdroj rozdielov medzi intervenčnou a kontrolnou skupinou, ktorý je s intervenciou korelovaný, ale nie ňou spôsobený.
- Keďže sú pri RCT jednotlivci rozdeľovaní do skupín náhodne, nemali by medzi skupinami (pri dostatočne veľkej vzorke jednotlivcov - zákon veľkých čísel) takéto systematické rozdiely spôsobujúce skreslenia existovať. Tým pádom by mala byť jedinou odlišnosťou medzi skupinami práve intervencia - a to je dôvod, prečo sú RCT zlatým štandardom pre určovanie kauzálnych efektov intervencií.

Nedostatky RCT

- Pre mnohé otázky nemôžeme RCT použiť, napr. preto, že experiment by bol príliš drahý, trval príliš dlho, alebo by bol eticky nepriechodný (napr. niektoré nové medicínske procedúry)
- Navyše, aj metóda RCT má svoje nedostatky. Prvý je, že výsledky sa merajú zo vzorky jedincov, často dobrovoľníkov, ktorá môže byť trochu odlišná od charakteristík všeobecnej populácie. Napríklad môžu mať účastníci štúdie menšiu averziu k riziku, alebo môžu byť vážne (zúfalo) chorí. Tým pádom sa môže stať, že výsledok RCT nebude validný pre priemerného človeka z populácie.
- Druhým problémom je prirodzený úbytok (attrition). Jednotlivci môžu experiment opustiť skôr, ako sa skončí. Ak by experiment opúšťali náhodne, nebol by to problém, lebo by vzorky ostali vyvážené. Predpokladajme však, že intervencia má pozitívne efekty na polovicu intervenčnej skupiny a negatívne na druhú polovicu, a že tí, na ktorých pôsobí intervencia negatívne, experiment opustia. Keďže v experimente ostanú iba tí, ktorým intervencia pomáha, môže sa stať, že na konci štúdie budeme nesprávne konštatovať, že intervencia má pozitívny efekt.
- prirodzený úbytok (attrition): zmenšenie vzorky v priebehu experimentu. Toto zmenšenie, ak nie je náhodné, môže viesť k skresleniu výsledkov.

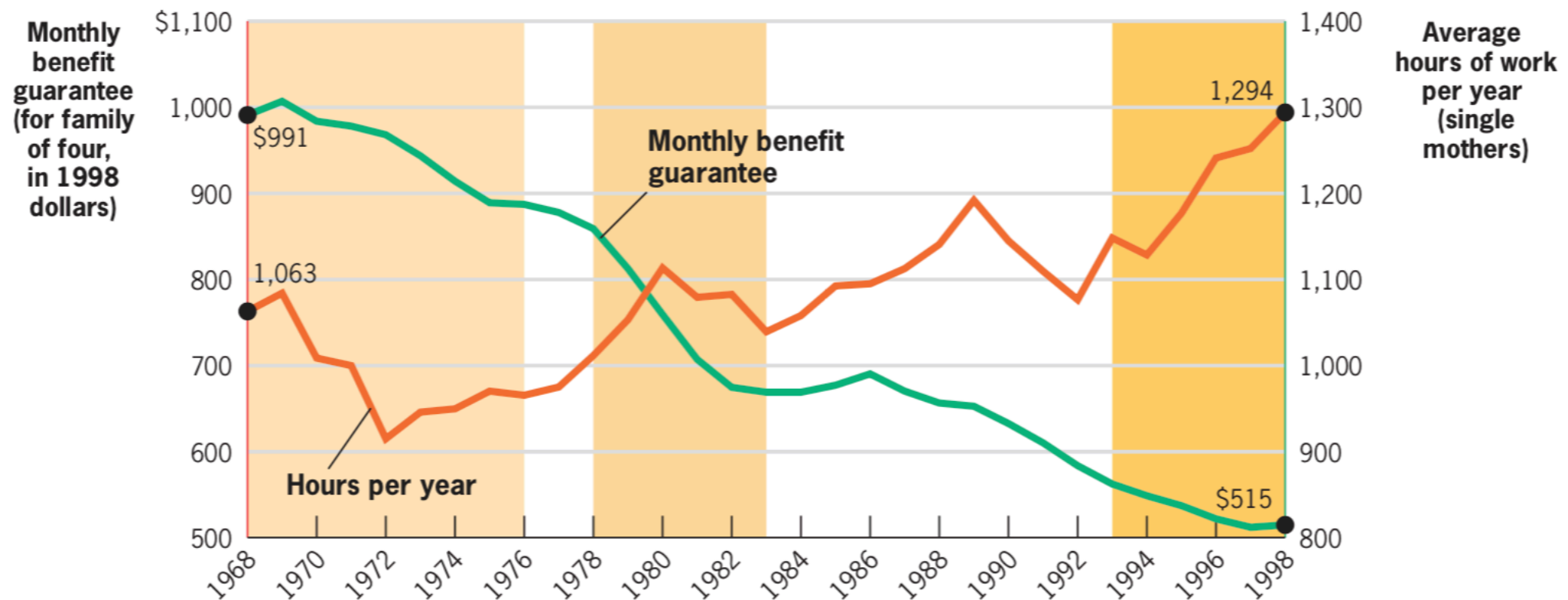
Určovanie kauzálnosti z neexperimentálnych dát

- Ak nie sú k dispozícii dáta z RCT, musíme sa spoliehať na klasické dáta z pozorovaní (observácií)
- Observačné dáta: dáta vygenerované správaním jednotlivcov v “reálnom svete”, teda mimo zámerne nadizajnovných experimentov
- Napríklad pri nových liekoch môžeme analyzovať, namiesto RCT dát, dáta o tom, kto liek užíval a aké boli výsledky
- Samozrejme, observačné dáta sú skreslené (napr. vitamíny užívajú iba špecifické typy ľudí, nie všetci), našťastie však máme k dispozícii niekoľko nástrojov, ktorými vieme so skreslením pracovať tak, aby bolo možné (s určitou nutnou dávkou rezervy) určovať kauzálnosti aj v takýchto, neexperimentálnych dátach

Analýza časových radov

- Pri určovaní kauzálnosti z observačných dát sa často využíva analýza časových radov, teda analýza zmien dvoch premenných v čase

■ FIGURE 3-1



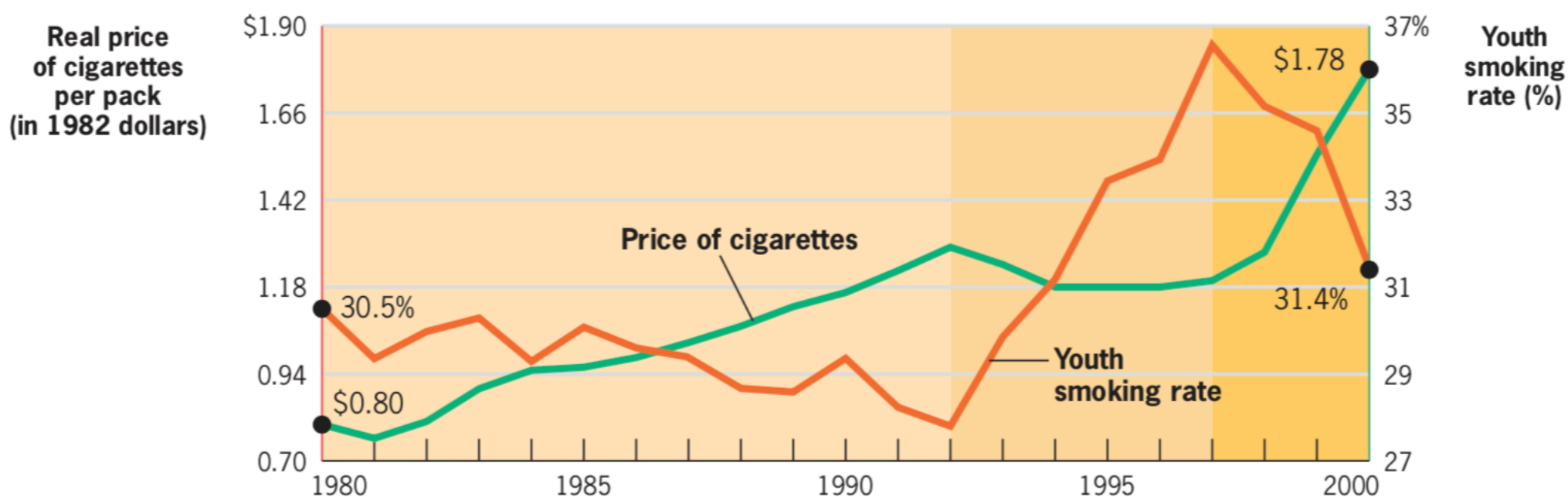
Average Benefit Guarantee and Single Mother Labor Supply, 1968–1998 • The left-hand vertical axis shows the monthly benefit guarantee under cash welfare, which falls from \$991 in 1968 to \$515 in 1998. The right-hand vertical axis shows average hours of work per year for single mothers, which rises from 1,063 in 1968 to 1,294 in 1998. Over this entire 30-year period, there is a strong negative correlation between the average benefit guarantee and the level of labor supply of single mothers, but there is not a very strong relationship within subperiods of this overall time span.

Analýza časových radov

- Nedostatky
 - Ani veľmi silná korelácia premenných v čase ešte nemusí znamenať kauzálny vzťah, keďže v hre môžu byť aj iné faktory, ktoré analýza nemusí zachytiť. Takéto faktory môžu spôsobovať skreslenia, ak sú tiež korelované so sledovanými premennými.
- Kedy je užitočná?
 - Hlavne ak v dátach nájdeme výrazné skoky, ktoré nie sú spojené s tretími faktormi. Naopak, časové rady nám nedokážu veľa povedať, ak nájdeme dlhé a pomerne nevýrazné trendy.

Analýza časových radov

■ FIGURE 3-2



Real Cigarette Prices and Youth Smoking, 1980–2000 • The left-hand vertical axis shows the real price of cigarettes per pack, which rises from \$0.80 in 1980 to \$1.78 in 2000. The right-hand vertical axis shows the youth smoking rate (the share of high school seniors who smoke at least once a month), which fell from 1980 to 1992, rose sharply to 1997, and then fell again in 2000 to roughly its 1980 level. There is a striking negative correspondence between price and youth smoking within subperiods of this era.

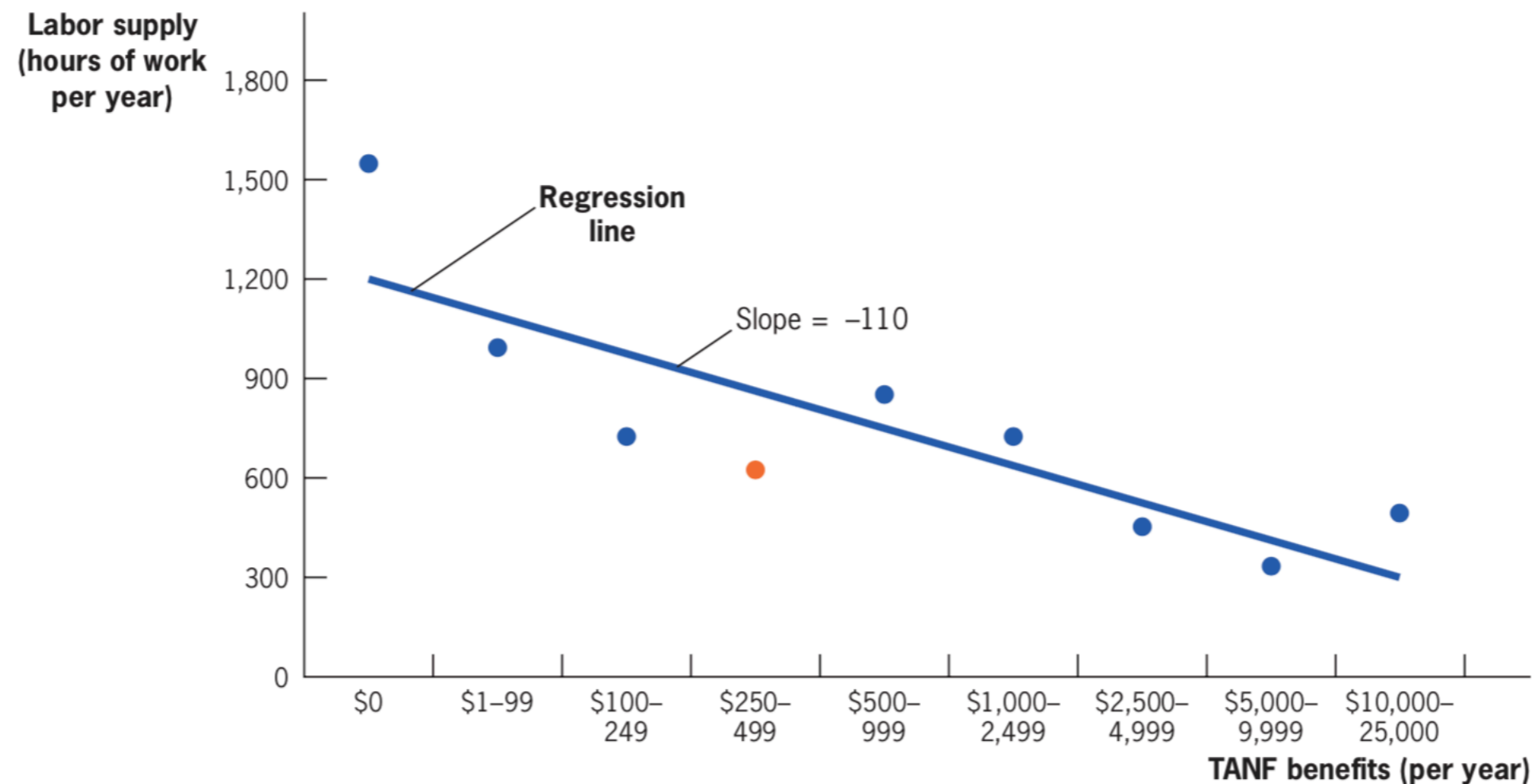
Source: Calculations based on data on smoking from Monitoring the Future survey and on tobacco prices from the Tobacco Institute.

Prierezová regresná analýza

- Ďalším postupom pre identifikovanie kauzálnych efektov je prierezová regresná analýza, ktorá vyhodnocuje vzťah dvoch premenných, pričom drží ostatné faktory konštantné. “Prierezová” znamená porovnanie množstva jednotlivcov v jednom časovom bode (teda nie vo viacerých obdobiach ako to robia časové rady)
- Regresná analýza popisuje (a kvantifikuje) vzťah medzi premennou, ktorej vývoj sa snažíme objasniť (závislá premenná) a súborom premenných, o ktorých sa domnievame, že by ju vysvetliť mohli (nezávislé premenné)
- Odhad takéhoto vzťahu zobrazuje regresná priamka. Keďže priamka, ktorá by prešla všetkým bodom v dátovom súbore neexistuje (museli by sme nakresliť krivku), lineárna regresia hľadá priamku, ktorá je k bodom čo najbližšie
- Technicky priamka identifikuje súčet najmenších štvorcov, teda kvadrátov vzdialenosti jednotlivých bodov od priamky. Táto technika je tým pádom citlivá na body, ktoré sú od nej príliš ďaleko (outliers), keďže im dáva veľmi vysokú váhu a teda vplyv. Niekedy je preto nutné zvážiť aj postupy, ktoré budú k vzdialeným pozorovaniam menej citlivé.

Prierezová regresná analýza

■ FIGURE 3-4



TANF Benefit Income and Labor Supply of Single Mothers, Using CPS Data • Using data from the CPS, we group single mothers by the amount of TANF income they have. Those who are receiving the lowest level of TANF income are the ones providing the highest number of work hours.

Source: Calculations based on data from Current Population Survey's annual March supplements.

Prierezová regresná analýza

- Vzťah medzi dvoma premennými, ktorý aproximujeme regresnou priamku nemusí byť kauzálny. Preto výsledky regresie neinterpretujeme slovami “x-percentný pokles/nárast premennej A spôsobuje y-percentný pokles/nárast premennej B”, ale “x-percentný pokles/nárast premennej A je spojený s y-percentný pokles/nárast premennej B”
- Regresná analýza má oproti korelácii výhodu v tom, že dokáže pracovať so skreslením, vďaka možnosti zahrnúť do nej kontrolné premenné. Kontrolné premenné vnesú do analýzy iné ako intervenčné rozdiely medzi jedincami vo vzorke, a tým pádom zostávajúcu koreláciu medzi A a B je možné o trochu spoľahlivejšie označiť za kauzálnosť.
- Použitie kontrolných premenných však nikdy nevyrieši problém skreslenia úplne, keďže väčšinou nemáme k dispozícii všetky charakteristiky, ktoré by mohli závislé premenné ovplyvňovať.

Prierezová regresná analýza

- $y = \alpha + \beta x + e$
- kde
 - α = konštanta (hodnota pre $x = 0$)
 - β = koeficient sklonu, vyjadrujúci zmenu y pri jednotkovej zmene x
 - e = odchýlka, ktorá vyjadruje rozdiel medzi skutočnou hodnotou premennej a jej hodnotou odhadovanou modelom

Table XXX: Linear regression analysis (Ordinary least squares regression)

Dependent variable	(1)	(2)	(3)	(4)	(5)
	Total	Charity	Subjects	U-supply	O-supply
Part B – Part A (Simulations)	-1.57	-2.20	0.63	-0.02	-0.09
Wait	-56.14*** (4.73)	-67.13*** (4.70)	10.99*** (2.12)	0.45*** (0.03)	-1.65*** (0.18)
Demand	4.12 (2.70)	0.77 (3.63)	3.36 (2.02)	-0.01 (0.03)	-0.40** (0.18)
Supply	7.11*** (2.52)	0.19 (3.24)	6.92*** (1.86)	-0.00 (0.02)	-0.75*** (0.17)
Wait * Demand	-4.10 (6.61)	-0.78 (6.60)	-3.31 (3.07)	0.01 (0.05)	0.40 (0.27)
Wait * Supply	7.73 (5.60)	13.57** (5.61)	-5.84* (2.96)	-0.09** (0.04)	0.54** (0.26)
Subjects	-0.14 (3.36)	-0.58 (5.06)	0.44 (2.70)	0.00 (0.03)	0.06 (0.23)
Wait * Subjects	-44.02*** (7.54)	-44.91*** (8.24)	0.89 (3.66)	0.30*** (0.06)	0.65* (0.35)
Demand * Subjects	-1.44 (4.92)	1.09 (6.49)	-2.53 (3.28)	-0.01 (0.04)	0.21 (0.29)
Supply * Subjects	-12.62* (6.60)	-9.80 (7.43)	-2.82 (3.71)	0.07 (0.05)	0.24 (0.33)
Wait * Demand * Subjects	1.83 (10.52)	-3.85 (10.72)	5.68 (5.19)	0.03 (0.07)	-0.42 (0.49)
Wait * Supply * Subjects	24.50** (11.84)	21.38 (13.53)	3.12 (5.74)	-0.14 (0.09)	-0.26 (0.51)
Constant	20.44*** (3.44)	21.78*** (3.92)	-1.33 (1.60)	-0.10*** (0.03)	0.03 (0.15)
R²	0.96	0.96	0.66	0.96	0.77
N	72	72	72	72	72
Seed dummies	Yes	Yes	Yes	Yes	Yes
p-values					
D = S	0.20	0.86	0.08	0.87	0.06
W*D = W*S	0.04	0.01	0.41	0.02	0.58
D*S = S*S	0.10	0.12	0.92	0.12	0.90
W*D*S = W*S*S	0.06	0.05	0.66	0.05	0.75

Note: Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

Kvázi-experimenty

- Ako už bolo povedané, pri výskume nemôžeme pri každej otázke použiť RCT. Diskutovali sme o alternatívach k RCT (časové rady, regresná analýza), ale zároveň sme uvideli, že tieto metódy majú nedostatky s odstraňovaním problému skreslenia.
- Je teda nejaká ďalšia možnosť ako vyšetriť kauzálne vplyvy aj bez použitia RCT?
- Empirický výskum sa v poslednom období začal zameriavať na riešenie, ktoré predstavuje akúsi strednú cestu: kvázi-experimenty. Ide o situácie, kedy zmeny v ekonomických podmienkach (napr. zmena zákona) prirodzene vytvoria takmer identické intervenčné a kontrolné skupiny. Pri kvázi-experimentoch je teda náhodné rozdelenie vytvorené namiesto experimentátora externými vplyvmi.
- Ani pri kvázi-experimentálnych štúdiách si nikdy nemôžeme byť istí, že z porovnania intervenčnej a kontrolnej skupiny sú vytlačené všetky skreslenia. Existujú však dva silné argumenty, prečo o kauzálnom efekte hovoriť môžeme. Prvý z nich je intuitívny - keďže intervenčná a kontrolná skupina boli pred zmenou veľmi podobné, je pravdepodobné, že skreslenia by mali byť minimálne. Druhý je štatistický - stále môžeme pre vyšetrenie skreslenia použiť alternatívne kontrolné skupiny.

Kvázi-experimenty

- Metóda rozdielu v rozdieloch (difference-in-differences) — technika spájajúca časové rady a prierezovú analýzu, čím odstraňuje nedostatky oboch z nich. Porovnáваме zmenu v populácii A so zmenou v populácii B, pričom kontrolujeme faktory času, ktorý skresľuje analýzu časových radov v rámci populácie A. Tiež porovnáваме zmenu v rámci každej populácie, namiesto porovnania dvoch populácií v jednom bode, čím kontrolujeme skreslenia prierezovej analýzy dvoch populácií.
- Vyšetrenie zmeny premennej X
 - 2 obdobia (Y,Z)
 - 2 populácie (A,B)
 - v období Y je politika rovnaká pre A aj B
 - v období Z je nová politika pre A, zatiaľ čo politika pre B sa nemení
- $x(\text{populácia A, rok Y}) - x(\text{populácia A, rok Z}) = \text{Efekt intervencie} + \text{Skreslenie}$
- $x(\text{populácia B, rok Y}) - x(\text{populácia B, rok Z}) = \text{Skreslenie}$
- Rozdiel = Efekt intervencie

Kváci-experimenty

■ TABLE 3-1

Using Quasi-Experimental Variation

Arkansas

	1996	1998	Difference
Benefit guarantee	\$5,000	\$4,000	-\$1,000
Hours of work per year	1,000	1,200	200

Louisiana

	1996	1998	Difference
Benefit guarantee	\$5,000	\$5,000	\$0
Hours of work per year	1,050	1,100	50

In Arkansas, there is a cut in the TANF guarantee between 1996 and 1998 and a corresponding rise in labor supply, so if everything is the same for single mothers in both years, this is a causal effect. If everything is not the same, we can perhaps use the experience of a neighboring state that did not decrease its benefits, Louisiana, to capture any bias to the estimates.

Obmedzenia experimentálnych pozorovaní

- Hoci RCT a kvázi-experimenty dokážu pomerne spoľahlivo odhaliť kauzálnosť, majú dve dôležité obmedzenia. Prvým je, že odhadujú iba kauzálny dopad konkrétnej intervencie. Povedzme, že sme zistili, že zníženie premennej A o 15 percent zvyšuje premennú B o 5 percent. To nám síce hovorí o efekte zníženia A o 15 percent, ale nehovorí to nič o tom, čo by sa stalo, ak by sme znížili A o napríklad 30 percent, alebo zvýšili o 10. Inými slovami, z pozorovania jednej zmeny v prostredí nevieme extrapolovať predpovede o iných zmenách v tom istom prostredí. Máme teda presnú odpoveď na jednu špecifickú otázku, ale nie všeobecný záver o tom, aké efekty by priniesli iné intenzity zmeny A.
- Druhým obmedzením je, že experimenty nám síce prezradia, čo intervencia zmení a ako, ale často nám nevedia povedať prečo - teda akým konkrétnym mechanizmom. Často by sme chceli poznať štrukturálne odhady, teda čo presne sa zmenilo v rozhodnutiach jednotlivcov (napríklad aby sme vedeli rozdeliť substitučný a dôchodkový efekt), ale experimenty nám dokážu poskytnúť iba tzv. redukované odhady.
- Redukované odhady ukazujú vplyv jednej zmeny na celkovú odozvu v správaní. Druhá nevýhoda experimentov je teda spojená s prvou - ak by sme dokázali rozumieť základnej štruktúre rozhodovania, potom by sme vedeli povedať viac o tom, ako budú ľudia reagovať na rôzne typy tej istej intervencie.

Štrukturálne modelovanie

- Problémy predostreté na predchádzajúcom slide viedli k vzniku štrukturálneho odhadovania. Touto metódou sa empirickí ekonómovia snažia odhadnúť nielen redukované odozvy na zmeny v prostredí, ale celé funkcie úžitku a to tak, že štiepia celkový efekt na viacero menších efektov.
- Štrukturálne modely sú potenciálne veľmi užitočným komplementom k experimentálnym analýzám, tieto modely sa však zostavujú oveľa problematickejšie ako redukované, pretože hoci používajú rovnaký objem dát, snažia sa z nich rozpoznať viac informácií.

Kvantitatívna podpora kauzálnosti

- Model nemusí hovoriť pravdu, musí však byť užitočný
- Čím viac z nasledujúceho je splnené, tým spoľahlivejšie môžeme hovoriť o kauzálnosti
 - Efekt je výrazný. Nie, že je iba štatisticky významný ale že je aj veľký z praktického hľadiska.
 - Efekt je konzistentný. Podobný efekt namerali pre iné subjekty, v inom štáte za iných podmienok.
 - Efekt je špecifický. Skúmaný faktor ovplyvňuje len skúmanú odozvu y a nie 10 ďalších vecí.
 - Efekt rešpektuje časovú následnosť. Ak malo x spôsobiť y , tak x muselo nastať skôr.
 - Efekt je monotónny. Viacej x spôsobuje výraznejšiu zmenu v y .
 - Efekt je hodnoverný. Experti v danej oblasti poznajú mechanizmus pomocou ktorého mohlo x spôsobiť y a tento mechanizmus je hodnoverný.
 - Efekt je potvrdený experimentom.

I USED TO THINK
CORRELATION IMPLIED
CAUSATION.



THEN I TOOK A
STATISTICS CLASS.
NOW I DON'T.



SOUNDS LIKE THE
CLASS HELPED.



WELL, MAYBE.

Viac informácií

- Pozrite sa na tento text, je napísaný veľmi pekne a nie príliš zoširoka, ani príliš technicky.
- <https://towardsdatascience.com/everything-you-need-to-know-about-hypothesis-testing-part-i-4de9abebbc8a>