

Sonja Drobnič

DINAMIČNA SOCIOLOŠKA ANALIZA: METODE IN PODATKI

Prehod od statičnega k dinamičnemu pristopu v sociologiji zahteva nove metode, s katerimi je možno zajeti dinamiko družbenih procesov. Ena od najobetavnejših metod je analiza zgodovine dogodkov, ki v raziskovanje eksplicitno vnaša novo dimenzijo - čas. Članek predstavi osnovna načela analize zgodovine dogodkov (AZD), naniza nekaj najpogostejših statističnih modelov, ki se uporabljajo v AZD, predstavi metodo zbiranja podatkov ob pomoči življenjepisnega koledarja ter na praktičnih primerih prikaže, kako pripravimo podatke in jih oblikujemo v datoteko, primerno za dinamično analizo.

Transition from static to dynamic approach in sociology requires new methods which are able to deal with the dynamics of social processes. One of the most promising methods is Event History Analysis which explicitly takes into account time dimension. In this paper, basic principles of the event history analysis will be introduced, and some of the most commonly used statistical models will be presented. It will be suggested that the Life History Calendar is an excellent tool for collecting data on life histories. Practical examples will be included to illustrate how to manage the data, and create the event oriented data file.

analiza zgodovine dogodkov, metodologija empiričnega raziskovanja, kvantitativne metode, statistični modeli

1. UVOD

Sociološko razumevanje sveta in družbenih pojavov niha med dvema skrajnostima. Posamezne teorije v zgodovini sociološke misli poudarjajo predvsem red, stabilnost in ekvilibrium; v okviru teh teorij je vsaka sprememba problematična, saj pomeni nevarnost za družbeni red, zato ji sledi takojšen poskus vzpostavljanja ponovnega ravnotežja na novi ravni. Druga skrajnost so teorije, ki razumejo družbo kot konstanten proces sprememb, prevratov, kolektivnega nasilja in revolucij. Večino socioloških pristopov je možno uvrstiti nekam na sredino med obema skrajnostima; to, ali so posamezni sociologi bližje enemu ali drugemu ekstremu, je navadno odvisno od problematike, s katero se ukvarjajo, ter teoretskega okvira, ki ga uporabljajo. Kot ugotavlja Bottomore:

"Resnični problem je oblikovati koncept družbene strukture, ki upošteva elemente regularnosti in reda v družbenem življenju ter hkrati ne zanemara toka zgodovinskega delovanja posameznikov in družbenih skupin, ki vzdržuje, spreminja, ali pa razbija ta red" (1975:160).

K Bottomorjevemu razmišljanju bi lahko dodali, da je poleg oblikovanja teoretičnih konceptov ravno tako ali še bolj problematično vprašanje empiričnega proučevanja družbenih pojavov. Za sodobne sociološke teorije lahko rečemo, da se čedalje bolj ukvarjajo z družbenimi spremembami - tako s tistimi, ki vodijo do radikalnih družbenih prevratov, kot tudi s spremembami in dogodki, ki sestavljajo vsakodnevno družbeno

dinamiko. Empirično družboslovno raziskovanje pa le počasi sledi temu razvoju. Večina socioloških metod je prirejenih za statično analizo in razvoj na tem področju je veliko počasnejši kot na teoretskem. Empirične sociološke raziskave si še vedno zastavljajo predvsem vprašanja o statičnih družbenih odnosih oziroma povezavah med posameznimi dejavniki v eni časovni točki. Prevladujoč je cross-section raziskovalni dizajn, torej enkratni posnetek stanja v določenem trenutku; dinamične družboslovne analize so šele v povojih.

Eden od pomembnih razlogov za ta zaostanek so gotovo težave s podatki; podatkovne zahteve so za dinamično analizo neprimerno višje, primerne podatkovne baze pa redke. Vendar Tuma in Hannan (1984) ugotavljata, da pomanjkanje podatkov lahko le delno upravičuje metodološko zaostajanje pri proučevanju dinamičnih procesov. Tudi kadar so podatki na voljo v obliki časovnih vrst ali panela, in to ni redko, se njihova časovno-dinamična struktura pogosto ignorira. Podatki se tretirajo kot cross-section, z nekaj dodatnimi metodološkimi komplikacijami, povezanimi z avtokorelacijo. Poudarek je navadno na merjenju spremembe med dvema ravnotežnima stanjema, na primer spremenjene vrednosti določene variable v dveh zaporednih panelnih valovih, ne pa na proučevanju dinamičnega procesa, ki je privedel do te spremenjene vrednosti.

Za analizo dinamičnih pojavov so potrebni novi prijemi in nove metode. Študij procesa namesto študija stanja oziroma spremembe stanja zahteva eksplicitno vključitev nove dimenzije - časa. Medtem ko je statično analizo možno izvajati na podatkih, ki se nanašajo na eno časovno točko, zahteva dinamična analiza informacije o istih enotah opazovanja skozi čas.

V tem prispevku se bomo v glavnem omejili na eno najobetavnejših metod za proučevanje dinamičnih družbenih procesov: "analizo zgodovine dogodkov" (Event History Analysis). Namen članka je v osnovnih obrisih prikazati načela analize zgodovine dogodkov, predstaviti nekaj najpogostejše uporabljenih modelov ter opozoriti na probleme, na katere je potrebno biti pozoren pri obdelavi podatkov. Nato sledi prikaz podatkovne osnove, ki jo taka analiza zahteva, ter predstavitev "življenjepisnega koledarja", ki se je izkazal kot nepogrešljiv praktični pripomoček pri zbiranju podatkov.

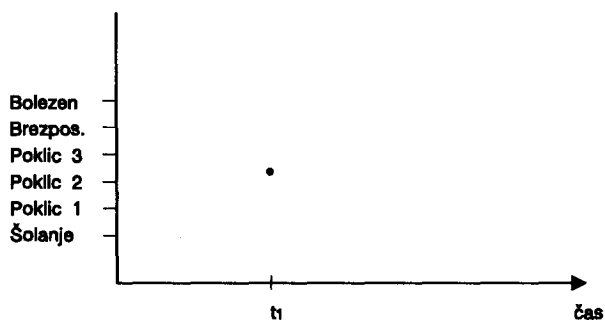
Da bi zadostili podatkovnim potrebam za izvajanje analize zgodovine dogodkov, so nujno potrebne spremembe v sistemu zbiranja podatkov. Vendar je potrebno poudariti, da te spremembe v nobenem primeru ne onemogočajo "klasične" podatkovne obdelave. Z nezahtevnim logičnim premislekom in tehnično preprostim "rekodiranjem" je namreč možno podatke, zbrane v obliki življenjepisnega koledarja, prevesti v običajno cross-section obliko.

2. OSNOVE ANALIZE ZGODOVINE DOGODKOV ¹

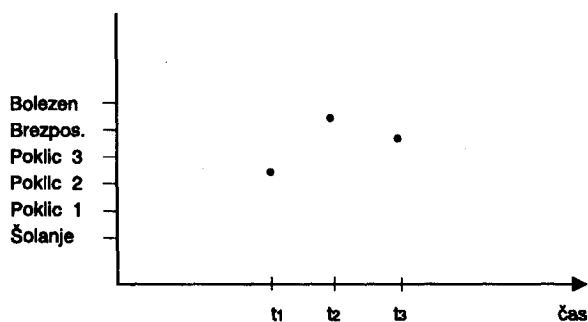
2.1. Prednosti AZD raziskovalnega dizajna

V družbenih znanostih raste zanimanje za analizo zgodovine dogodkov (AZD) izredno hitro, saj zgodovina dogodkov navadno bolje odseva dinamično naravo empiričnega dogajanja kot pa tradicionalni podatki iz cross-section študij ali panelni podatki ali časovne serije. Da bi videli, kaj novega prinaša AZD, si za začetek ogledimo skico treh različnih raziskovalnih dizajnov: anketiranje v eni časovni točki ali cross-section dizajn, panelna raziskava in raziskava zgodovine dogodkov (graf 1).

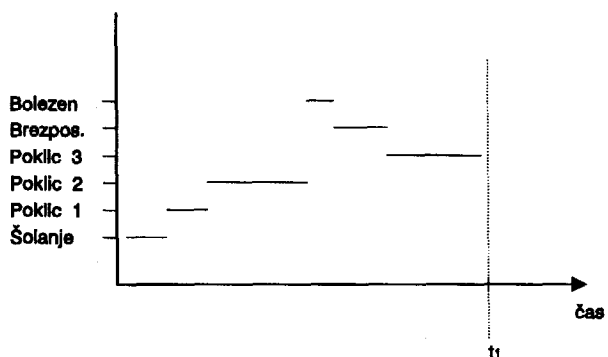
Cross-section dizajn



Panelni dizajn



Dizajn zgodovine dogodkov



Graf 1

Proučevanje zaposlitvene kariere posameznika na osnovi cross-section vzorca, panelne raziskave in dizajna zgodovine dogodkov (prirejeno po Blossfeldu, Hamerleju in Mayerju 1989).

Skica prikazuje zaposlitveno kariero posameznika, kot jo zajamemo z različnimi raziskovalnimi dizajni. S cross-section anketiranjem je celotna izobraževalna in zaposlitvena kariera posameznika predstavljena z informacijo v eni časovni točki; zvedemo le, kakšno je stanje v času anketiranja. Panelna raziskava zajame več informacij, saj posameznikov zaposlitveni položaj opazujemo v treh časovnih točkah. Vendar iz podatkov ni jasno, ali so se pred opazovanji ali med njimi zgodile kakšne spremembe, niti ne, kako dolge so posamezne epizode v zaposlitveni karieri. Šele raziskovalni dizajn, ki zajame zgodovino dogodkov, registrira vse spremembe stanja in natančen čas teh sprememb. Tak raziskovalni dizajn daje pregled nad celotnim potekom dogodkov in omogoča podrobno rekonstrukcijo zaposlitvene kariere v katerikoli časovni točki.

Razlike med temi tremi raziskovalnimi dizajni niso samo količinske, se pravi, da vsak naslednji dizajn zagotovi večjo količino informacij. Kar je pomembno, so predvsem kakovostne razlike. Anketiranje v eni časovni točki temelji na domnevi, da je proces, ki ga proučujemo, sorazmerno stabilen skozi čas. Če je ta predpostavka v realnosti izpolnjena, je cross-section raziskovanje neproblematično. Če pa prihaja v realnosti do fluktuacij in sprememb, "posnetek" stanja v eni časovni točki ne daje zadovoljive slike dogajanja, saj bodo rezultati odvisni od specifičnih razmer v času anketiranja. Pa tudi če je empirično dogajanje stabilno skozi čas, sta panelni dizajn in zgodovina dogodkov bolj informativna, saj z njima lahko stabilnost empirično demonstriramo in potrdimo, ne pa le domnevamo.

Panelna raziskava generira več informacij kot cross-section dizajn in že omogoča analizo sprememb, vendar je popolnoma ustrezna le takrat, kadar se spremembe dogajajo v jasno definiranih časovnih točkah, ki se pokrivajo z intervali anketiranja oziroma registriranja podatkov.² Panelni podatki pa ne omogočajo rekonstrukcije kontinuiranega procesa, če se spremembe lahko zgodijo v katerikoli časovni točki. Adekvaten študij sprememb v kvalitativnih spremenljivkah, ki se lahko zgodijo v poljubni časovni točki, omogoča le raziskovalni dizajn, ki rekonstruira zgodovino dogodkov. AZD je še posebej nepogrešljiva, če želimo analizirati kompleksne, med seboj povezane procese, pri katerih je zaradi vzporednega poteka in feedback učinkov težko odkriti vzročne povezave.

2.2. Dogodki

Dogodki so temelj večine družboslovnih raziskav, tudi če zaradi statične narave empiričnega raziskovanja raziskovalci tega ne izražajo eksplicitno. V medicinskih in epidemioloških študijah, v katerih ima analiza zgodovine dogodkov daljšo tradicijo kot v družboslovju, se koncept "zdravstveno stanje posameznika" operacionalizira kot zaporedje dogodkov, ki opisujejo prehod iz enega stanja v drugo, na primer obolelost, hospitalizacija, operacija, ozdravitev, smrt. Delovno kariero lahko opišemo kot niz dogodkov: vstop v prvo službo, napredovanje, izguba zaposlitve, nova zaposlitev, upokožitev. Tudi če je naše raziskovanje osredotočeno le na eno fazo, na primer stanje brezposelnosti, je implicitno jasno, da to stanje omejujeta dva dogodka: začetek iskanja dela (kot posledica izgube prejšnje zaposlitve ali vstopa na trg delovne sile), ter konec tega stanja (nova zaposlitev, izstop s trga delovne sile).

Poznavanje zaporedja dogodkov in časovnih točk, v katerih se dogodki "zgodijo", vnaša v analizo novo dimenzijo - čas. V našem primeru zaposlitvene kariere nam te informacije na primer omogočajo razločevati med kratkotrajno in dolgotrajno brezposelnostjo, ter dajejo vpogled v zaposlitveno dinamiko (morebitna ponavljajoča se

brezposelnosti); to ostane pri cross-section analizi prikrito. Te vrste informacij so izjemno pomembne predvsem, če služijo kot osnova za oblikovanje socialne politike. Če imamo na primer dve državi ali dve regiji in je v obeh primerih stopnja brezposelnosti 10%, vendar je v prvem brezposelnost "enakomerno" porazdeljena med prebivalstvom, tako da je večina ljudi med delovno kariero za krajši čas brezposelna, v drugem pa je koncentrirana med manjšim delom marginalnega prebivalstva, pri katerem se med celotno delovno kariero epizode zaposlenosti izmenjujejo z obdobji brezposelnosti, bo to zahtevalo radikalno drugače zasnovane zaposlitvene in socialne programe. Osnovo za oblikovanje socialnih programov nam da šele poznavanje dinamike brezposelnosti v družbi in poznavanje dinamike tipičnih zaposlitvenih karier; samo agregatni podatki v določeni časovni točki ne zadostujejo za te namene. Poleg tega nam poznavanje zgodovine dogodkov omogoča, da vključimo v analizo tudi časovno-variabilne neodvisne spremenljivke; o tem bo več govora pozneje.

Za ponazoritev omenimo še nekaj področij, na katerih se za proučevanje dogodkov in vzrokov zanje čedalje pogosteje uporablja analiza zgodovine dogodkov. Demografija je klasičen primer; demografi se tradicionalno zanimajo za rojstva, smrti, poroke, ločitve, migracije. Kriminologi se ukvarjajo s prekrški, kriminalnimi dejanji, aretacijami, sodbami, odpustitvami iz zapore, povratništvom. Politologe zanimajo družbena gibanja, demonstracije, revolucije, vojne, spremembe političnih režimov. V vsakem od teh primerov sestavlja dogodek kvalitativna sprememba, ki se zgodi v določeni časovni točki. Pojma "dogodek" ne uporabljamo za opis postopnega spreminjanja neke kvantitativne spremenljivke. Sprememba mora odsevati sorazmerno ostro disjunkcijo med "poprejšnjim" in "novim" stanjem.

Morda je potrebno poudariti, da so metode analize zgodovine dogodkov uporabne za različno koncipirane družbene procese. Včasih so predmet raziskovanja ponavljajoči se dogodki, kot na primer stavke ali demonstracije, ali, kot praktičen in nazoren primer, ponavljajoče se avtomobilske nesreče na določenem odseku avtoceste. V takem primeru je trajanje dogodka samega zanemarljivo glede na dolžino čakalnega obdobja (waiting time) med dogodki; običajno koncipiramo dogajanje, kot da oseba ali sistem, ki ga proučujemo, "doživlja" dogodke. Število dogodkov je sorazmerno visoko, rezultat tega pa je, da nobeno posamezno stanje (recimo stanje, ko sistem doživi natančno 12 stavk ali ko se zgodi natančno 20 prometnih nesreč) ni substancialno zanimivo. Kar nas v tem primeru zanima, je vsesplošna stopnja pojavljanja dogodkov, ne pa posamezen prehod stanja N v stanje $N+1$.

V drugem primeru pa nas zanima predvsem prehod med dvema stanjema oziroma sprememba med strukturno različnimi stanji, na primer sprememba zakonskega statusa (neporočen, poročen, ločen), sprememba v družinskem statusu (nima otrok, ima otroke) ali sprememba v zaposlitvenem statusu (zaposlen, brezposeln, upokojen). V tem primeru - prehod med dvema stanjema - je trajanje vsakega posameznega stanja netrivialno. Proces, ki nas zanima, je ravno sprememba stanja oziroma prehod iz ene oblike v drugo. Število stanj je omejeno in vsako stanje je substancialno zanimivo, zato v analizi modeliramo posamezne stopnje prehoda med dvema stanjema.

Naj bo proces, ki ga proučujemo, prehod med posameznim stanji ali ponavljajoči se dogodki, oba primera lahko predstavimo kot stohastični proces.

2.3. Osnovni statistični koncepti

Osnovni model analize zgodovine dogodkov proučuje dolžino časovnih intervalov med zaporednimi kvalitativnimi spremembami stanja v opazovanem obdobju. Z drugimi besedami, AZD proučuje časovni razmik med dogodki. Časovno obdobje med zaporednimi dogodki imenujemo **epizoda** (spell, episode). Časovni interval od dogodka N-1 do dogodka N imenujemo **čakalno obdobje** (waiting time).

Najpreprostejši primer AZD je merjenje prehoda z začetnega v končno stanje; tak preprost primer so tablice smrtnosti. Vendar je mnogo konceptov, razvitih za enoepizodni primer, možno aplicirati na bolj kompleksne situacije, za katere so značilni ponavljajoči se dogodki in/ali kompetitivna tveganja.³ Trajanje ali dolžino epizode popisuje v statističnem modelu nenegativna stohastična spremenljivka T. Če je čas, ko se dogodki pojavljajo, merjen v vsaki časovni točki (zvezno), je količina T zvezna stohastična spremenljivka.

Porazdelitev verjetnosti trajanja epizode lahko specificiramo kot porazdelitveno funkcijo

$$F(t) = P(T < t)$$

ki specificira verjetnost, da je slučajna spremenljivka T manjša od neke količine t. Odgovarajoča gostota porazdelitve verjetnosti je

$$f(t) = dF/dt$$

Eden od zelo uporabnih načinov za predstavitev sklopa zgodovin dogodkov je **funkcija preživetja** (survivor function)

$$S(t) = P(T \geq t)$$

kjer je T slučajna spremenljivka, ki popisuje dolžino čakalnega obdobja v določenem stanju. Funkcija preživetja izraža verjetnost, da posameznik ostane v določenem stanju (preživi) do časa t, oziroma, da se dogodek še ni zgodil in se epizoda nadaljuje.

Funkcija preživetja S(t) je eden od alternativnih načinov za specifikacijo porazdelitve slučajne količine T. Njeno razmerje do porazdelitvene funkcije F(t) je: $S(t) = 1 - F(t)$

Posebej uporabna pri analizi dolžine trajanja časovnih epizod je **stopnja tveganja** h(t) (hazard rate⁴), ki jo lahko interpretiramo kot neopazovano stopnjo (oziroma hitrost) pojavljanja dogodkov,

$$h(t) = f(t)/S(t)$$

oziroma precizna definicija

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq T < (t + \Delta t) \mid t \leq T)$$

Formalno interpretiramo stopnjo tveganja kot verjetnost, da se epizode končajo v intervalu [t, t+Δt]- pod pogojem, da se dogodek ni zgodil pred začetkom intervala oziroma da je posameznik ali sistem v tem intervalu še podvržen tveganju.⁵

Že iz te skraćene predstavitve je razvidno, da lahko epizode opišemo z različnimi funkcijami. Če poznamo eno od njih, lahko iz nje izpeljemo druge. Izberemo tisto specifikacijo, ki je zaradi ugodnosti pri ocenjevanju in/ali lažje interpretacije primernejša.

2.4. Cenzurirani podatki

Eden najbolj perečih problemov pri analizi zgodovine dogodkov so **cenzurirani** podatki (censored data). Kadar imamo na voljo podatke za obdobje med t_1 in t_2 , vendar zavzema kvalitativna spremenljivka $y(t)$ vrednosti tudi v času pred t_1 ali po t_2 , pa teh vrednosti ne poznamo, pravimo, da so podatki o zgodovini dogodkov cenzurirani. Če nam ni znano dogajanje za $t < t_1$, so podatki levo cenzurirani; če nam manjkajo informacije za $t > t_2$, so desno cenzurirani. Podatkovne baze so skoraj vedno desno cenzurirane, saj je zaključek opazovanega obdobja, ki se navadno ujema s časom zbiranja podatkov, običajno povsem arbitraren. Pogosto so podatki tudi levo cenzurirani, če začetek opazovanega obdobja pade v sredino epizode.

Poglejmo si problem cenzuriranih podatkov na primeru študija dolžine brezposelnosti. Recimo, da smo podatke v obliki zgodovine dogodkov rekonstruirali za 5 let, in sicer za obdobje od junija 1987 do junija 1992, ko je potekalo zbiranje podatkov na terenu. V tem petletnem časovnem obdobju so nekateri delavci postali brezposelni, vendar so tudi že našli drugo zaposlitev - lahko tudi po večkrat. Če se je epizoda brezposelnosti začela in končala znotraj opazovanega obdobja in torej poznamo datum začetka in konca brezposelnosti, je epizoda necenzurirana in podatki niso problematični. Vendar je zelo verjetno, da bomo med anketiranimi osebami v raziskavi našli tudi na primere, ko so bili ljudje junija 1987, ko se je naše opazovanje začelo, že brezposelni. Datum začetka opazovanja torej ni datum začetka epizode; zgodovina je neznana na levi strani, epizode so levo cenzurirane. Druga možnost je, da osebe, ki so izgubile zaposlitev v obdobju opazovanja, še niso našle zaposlitve v času anketiranja. Zanje poznamo datum začetka epizode, ne pa datuma konca brezposelnosti. Epizode so desno cenzurirane. In še kombiniran primer: oseba, ki je postala brezposelna pred začetkom opazovanja ter po petih letih v času terenske raziskave še ni našla zaposlitve, generira eno samo epizodo, ki je hkrati levo in desno cenzurirana.

Bolj uspešni kot pri reševanju problema levo cenzuriranih podatkov smo pri desno cenzuriranih. Prvotna strategija, ki so jo uporabljali raziskovalci, je bila izključitev desno cenzuriranih epizod iz analize. Vendar se je pokazalo, da taka izključitev lahko privede do pristranskih ocen. Danes se za ocenjevanje modelov uporablja metode, ki upoštevajo cenzurirane podatke pod predpostavko, da velja enak stohastični model za vse epizode, tako za necenzurirana kot za cenzurirana opazovanja. Taka metoda je metoda maksimalne verjetnosti (maximum likelihood method ali ML). Da bi jo lahko uporabili, potrebujemo pri analizi nekaj dodatnih informacij. Oblikovati moramo posebno slamnato spremenljivko d , tako imenovani cenzurni indikator, ki jo priključimo k vsaki epizodi; ta spremenljivka pove ali je epizoda desno cenzurirana ali ne.

$$d_i = \begin{cases} 1 & \text{če } t_i \text{ ni cenzuriran} \\ 0 & \text{če je } t_i \text{ cenzuriran} \end{cases} \quad (i = 1, \dots, n)$$

V primeru levo cenzuriranih podatkov - kadar manjkajo informacije za čas pred t_1 - pa žal ni mogoče izpeljati verjetnosti poteka pretekle zgodovine iz modela, kot je to možno pri desno cenzuriranih podatkih. Tu si pomagamo na ta način, da poskušamo rekonstruirati čimveč fragmentov iz obdobja pred t_1 ter te drobce informacij uporabiti pri ocenjevanju modelov. Če se vrnemo k primeru študija dolžine brezposelnosti, tu za osebe, ki so že brezposelne ob začetku opazovanega obdobja (junij 1987), poizvemo, kdaj se je ta epizoda brezposelnosti začela, ter to informacijo kot kovariato uporabimo pri ocenjevanju parametrov.⁶

3. MODELI IN OCENJEVANJE

3.1. Modeliranje

V tem poglavju predstavljamo nekaj napogostej uporabljanih modelov, ki opisujejo porazdelitev čakalnega obdobja med dogodki. Ker se v večini aplikacij analize zgodovine dogodkov v družboslovju uporabljajo zvezni modeli, se bomo omejili nanje. Omenimo le, da dajejo tudi metode, ki se uporabljajo, kadar je čas merjen v diskretnih enotah, podobne rezultate. Vendar so za družbene in ekonomske znanosti konceptualno primernejši zvezni modeli, saj pri družbenih pojavih običajno ni nikakršnih "naravnih" časovnih točk zanje oziroma za dogodke; dogodki se lahko zgodijo poljubno na katerikoli točki časovnega intervala.

Analiza podatkov je pravzaprav iskanje ekonomičnega statističnega modela, ki se kar se le da približa realnosti ter omogoča vsebinsko interpretacijo. Osnovni namen analize zgodovine dogodkov je modelirati dolžino časa, ki preteče do pojava dogodka oziroma časa med posameznimi dogodki. Za opis trajanja čakalnega obdobja lahko služijo številne porazdelitvene funkcije. Med njimi izberemo tisto, ki najbolj ponazarja konkretno empirično dogajanje.

V kakšnem smislu je empirično dogajanje (pojavljanje dogodkov) odvisno od časa? Klasičen primer v teoriji verjetnosti je metanje kocke; pri tem je verjetnost, da se pojavi na primer število tri, v vsakem poskusu enaka. Tveganje, da bomo dobili trojko, ni odvisno od tega, koliko časa že mečemo kocko ali koliko časa je preteklo, odkar se je ta številka zadnjič pojavila. Če izračunamo stopnjo tveganja za pojavljanje števila tri, ta ni odvisna od časa, pač pa je konstantna skozi čas oziroma modificirana le s slučajno variabilnostjo.

V večini primerov v družboslovju pa je smiselno domnevati, da je stopnja tveganja funkcija časa in se zato spreminja skozi čas, bodisi v odvisnosti od preteka časa od zadnjega dogodka bodisi od starosti posameznika oziroma sistema, ki dogodke doživlja. Podatki, ki so na voljo, na primer kažejo, da po 25. letu starosti stopnja tveganja, da bo oseba aretirana zaradi kriminalne dejavnosti, konstantno upada (Allison 1984). Po drugi strani pa je očitno, da se stopnja tveganja upokojitve s starostjo povečuje. Kiefer (1988) navaja, da je zaključek zaposlitvene epizode negativno povezan z dolžino čakalnega obdobja, vsaj po preteku prvih nekaj mesecev po zaposlitvi. Čim dlje je torej oseba zaposlena na določenem delovnem mestu, tem manjše je tveganje, da bo izgubila zaposlitev. Nasprotno pa je za kolektivna gibanja značilno, da se čas med posameznimi protesti ali demonstracijami po prvotnem izbruhu skrajšuje; kolektivne akcije se pojavljajo v ciklih ali valovih, kot da bi bil proces "nalezljiv" (Olzak 1989). V sklopu takega cikla se čas med sekvenčnimi dogodki skrajšuje.

Včasih težko govorimo o monotonem povečevanju ali zmanjševanju tveganja skozi čas, ker se "smer" spreminja. Pač pa lahko tveganje ponazorimo s funkcijo, ki ima obliko črke U. Tak primer je stopnja smrtnosti. Ta je sporazmerno visoka takoj po rojstvu, nato močno upade ter se začne spet povečevati v poznih srednjih letih.

Iz teh primerov je razvidno, da je oblika funkcije, ki ponazarja stopnjo tveganja $h(t)$, ena najpomembnejših značilnosti različnih zveznih modelov, s katerimi opisujemo družbene procese. Stopnja tveganja popolnoma determinira porazdelitev verjetnosti trajanja časa do dogodka oziroma trajanja epizode med dogodki, kadar gre za ponavljajoče se dogodke.

3.2. Pogosto uporabljeni parametrični modeli

EkspONENTNA funkcija se izredno pogosto uporablja kot model za podatke, ki popisujejo časovne epizode. Z eksponentnim modelom je lahko ravnati in ga je lahko interpretirati. Primeren je predvsem za analizo epizod, ki ne kažejo velike variabilnosti.⁷

Stopnja tveganja, funkcija preživetja in funkcija gostote za eksponentno porazdelitev s parametrom $r > 0$ so:

$$\begin{aligned}h(t) &= r \\S(t) &= \exp(-rt) \\f(t) &= r \exp(-rt)\end{aligned}$$

Stopnja tveganja v eksponentnem modelu je konstanta; torej ni odvisna od časa oziroma pretekle zgodovine. Zato eksponentno porazdelitev imenujemo tudi "funkcija brez spomina". Ker je ta funkcija zares preprosta, se pogosto uporablja kot prva aproksimacija pri iskanju boljšega modela.

Iz eksponentne funkcije je tudi zelo plastično razvidno razmerje med funkcijo tveganja in funkcijo preživetja. Pričakovana srednja vrednost dolžine epizode je $E(T) = 1/r$. Čim večje je torej "tveganje", da se bo dogodek zgodil, tem krajša bo pričakovana dolžina čakalnega obdobja. In nasprotno, z zmanjševanjem tveganja se podaljšuje čakalno obdobje.

Ker pa je eksponentna porazdelitev odvisna od enega samega parametra r , ni zelo fleksibilna; zato tudi ni verjetno, da bo ustrezno opisovala podatke, če vzorec vsebuje hkrati zelo dolge in zelo kratke epizode.

Weibullova porazdelitev je že bolj fleksibilna in dobro opisuje nekatere empirične družbene pojave. Vsebuje dva parametra ($r > 0$, $a > 0$). Funkcije tveganja, preživetja in gostote so

$$\begin{aligned}h(t) &= ra (rt)^{a-1} \\S(t) &= \exp [-(rt)^a] \\f(t) &= ra (rt)^{a-1} \exp [-(rt)^a]\end{aligned}$$

Weibullov model kaže, da je stopnja tveganja eksponencialna funkcija čakalnega obdobja. Tveganje je monotono padajoča funkcija časa pri $a < 1$, rastoča pri $a > 1$ ter se zreducira na konstanto, če je $a = 1$. Eksponentna porazdelitev je torej poseben primer Weibullove porazdelitve, ko je parameter $a = 1$.

Gompertzov model je v načelu podoben Weibullovemu, le da stopnja tveganja raste oziroma pada linearno s časom in ne linearno z logaritmom časa, kot v Weibullovem modelu. Uporabo enega ali drugega navadno determinira software, ki je na voljo za ocenjevanje modelov. Običajno namreč teorija ali empirični podatki niso zmožni diskriminirati med obema modeloma.

Poglejmo si še primer nemonotonega tveganja. Log-logistična porazdelitev ima dva parametra, $r > 0$ in $a > 0$.

$$\begin{aligned}h(t) &= ra (rt)^{a-1} / 1 + (rt)^a \\S(t) &= 1 / 1 + (rt)^a \\f(t) &= ra (rt)^{a-1} / [1 + (rt)^a]^2\end{aligned}$$

Pri $a > 1$ tveganje skozi čas sprva raste, nato upade. Če je $0 < a \leq 1$, funkcija tveganja pada z dolžino epizode.

3.3. Ocenjevanje modelov

Kako med alternativnimi modeli izberemo "pravega"? Odločitev je odvisna od številnih dejavnikov, ki pa jih je težko nanizati v obliki proceduralnega recepta. Nedvomno je potrebno loviti ravnotežje med tem, kako lahko je problem rešiti matematično, kako model ustreza teoretičnim predpostavkam in, seveda, kako se sklada z empiričnimi dejstvi.

Ključna za odločitev pri parametričnih modelih je ugotovitev, kako je stopnja tveganja odvisna od časa. Torej moramo najprej razločevati med eksponentnim modelom, ki specificira konstantno stopnjo tveganja, ter vsemi drugimi porazdelitvami. Z matematičnega vidika in z vidika softwera, ki ga uporabljamo pri ocenjevanju modelov, je eksponentni model zelo privlačen. Zato ga pogosto uporabimo kot prvo aproksimacijo "pravega" modela; iskanje boljšega modela torej nadaljujemo s te izhodiščne točke. Pomagamo si tudi s teorijo, ki pogosto generira jasne hipoteze o tem, kako se tveganje spreminja skozi čas. In seveda z empiričnimi dejstvi, zajetimi v podatkih, ki jih obdelujemo. V tem primeru si navadno pomagamo z grafičnimi metodami, ki služijo presoji o tem, ali se dogodki skozi čas pojavljajo po eksponentni porazdelitvi ali ne. Pri tem je potrebno paziti tudi na vpliv drugih neodvisnih spremenljivk ali kovariat, ki lahko vplivajo na odvisno spremenljivko. Učinki teh drugih karakteristik - populacijska heterogenost - se namreč lahko v prikriti obliki kažejo kot časovni učinki. V takem primeru bodo grafične metode pokazale padajočo stopnjo tveganja tudi, če je stopnja tveganja v resnici konstantna skozi čas.⁸

Modeli, pri katerih stopnja tveganja ni konstantna skozi čas, so na primer Weibullovo model, Gomerzov model, model gamma, log-normalni in log-logistični model. Eden najučinkovitejših načinov za iskanje najboljšega modela izhaja iz dejstva, da je eksponentna porazdelitev poseben primer številnih drugih modelov, med drugim Weibullovega, Gomerzovega in gamma modela. Pravimo tudi, da eksponentni model "gnezdi" v teh drugih modelih, zato je možno rezultate primerjati in jih testirati. V postopku izbire ocenimo eksponentni model in enega od teh drugih modelov po metodi ML (maximum likelihood). Nato primerjamo, kako dobro se podatki prilegajo tem modelom. Prileganje testiramo s primerjanjem vrednosti za log-likelihood v obeh ocenah. Eksponentno porazdelitev zavrnemo kot neustrezno, če se vrednost za log-likelihood pri eksponentnem modelu signifikantno razlikuje od vrednosti pri alternativnem modelu.⁹

Če zavrnemo eksponentni model, se nato odločamo med monotonimi modeli, kjer tveganje vedno raste ali pada, ter nemonotonimi modeli. Med nemonotonimi porazdelitvami se veliko uporablja log-logistični in log-normalni model; pri obeh funkcija tveganja sprva raste in nato pade. Žal ni na voljo primernih parametričnih

modelov, ki bi predstavljali funkcijo tveganja v obliki črke U. Če se porazdelitev močno odmika od monotonosti, je bolje presedlati na semiparametrične modele oziroma Coxov model proporcionalnega tveganja (proportional hazards model) (Allison 1984).

3.4. Vključitev kovariat

Poleg časovne dimenzije nas pri raziskovanju navadno zanima tudi, kako različne kovariate oziroma eksplanatorni dejavniki vplivajo na družbene procese. V statistični analizi zato poskušamo ugotoviti, kakšen je kvantitativen vpliv teh spremenljivk na stopnjo tveganja. Primer: pri proučevanju brezposelnosti nas ne zanima le, ali se z daljšanjem brezposelnega staža verjetnost, da bo iščoč našel zaposlitev, povečuje ali zmanjšuje, ampak nas zanima tudi, kako na stopnjo tveganja vplivajo drugi dejavniki, kot na primer starost, spol, izobrazba, poklic, družinski status ali poprejšnja zaposlitvena kariera brezposelne osebe.

Te spremenljivke ali kovariate so, kot je razvidno v zgornjem primeru, lahko kvantitativne ali kvalitativne. V AZD tretiramo kvantitativne spremenljivke x_j na enak način kot v tradicionalni multipli regresijski analizi. Za kvalitativne spremenljivke pa oblikujemo slamnate variable, ki zavzamejo vrednost 1, če je določena karakteristika ali pojav navzoč, ter vrednost 0 ob odsotnosti te karakteristike ali pojava. Enako kot v tradicionalni regresijski analizi lahko v model vključimo in ocenjujemo tudi interakcijske učinke neodvisnih spremenljivk. Konkretno poteka vključitev kovariat v model tako, da vrednosti spremenljivk, merjene ob začetku epizode, priključimo epizodi. Če ena oseba v toku zgodovine, ki jo proučujemo, generira več epizod, seveda priključimo kovariate vsaki epizodi posebej.

V praksi se predpostavlja, enako kot v regresijski analizi, da je vpliv kovariat linearen v parametrih. Specifično za AZD je, da navadno ne predpostavljamo, da kovariate vplivajo neposredno na dolžino čakalnega obdobja, temveč na funkcijo časovne spremenljivke T , na primer $\ln T$. Preprosto povedano, odvisna spremenljivka y je v AZD regresijskem modelu $y = \ln T$.

Druga pomembna razlika v primerjavi z metodami, ki se uporabljajo v tradicionalnih cross-section študijah, je, da se v AZD modelih nekatere kovariate lahko skozi čas spreminjajo in so torej časovno odvisne. Za časovno odvisne variable, ki se spreminjajo enakomerno zvezno skozi čas, kot na primer starost, je dokaj preprosto izračunati vrednost spremenljivke ob začetku vsake epizode in jo nato priključiti ustrezni epizodi. Vendar je večina spremenljiv, s katerimi imamo opravka v družboslovju, diskretnih. Kako torej ugotoviti, na kakšen način diskretne spremembe, kot na primer selitev, ločitev zakona, rojstvo otroka, nenadna obolelost, izguba zaposlitve, sprememba političnega režima in podobno, vplivajo na stopnjo tveganja v procesu, ki ga proučujemo?

Te vrste časovno odvisnih spremenljivk lahko priključimo AZD epizodam na več načinov. Preprostejši način je, da izmerimo vrednosti kovariat na začetku vsake epizode ter jih priključimo epizodi. To pomeni, da velja na začetku epizode merjena vrednost spremenljivke prek celotne epizode, ne glede na to, če je medtem nastala kakšna sprememba ali ne. Natančnejši, toda zahtevnejši postopek je, da epizodo, med katero se je sprememba zgodila, razcepimo na dva ali več delov ter jih ocenjujemo ločeno. Več o vključevanju časovno odvisnih kovariat in cepljenju epizod bo govora v poglavju o strukturi podatkov za analizo zgodovine dogodkov.

4. PODATKOVNA STRUKTURA

4.1. Ureditev podatkov

Shranjevanje podatkov in priprava le-teh za statistično analizo sta ena najpomembnejših faz AZD. Struktura podatkov je veliko bolj kompleksna kot v primeru cross-section raziskovalnega dizajna, kjer so podatki praviloma shranjeni v obliki "četverokotnika" (po stolpcih spremenljivke, po vrsticah primeri); ko so taki cross-section podatki že enkrat vnešeni v računalnik in "očiščeni", so raziskovalcem vedno na voljo, že popolnoma pripravljeni za analizo.

Če gre za podatke o dogodkih, so zadeve bolj zapletene, saj je potrebno pri vsakem raziskovalnem problemu pretehtati, kakšna podatkovna struktura je potrebna za analizo. Možno je torej, da bo vsak raziskovalni problem zahteval specifično organizirane podatke in torej posebno datoteko, pripravljeno na osnovi prvotne, originalne baze podatkov. Včasih zahteva že majhna modifikacija raziskovalnega vprašanja, ki vodi do nove definicije "stanja" ali nove definicije epizode, fundamentalno reorganizacijo podatkov.¹⁰

Popolna zgodovina dogodkov za posameznika v določenem časovnem obdobju zahteva poznavanje tehle informacij:

y_{i0}	začetno stanje
n_i	število epizod v opazovanem obdobju
$t_{i1} \leq t_{i2} \leq \dots \leq t_{in}$	časovne točke, na katerih se je zgodila sprememba stanja oziroma se je zgodil dogodek
$y_{i1}, y_{i2}, \dots, y_{in}$	nova stanja, ki ustrezajo gornjim časovnim točkam
d_i	indikator, ki kaže, ali je n_i -ta epizoda cenzurirana ali ne
$x_{i1}, x_{i2}, \dots, x_{in}$	vektor kovariat, merjen na začetku vsake epizode

O AZD podatkih govorimo, kadar vsak zapis v datoteki ustreza natančno enemu dogodku ali eni epizodi. Če vsak posameznik oziroma enota opazovanja generira samo eno epizodo, potem število zapisov v datoteki ustreza številu enot opazovanja. Kadar pa imamo opravka s ponavljajočimi se dogodki, kar je v družboslovju prej pravilo kot izjema, potem število zapisov v datoteki ustreza vsoti vseh enot-specifičnih epizod.

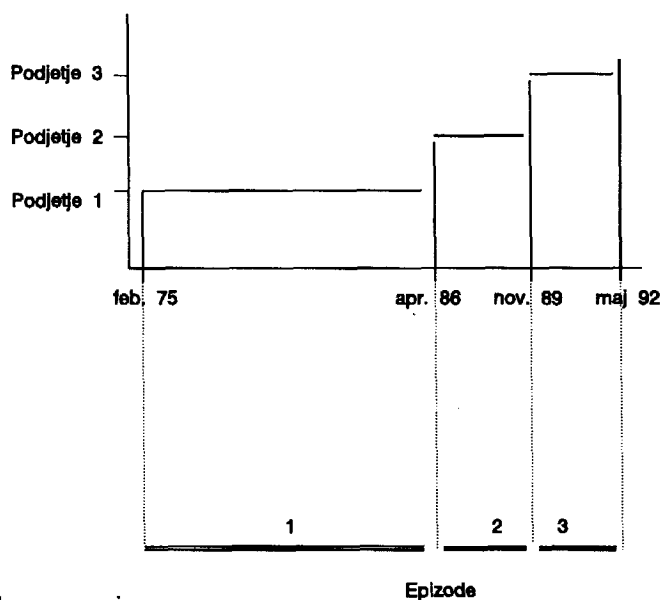
Vsaka epizoda je torej popolno opisana, če poznamo začetni čas, končni čas, začetno stanje in končno stanje. Kot že rečeno, se v AZD namesto analiziranja začetnega in končnega časa običajno uporablja trajanje oziroma dolžina obdobja med tema dvema točkama kot odvisna spremenljivka. Če sta začetno in končno stanje enaki, je epizoda desno cenzurirana; to informacijo vsebuje cenzurni indikator d , ki ga priključimo vsaki epizodi. Epizodam lahko priključimo tudi kovariate, ki so neodvisne od časa (x), ali pa so, če opisujejo zgodovino procesa ali druge paralelne procese, časovno variabilne ($x(t)$).

Poglejmo si na praktičnem primeru, kako morajo biti pripravljeni podatki, da jih lahko uporabimo pri analizi zgodovine dogodkov. Vzemimo za ilustracijo hipotetični primer študije medorganizacijske zaposlitvene mobilnosti; zanima nas, kako pogosto in zakaj ljudje prostovoljno menjajo službo. Da bi poenostavili problem, predpostavimo, da delavcev ne odpuščajo, da se delavci ne upokojujejo in ne zapuščajo trga delovne sile. Fluktuacija je le prostovoljna. Dogodek, ki nas zanima, je torej zamenjava službe oziroma delovne organizacije. Predpostavimo tudi, da podatki niso levo cenzurirani.

Podatke smo zbrali maja 1992 v anketi, in sicer v mesečnih intervalih. Iz teh podatkov rekonstruiramo zgodovino medorganizacijske mobilnosti za prvo osebo: ta oseba se je prvič zaposlila februarja 1975 v podjetju 1. Aprila 1986 je zamenjala službo in šla v podjetje 2, novembra 1989 pa se je zaposlila v podjetju 3, kjer je v času anketiranja (maj 1992) še vedno zaposlena. Dodatno še zvemo, da je to ženska, rojena oktobra 1950, ima za sabo 14 let šolanja, prvič se je poročila februarja 1985. Ti podatki nam omogočajo analizirati vpliv časa na proces fluktuacije (ali se z daljšim zaposlitvenim stažem v podjetju tveganje fluktuacije povečuje ali zmanjšuje ali je neodvisno od dolžine zaposlitve), ter vpliv dodatnih dejavnikov, ki tudi lahko vplivajo na fluktuacijo: spol, starost, izobrazba, zakonski stan.

Graf 2 prikazuje sliko medorganizacijske mobilnosti za to anketiranko. Oseba generira tri epizode. Epizoda številka 1 se začne februarja 1975, ko se je anketiranka zaposlila v prvem podjetju, ter traja do aprila 1986, ko je zamenjala službo in odšla v drugo podjetje. Epizoda je popolna, saj poznamo začetni in končni datum ter začetno in končno stanje. Druga epizoda se začne s prihodom v drugo podjetje ter se konča novembra 1989, ko je anketiranka spet zamenjala službo. Tudi druga epizoda je necenzurirana. Tretja epizoda se začne novembra 1989 s prihodom v tretje podjetje ter konča maja 1992, ko je bila anketa izvedena. Anketiranka je še vedno v podjetju št. 3. Zaključek epizode je torej "umeten"; epizoda je desno cenzurirana, kar moramo naznačiti s cenzurnim indikatorjem, ko izvajamo statistično analizo.

Medtem ko nam graf zelo plastično in razumljivo predstavi zaposlitveno zgodovino posameznika, pa računalnik, ki bo izvajal analizo, podatkov ne zna "prebrati" v tej obliki. Zato jih moramo pripraviti na ustrezen način.



- 1 = epizoda št.1, necenzurirana
- 2 = epizoda št.2, necenzurirana
- 3 = epizoda št.3, desno cenzurirana

Graf 2 Študija zaposlitvene fluktuacije: grafična predstavitev epizod, ki opisujejo medorganizacijsko mobilnost

Tabela 1 kaže, kako je strukturirana datoteka za analizo stopnje tveganja medorganizacijske fluktuacije. Začetek prve epizode je februar 1975. Ker je ta oblika datuma nerodna za računalniško branje, kodiramo čas drugače. Vse datume kodiramo na enoten način, na primer po številu mesecev v tem stoletju. Februar 1975 torej predstavlja vrednost 902 ($(75 \cdot 12) + 2 = 902$), začetno stanje je zaposlitev v podjetju 1. Ta epizoda se konča v času 1036 ($(86 \cdot 12) + 4$), in sicer je končno stanje zaposlitev v podjetju 2.

Tabela 1 Primer AZD podatkov, pripravljenih za analizo fluktuacije.

Oseba št.	Čas začet.	Čas konec	Trajanje	Stanje začet.	Stanje konec	Cenzur. indik.	Št. epiz.	Spol	Starost	Delov. doba	Izobr.	Prva poroka	Zakon. stan	...
1	902	1036	134	1	2	1	1	2	292	0	14	1022	0	...
1	1036	1079	43	2	3	1	2	2	426	134	14	1022	1	...
1	1079	1109	30	3	3	0	3	2	469	177	14	1022	1	...
2	725	998	273	1	2	1	1	1	228	0	8	718	1	...
.	.	.	.	.	.	.	.	.	.	.	.	.	.	...
.	.	.	.	.	.	.	.	.	.	.	.	.	.	...

Trajanje epizode je 134 mesecev (1036-902). Cenzurni indikator kaže, da epizoda ni cenzurirana, saj se začetno in končno stanje razlikujeta. V datoteki imamo še druge spremenljivke, za katere predvidevamo, da lahko vplivajo na delovno fluktuacijo: spol (2=ženski), izobrazba (14 let šolanja). Na osnovi rojstnih podatkov izračunamo starost anketiranke na začetku vsake epizode, izraženo v mesecih. Ker na fluktuacijo lahko vplivata tudi dolžina delovne dobe in zakonski stan, vključimo tudi ti dve spremenljivki. Delovna doba ob začetku prve zaposlitve je 0 mesecev, iz datuma poroke (1022) pa lahko oblikujemo novo spremenljivko, zakonski stan; ta bo imela v tej epizodi vrednost 0, saj oseba ob začetku epizode še ni poročena.

Naslednji zapis v datoteki generira ista oseba (tabela 1). Čakalno obdobje se začne aprila 1986 ($(86 \cdot 12) + 4 = 1036$) in traja 43 mesecev. Končno stanje se razlikuje od začetnega; epizoda ni cenzurirana. Spol in izobrazba se med opazovalnim obdobjem ne spreminjata, vrednosti so enake kot v prvi epizodi. Starost je izračunana ob začetku epizode, prav tako kumulativna delovna doba. Ker je datum poroke pred datumom začetka epizode, vemo, da je oseba ob začetku epizode poročena (zakonski stan = 1).

Tretji zapis oziroma tretja enota analize je epizoda, katere datum začetka in začetno stanje sta znana, datum morebitne spremembe stanja pa ne; zaključek epizode narekuje datum anketiranja, zato je epizoda desno cenzurirana (cenzurni indikator=0). To je tretja epizoda v sekvenci, ki jo generira oseba št. 1. V zapis so vključene že znane časovno nevariabilne in časovno variabilne spremenljivke.

Četrty zapis v datoteki pa generira nova oseba; v tabeli 1 vidimo, da je to moški z osemletno izobrazbo, ki se je prvič zaposlil maja 1960 (725), ter ostal zaposlen v isti delovni organizaciji kar 273 mesecev. Sledijo še druge informacije, ki jih upoštevamo v analizi zgodovine dogodkov.

Bistveno pri pripravi podatkov za analizo zgodovine dogodkov je, da najprej konceptualno dojamemo, da se podatkovni zapis ne nanaša na posameznika, kot smo navajeni v cross-section analizi, ampak na dogodek oziroma epizodo. Le na ta način namreč lahko analiziramo družbene procese, katerih inherentna lastnost sta trajanje in razvoj v časovni dimenziji. Atributi posameznika, na primer sociodemografske

značilnosti, so lahko pomembni in jih zato kot kovariate vključujemo v analizo; vendar jih pri strukturiranju podatkov razumemo kot attribute časovnih epizod, ki jih ta posameznik generira, ne pa kot attribute osebe.¹¹

4.2. Cepitev epizode: diskretna kovariata

V tem pregledu AZD smo že nekajkrat mimogrede naleteli na problem, ki ga povzročajo časovno variabilne spremenljivke. Tehnično govorimo o časovno nevariabilnih spremenljivkah takrat, kadar so kovariate merjene na začetku epizode ter njihova vrednost med epizodo ostane nespremenjena. Večkrat pa imamo opravka z atributi posameznika oziroma sistema, ali pa z vzporednimi procesi, ki se spreminjajo med trajanjem epizode. Ti dogodki ali vzporedno potekajoči procesi vplivajo na dolžino epizode, saj spremenijo stopnjo tveganja, vendar ta vpliv, če imamo na voljo samo vrednosti merjenja na začetku epizode, v analizi ni upoštevan. Kaj storiti v tem primeru?

Če domnevamo, da ima neka kovariata odločujoč vpliv na proces, ki ga proučujemo, je ustaljena metoda, ki jo uporabljamo pri ML (maximum likelihood) proceduri, cepljenje epizod v podepizode. Poglejmo si najprej primer diskretne časovno variabilne kovariate.

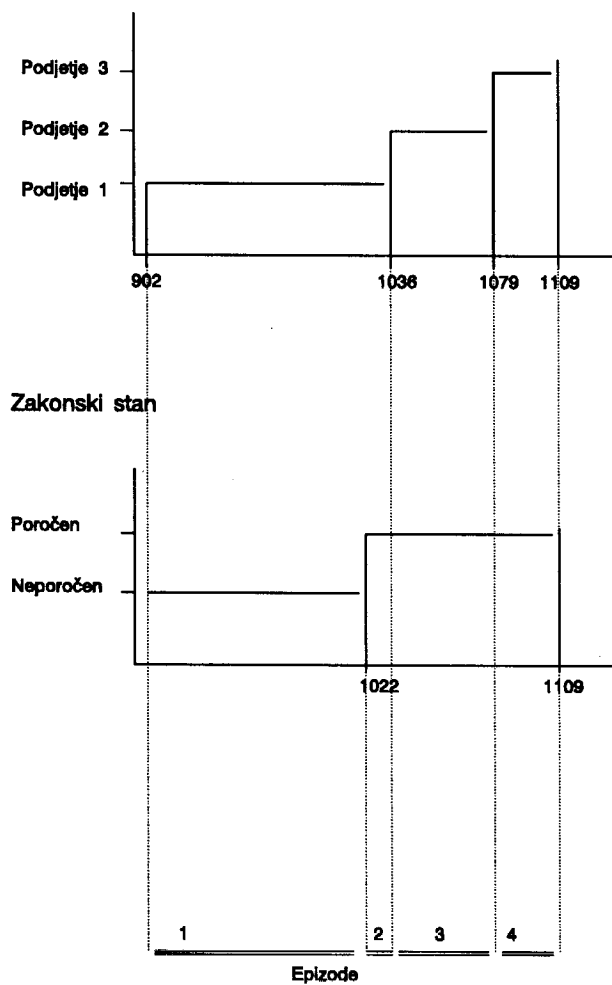
V našem primeru medorganizacijske mobilnosti je taka diskretna časovno variabilna spremenljivka zakonski stan. Spremembo stanja označuje diskreten dogodek - poroka, ločitev, smrt - ki se lahko zgodi v katerikoli časovni točki. Malo verjetno je, da se bo prehod med dvema stanjema (neporočen/poročen, poročen/ločen) zgodil ravno na začetku zaposlitvene epizode. Najverjetneje se bo dogodek zgodil znotraj epizode, tako kot je bilo to tudi v našem primeru v tabeli 1. Če domnevamo, da zakonski stan vpliva na fluktuacijo - recimo, da so poročeni delavci bolj stabilni v zaposlitvi - potem je neupoštevanje dejstva, da je prišlo do spremembe zakonskega stanja med prvo epizodo, problematično.

Da bi vzeli v obzir vzporedni proces ali več procesov, ki vplivajo na rezultate, razcepimo opazovano trajanje čakalnega obdobja t_i na osnovi s_i dogodkov v s_i+1 neodvisnih podepizod. Primer cepljenja zaposlitvenih epizod na osnovi spremembe zakonskega stana je predstavljen v grafu 3.

Zgornji del slike prikazuje proces medorganizacije mobilnosti, kot ga že poznamo iz grafa 2. Oseba je med zaposlitveno kariero dvakrat zamenjala delovno organizacijo ter na ta način generirala tri zaposlitvene epizode. V sredini grafa je ponazorjen vzporedni proces, ki poteka istočasno z dogajanjem na zaposlitvenem področju. Gre za spremembo zakonskega statusa. Oseba, ki je bila ob začetku opazovanega obdobja samska, je prestopila v stanje "poročen" v časovni točki 1022, ki pride znotraj prve epizode. Da bi lahko analizirali vpliv spremembe v zakonskem stanu na fluktuacijo, razcepimo epizodo, v kateri se je sprememba zgodila, na dve podepizodi (graf 3, spodaj). Prva epizoda traja sedaj od začetka prve zaposlitve do poroke (902-1022). Zakonski stan med celo epizodo je zdaj "neporočen", kar ustreza realnemu stanju. Epizoda se ne konča s spremembo delovne organizacije, zato je desno cenzurirana. Stanje ob koncu epizode ostaja organizacija št. 1, cenzurni indikator je 0. Preostale kovariate so normalno priključene tej epizodi.

Druga epizoda se začne v času 1022 in traja do dogodka (zamenjava službe) v točki 1036. Končno stanje je organizacija št. 2, epizoda ni cenzurirana ($d=1$). Kot pravilno kažejo podatki, je oseba med celotno epizodo poročena. Preostale kovariate so priključene epizodi; nekatere med njimi (starost, delovna doba) izmerimo ob začetku epizode. Tabela 2 kaže, kako bodo urejeni podatki po cepitvi epizode.

Medorganizacijska mobilnost



Graf 3 Cepitev epizode pri diskretni časovno variabilni kovariati

Tabela 2 AZD podatki po cepitvi epizode pri diskretni časovno variabilni kovariati

Oseba št.	Čas začet.	Čas konec	Trajanje	Stanje začet.	Stanje konec	Cenzur. indik.	Št. epiz.	Spol	Starost	Delov. doba	Izobr.	Prva poroka	Zakon. stan	...
1	902	1022	120	1	1	0	1	2	292	0	14	1022	0	...
1	1022	1036	14	1	2	1	2	2	412	120	14	1022	1	...
1	1036	1079	43	2	3	1	3	2	426	134	14	1022	1	...
1	1079	1109	30	3	3	0	4	2	469	177	14	1022	1	...
2	725	998	273	1	2	1	1	1	228	0	8	718	1	...
.	.	.	.	.	.	.	.	.	.	.	.	.	.	...
.	.	.	.	.	.	.	.	.	.	.	.	.	.	...

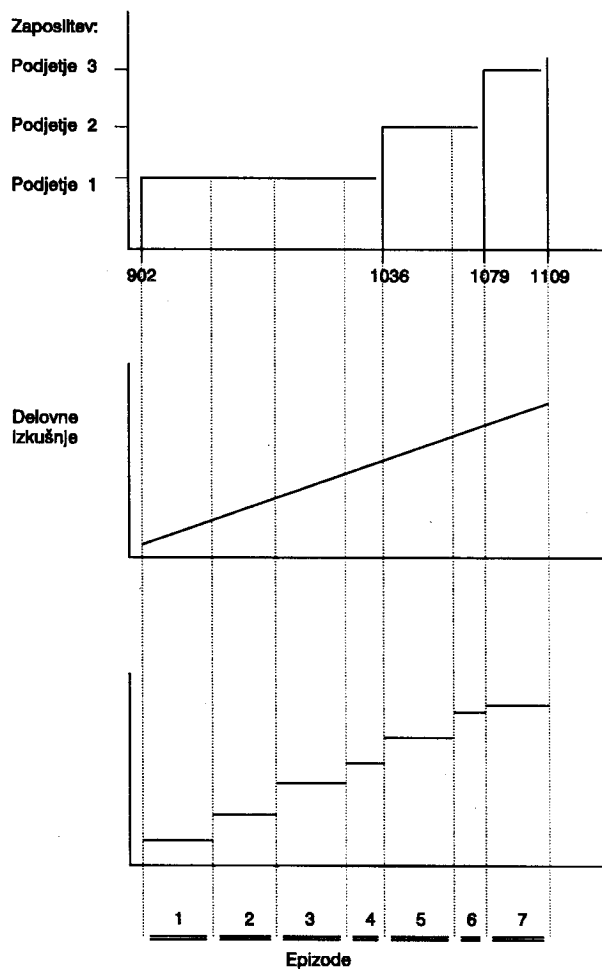
4.3. Cepitev epizode: zvezna kovariata

Zvezno časovno variabilno neodvisno spremenljivko navadno merimo ob začetku epizode in jo kot kovariato vključimo v analizo. Vendar to pomeni, da ostaja vrednost te spremenljivke nespremenjena med celotnim čakalnim obdobjem, kar ni realno. Zato je včasih priporočljivo, da cepimo epizodo tudi na osnovi zvezne časovno variabilne spremenljivke. Vrednosti za zvezno časovno variabilno spremenljivko izračunamo za več časovnih točk in nato umetno razcepimo epizodo v teh točkah.

Vzemimo kot primer dolžino delovne dobe, ki služi kot indikator delovnih izkušenj. V našem primeru interorganizacijske mobilnosti predvidevamo, da imajo delovne izkušnje neposreden vpliv na fluktuacijo; delovne izkušnje rastejo linearno s časom, ki ga posameznik preživi v delovnem razmerju. Iz tabele 1 je razvidno, da so delovne izkušnje merjene zelo grobo. V prvi epizodi, ki jo generira oseba 1, imajo vrednost 0, v drugi epizodi, ki jo generira ista oseba, vrednost nenadoma poskoči na 134 mesecev. V analizi prve epizode se torej še po enajstih letih službovanja upošteva 0 mesecev delovne dobe. Če je res, da delovne izkušnje ob kontroli drugih dejavnikov vplivajo na fluktuacijo, bodo rezultati v primeru, ki smo ga navedli, pristranski.

Graf 4 kaže primer cepitve zaposlitvene epizode na podlagi delovnih izkušenj. V tem primeru smo originalne epizode razcepili v maksimalno 36 mesecev oziroma tri leta dolge podepizode. Delovne izkušnje, merjene z dolžino delovne dobe, smo izmerili na začetku vsake podepizode. Razcepitev generira skupaj sedem podepizod. Le tiste epizode, ki se končajo z zamenjavo službe, so necenzurirane. Strukturo podatkov po cepitvi epizode prikazuje tabela 3.

Cepitev epizode v triletno intervale je v prikazanem primeru arbitrarna. V praksi nam pri odločitvi pomaga teorija. Če so teoretične predpostavke nejasne, pa vendarle menimo, da je cepitev epizode smiselna, si lahko pomagamo tako, da epizode razcepimo na več alternativnih načinov, na primer na dveletne, triletno in petletne podepizode. Nato ocenimo identičen model, ki vključuje spremenljivko "delovne izkušnje" na vseh treh AZD datotekah, ter testiramo, katera specifikacija bolj poveča eksplanatorno moč modela.



- 1 = epizoda št.1, desno cenzurirana
- 2 = epizoda št.2, desno cenzurirana
- 3 = epizoda št.3, desno cenzurirana
- 4 = epizoda št.4, necenzurirana
- 5 = epizoda št.5, desno cenzurirana
- 6 = epizoda št.6, necenzurirana
- 7 = epizoda št.7, desno cenzurirana

Graf 4 Cepitev epizode pri zvezni časovno variabilni kovariati

Tabela 3: AZD podatki po cepitvi epizode pri zvezni časovno variabilni kovariati.

Oseba št.	Čas začet.	Čas konec	Trajanje	Stanje začet.	Stanje konec	Cenzur. indik.	Št. epiz.	Spol	Starost	Delov. doba	Izobr.	Prva poroka	Zakon. stan	...
1	902	938	36	1	1	0	1	2	292	0	14	1022	0	...
1	938	974	36	1	1	0	2	2	328	36	14	1022	0	...
1	974	1010	36	1	1	0	3	2	364	72	14	1022	0	...
1	1010	1036	26	1	2	1	4	2	400	108	14	1022	0	...
1	1036	1072	36	2	2	0	5	2	426	134	14	1022	1	...
1	1072	1079	7	2	3	1	6	2	462	170	14	1022	1	...
1	1079	1109	30	3	3	0	7	2	469	177	14	1022	1	...
2	725	761	36	1	1	0	1	1	228	0	8	718	1	...
.	.	.	.	.	.	.	.	.	.	.	.	.	.	...
.	.	.	.	.	.	.	.	.	.	.	.	.	.	...

5. ZBIRANJE PODATKOV

5.1. Viri podatkov

Doslej smo si ogledali osnovna načela analize zgodovine dogodkov ter primerno obliko datoteke, ki tako analizo omogoča. Vendar so prvi pogoj za vse to kakovostni podatki. Zbiranje AZD podatkov je zahtevnejše, dolgotrajnejše in dražje kot zbiranje informacij o trenutnem stanju ali dogajanju. AZD podatki morajo omogočiti popolno rekonstrukcijo procesa; vsebovati morajo časovne točke, ko so se dogodki oziroma spremembe zgodile, ter registrirati spremembo, torej stanje pred dogodkom in po njem.

Ena od možnih tehnologij za zbiranje podatkov je tekoč registracijski sistem, ki stalno sledi procesu, ki ga proučujemo, ter sproti registrira vse relevantne aktivnosti in dogodke. Vendar se v praksi na ta način zbirajo podatki le v posebnih primerih, na primer za proučevanje delovanja organizacij ali institucij, ki navadno že tako ali tako rutinsko zbirajo informacije o svojih tekočih aktivnostih. Včasih se na tak način zbirajo tekoči podatki v obliki podrobnih dnevniških zapisov tudi v študijah, ki se ukvarjajo z vprašanjem, kako posamezniki izrabljajo čas, vendar gre pri tem za kratkotrajna obdobja, za nekaj dni ali kvečjemu nekaj tednov.

Drug način zbiranja podatkov za analizo zgodovine dogodkov je arhivski način. Številne organizacijske študije, ki proučujejo pojavljanje in izumiranje organizacijskih oblik (glej n.pr. Hannan in Freeman 1989), temeljijo na podatkih o datumu ustanovitve in propada organizacij, zbranih iz različnih arhivov. Ali pa se podatki o dogodkih, o katerih ni sistematičnih registrov, na primer o demonstracijah, nemirih, stavkah, zbirajo iz časopisnih poročil. Ker se večina časopisov arhivira in so dostopni še desetletja po izidu, so lahko dragocen vir podatkov o družbeno pomembnih dogodkih.

Vendar pa je potrebno veliko večino podatkov, ki jih uporabljamo v socioloških študijah, zbirati retrospektivno. Eden od možnih načinov je panelna študija. Z njo v sorazmerno kratkih intervalih anketiramo vzorec posameznikov. Primer tako študije je nemški Socioekonomski panel (Hanefeld 1987). Ta panelna študija se izvaja vsako leto

od leta 1984 naprej in sicer na reprezentativnem vzorcu gospodinjstev v ZRN. Kot pove že ime, ostaja vzorec gospodinjstev isti oziroma se spreminja le toliko, kolikor posamezniki in gospodinjstva izpadejo iz vzorca zaradi smrti, selitve v tujino ali zavračanja sodelovanja. Hkrati se v vzorec zajemajo nova gospodinjstva, ki se oblikujejo iz članov poprej zajetih gospodinjstev. Ob vsakem intervjuju se zbirajo tudi podrobnosti o dogodkih in aktivnostih med anketami, sicer bi dobili le cross-section posnetke iz življenja posameznikov. Taka panelna študija zagotavlja zelo kakovostne in izčrpne podatke, tako da postaja v ZRN Socioekonomski panel "najpopularnejša" družboslovna podatkovna baza, ki se uporablja za vrsto analiz dinamičnih sprememb v življenju posameznikov in gospodinjstev.

Taka študija je iz več razlogov praktično težko izvedljiva. Poleg visokih stroškov je za tako študijo problematičen tudi čas; potekati mora skozi daljše časovno obdobje, preden so podatki uporabni za analizo.

Cenejši in bolj praktičen način zbiranja AZD podatkov je standardna cross-section anketa, ki zbere retrospektivne podatke za daljše časovno obdobje v preteklosti, torej za več let ali celo desetletij. Vendar se pri tem odpira vprašanje, kako zagotoviti zanesljivost retrospektivno zbranih informacij, saj so le-te tem bolj podvržene spominskim napakam in ohlapnostim, čim daljše je obdobje, ki ga v raziskavi pokrivamo.

Kakovost retrospektivnih podatkov je možno povečati s skrbnim in premišljenim postopkom zbiranja. Blossfeld, Hamerle in Meyer (1989) poročajo, da so pilotne študije za obsežno German Life History Study (Nemška življenjepisna študija) pokazale, da retrospektivno zbrani podatki o objektivnih življenjskih dogodkih niso sistematično pristranski zaradi nesposobnosti respondentov, da bi odgovorili na zastavljena vprašanja ali zaradi spominskih vrzeli. Vendar pa sta za kakovost odgovorov ključnega pomena oblika in preciznost anketnega instrumenta. Predvsem se je pokazalo, da je nujno potrebno razdeliti življenjsko zgodovino v različne življenjske sfere, kot na primer izobraževanje, zaposlitvena kariera, družinski dogodki, selitve. Temu je sledilo obsežno editiranje podatkov ter preverjanje konsistentnosti po različnih življenjskih področjih.

Kot pomoč pri zbiranju podatkov so raziskovalci razvili različne tehnike; od teh je ena najbolj obetajočih uporaba življenjepisnega koledarja (Life History Calendar).

5.2. Življenjepisni koledar ¹²

Življenjepisni koledar je del vprašalnika ali pa ločen dokument, ki omogoča grafično predstavitev zgodovine več vrst dogodkov v življenjskem ciklu. Format koledarja je navadno mreža; ena dimenzija predstavlja detajlno razdelane vzorce obnašanja za pojav, ki ga proučujemo, druga dimenzija matrike pa je razdeljena v časovne enote, v katerih te vzorce obnašanja kodiramo. Anketa z informacijami, ki jih dobi od respondenta, na predpisan način izpolnjuje celice v tej mreži.

Zakaj ima življenjepisni koledar prednost pri zbiranju retrospektivnih podatkov pred klasično obliko vprašalnika? Prvič, respondentu pomaga rekonstruirati dogodke v njegovem življenju, ker mu omogoči vizualno in mentalno povezati časovne točke, v katerih so se različni dogodki zgodili. Nekateri dogodki in njihovi datumi, kot na primer selitev, poroka, rojstvo otroka, so posameznikom dobro v spominu in so pomembne oporne točke pri rekonstruiranju manj izpostavljenih dogodkov, kot so na primer podrobnosti v zvezi z zaposlitveno kariero. Koledarska oblika poveča preglednost in opozori tako anketaarja kot respondenta na morebitne časovne nekonsistentnosti pri dogodkih z različnih življenjskih področij.

Drugič je detaljne sekvenčne dogodke lažje zapisati oziroma kodirati z življenjepisnim koledarjem kot pa s konvencionalnim vprašalnikom. Predvsem stanja, ki se lahko spreminjajo pogosto in v različnih časovnih intervalih, je neprimerno bolj zapleteno kodirati s konvencionalnim vprašalnikom. V koledarju uporabljamo grafične simbole, s katerimi označimo začetni in končni datum (na primer mesec), ter jih povezujemo s črtami, ki označujejo tekoče aktivnosti.

Z zapisom oziroma označbo v koledarju delo seveda še ni končano, saj je potrebno v končni fazi koledar prevesti v računalniško berljivo datoteko, vendar je v fazi zbiranja podatkov, ko je najpomembneje dobiti od respondentov čimbolj natančne in zanesljive podatke, grafična predstavitev dogodkov, sprememb in stanj z življenjepisnim koledarjem skorajda nepogrešljiva.

5.3. Dizajn koledarja

Ena od pomembnih odločitev pri pripravi življenjepisnega koledarja je izbira časovne enote. Ali se bodo podatki nanašali na dan, teden, mesec ali leto, je odvisno od podatkovnih potreb v raziskavi. Časovna enota mora biti dovolj majhna, da bo omogočila razpoznati sekvence dogodkov in njihovo časovno medsebojno odvisnost. Dogodki, med katerimi so intervali čakalnega obdobja kratki, zahtevajo precej podrobne časovne enote. Po drugi strani pa je treba upoštevati tudi spominske sposobnosti respondentov za razločevanje kratkih časovnih enot ter praktične probleme pri pripravi koledarja, ki mora imeti obvladljiv obseg.

Izkušnje kažejo, da je mesec kot časovna enota v družboslovnih raziskavah enota, ki je hkrati dovolj podrobna za kodiranje večine družbenih procesov ter obvladljiva glede na spominske sposobnosti. Freedman et al. (1988) so uporabljali mesečne časovne enote za obdobje od 15. do 23. leta starosti anketirancev, torej za obdobje osmih let. McPherson, Popielarz in Drobnič (1992) so uporabljali mesečne intervale za 15-letno časovno obdobje - od 1974 do 1989 - (oziroma obdobje od 15. leta starosti naprej za mlajše respondente), v katerem so proučevali dinamiko vključevanja posameznikov v zaposlitvene in tako imenovane prostovoljne organizacije. Blossfeld, Hamerle in Mayer (1989) tudi poročajo, da so zaposlitvene kariere v Nemški življenjepisni študiji kodirane na mesečni osnovi, in sicer za obdobje od vključitve v prvo zaposlitev pa do izvedbe študije, kar za starejše respondente lahko pomeni obdobje več desetletij. Tudi nemški Socioekonomski panel zajema podatke v mesečnih intervalih.

Druga pomembna odločitev pri uporabi življenjskega koledarja je določiti obdobje, ki ga želimo zajeti. To obdobje je lahko definirano za vse enako, recimo zadnjih 10 ali 15 let, lahko pa je odvisno od posameznika, recimo od anketirančevega rojstva, od izpolnitve določene starosti, od njegove prve zaposlitve ali kaj podobnega. Če je časovno obdobje definirano vnaprej, je priporočljivo v intervjuju uporabiti kak vsesplošno svetovno ali nacionalno pomemben dogodek, za katerega je verjetno, da se ga večina ljudi dobro spominja. Težko je namreč pričakovati, da se bodo anketiranci na ukaz spomnili, kje so živeli, kaj so počeli ali kaj se jim je dogajalo na primer poleti 1977. Na neki način je potrebno sprožiti spominski proces. V ZDA so kot take sprožilne dogodke uporabljali umor Kennedyja, afero Watergate, konec vojne v Vietnamu. V Sloveniji bi morali ob določitvi opazovanega obdobja identificirati kak dogodek, ki bi ljudem lahko priklical v spomin zahtevano obdobje.¹³

Naslednja pomembna odločitev je izbira področij oziroma dimenzij raziskovanja z življenjepisnim koledarjem. Praviloma bodo te dimenzije seveda odsevale namen študije,

vendar morajo biti področja natančno definirana in njihovo število ne sme biti preobsežno. Morda je vredno še enkrat poudariti, da so edino smiselni "kandidati" za AZD študijo področja, za katera lahko definiramo objektivne dogodke oziroma spremembe in rekonstruiramo časovno shemo njihovega pojavljanja. Spremembe v vrednotah, mnenjih ali družbenih normah - čeprav jih tudi lahko interpretiramo kot indikatorje dinamičnih družbenih procesov - očitno ne zadostijo tem zahtevam.

V življenjepisni koledar lahko vključimo različne vrste spremenjivk - nominalne, ordinalne ali intervalne. Vendar pa raven merjenja vpliva na dizajn koledarja in vrsto informacij, ki jih kodiramo v koledar.

Pri vnašanju informacij v koledar mora raziskovalec ugotoviti "stanje" vsake aktivnosti v vsaki časovni enoti, označiti prehode med stanji ter indentificirati sukcesivne epizode vsake aktivnosti. Za nekatere aktivnosti je koristno dokumentirati vrsto dogodka, ki je privedel do prehoda iz enega stanja v drugo. Na primer, pri prehodu iz življenja z zakonskim partnerjem v "samski" bivalni aranžma je za kasnejšo analizo verjetno koristno identificirati ne le čas prehoda in spremembo bivalnega oziroma družinskega statusa, ampak tudi dogodek, ki je privedel do spremembe stanja: razveza zakona, smrt zakonca, odselitev partnerja zaradi nesporazumov, vendar brez formalne razveze, ločeno življenje zaradi drugih razlogov, n.pr. služenja vojaškega roka, ali pa šolanja oziroma zaposlitve partnerja (ali anketiranca) v drugem kraju.

Življenjepisni koledar mora torej zagotoviti, da bo vsaka epizoda, uporabljena v statistični analizi, popolno in nedvoumno opisana: začetni čas, končni čas, začetno stanje, končno stanje ter cenzurni indikator. Zato mora življenjepisni koledar za vsako časovno enoto zabeležiti, ali se je specifična aktivnost takrat začela, nadaljevala, končala, ali pa se ni dogajala.

Praktična vrednost AZD podatkov je močno odvisna od kakovosti dela na terenu. Anketaarji morajo biti dobro seznanjeni z metodo življenjepisnega koledarja in izvrstno izurjeni. Freedman et al. (1988) poročajo, da so uporabljali že usposobljene anketarje, kljub temu pa je trening za njihovo študijo zahteval šest dni intenzivnih priprav; začetno petdnevno urjenje ter dodaten dan za "ponavljanje", ko se je delo na terenu že začelo. To pomeni, urjenje trikrat daljše kot normalne priprave z izkušenimi anketarji v poprejšnjih etapah njihove študije.

Anketarji morajo biti usposobljeni ne le slediti instrumentu, ampak tudi sproti zaslediti neusklajenosti v koledarju ter nanje opozoriti respondenta. Ko respondent ve za napake, je v večini primerov sposoben rekonstruirati pravo sekvenco dogodkov. Tehniko odkrivanja napak in način feedback komuniciranja z respondenti je potrebno anketarje naučiti na uvodnem seminarju.

Življenjepisni koledar je posebej primeren za osebni intervju, kjer ima anketiranec tudi sam pregled nad dokumentom. Po nekaterih poročilih deluje koledar kot dodaten motivacijski dejavnik, saj anketirancem omogoča vizualno in mentalno rekonstrukcijo poteka njihovega življenja, kar je zanje zanimivo in privlačno. Če se koledar uporablja pri telefonskem intervjuju, se anketirancem ne razlaga o koledarju, podatke pa se na enak način vnaša v dokument.

Tudi če je osnovni "vprašalnik" že računalniško sprogramiran in se odgovore vnaša neposredno v računalnik, je priporočljivo tiste dele intervjuja, ki so v obliki, primerni za analizo zgodovine dogodkov, spremljati z življenjepisnim koledarjem.

Opombe

- 1 Kjer avtor(ji) ni(so) posebej naveden(i), smo za predstavitev analize zgodovine dogodkov uporabili predvsem naslednje vire: Allison (1984); Blossfeld, Hamerle in Mayer (1989); Hannan (1989); Kalbfleisch in Prentice (1980); Tuma in Hannan (1984). Za orientacijo potencialnim uporabnikom izvirm literature: Allison je pisan na elementarni ravni, primeren je za uvodno seznanjanje z AZD. Kalbfleisch in Prentice sta najbolj "tehnična" in strokovno dognana - knjiga zahteva določeno predznanje statistične teorije. Hannan in Tuma ter Blossfeld, Hamerle in Mayer pišejo bolj neposredno za družboslovce.
- 2 Formalizacijo problema dinamičnih prehodov med stanji na osnovi panelnih podatkov in empirični primer glej v Antončič (1992).
- 3 O kompetitivnih tveganjih (competing risks) govorimo, kadar se epizoda lahko konča z več vrstami dogodkov. Recimo, da nas zanima, kaj vpliva na odpuščanje delavcev. Epizode, ki jih proučujemo, opisujejo čas od zaposlitve do prenehanja dela v tem podjetju. Vendar je prenehanje zaposlitve lahko tudi posledica prostovoljne zapustitve dela, upokojitve, smrti, itd. Če nas torej zanima odpuščanje delavcev, moramo izračunati specifične stopnje za vsako vrsto prehoda: iz zaposlitve v neprostovoljno brezposelnost, iz zaposlitve v prostovoljno brezposelnost, iz zaposlitve v upokojitev itd... Vsota teh specifičnih stopenj nam da stopnjo tveganja za prenehanje dela v podjetju. Specifična tveganja so "kompetitivna" zato, ker različni načini prekinitve zaposlitve "tekmujejo" med seboj; delavec, na primer, ki sam zapusti delovno mesto, ne "tvega" več, da bo odpuščen.
- 4 Drugi izrazi za "hazard rate" v angleški literaturi: hazard function, intensity function, risk function, transition rate, mortality rate, failure rate.
- 5 Za boljše razumevanje si lahko funkcijo tveganja poenostavljeno predstavljamo kot verjetnost, da se bo dogodek zgodil v določenem času oziroma časovnem intervalu. Vendar se moramo zavedati, da vrednosti funkcije tveganja formalno niso (pogojne) verjetnosti. So sicer vedno nenegativne, lahko pa so večje od 1.
- 6 Sicer pa so posledice levo cenzuriranih podatkov odvisne tudi od modela, ki ga uporabljamo. V primeru eksponentnega modela, ki specifična stopnjo tveganja kot konstantno (neodvisno od časa), levo cenzuriranje ni problematično. V tem primeru smemo levo cenzurirano epizodo obravnavati na enak način, kot če bi se v resnici začela z nastopom opazovanega obdobja.
- 7 Kot pravi Kiefer (1988), se da eksponentni model v AZD primerjati z linearnim regresijskim modelom v regresijski analizi, ki je tudi preprost za ocenjevanje in povsem ustrezen, če podatki ne variirajo do te mere, da bi izražali pomembno nelinearnost.
- 8 Če pa nam grafični test kaže rastočo stopnjo tveganja, ni nevarnosti, da bi bil rezultat izmalichen zaradi populacijske heterogenosti.
- 9 Statistika, na kateri temelji test, je $L_q = 2(\ln L_1 - \ln L_2)$, kjer je $\ln L_1$ log-likelihood vrednost v prvem modelu in $\ln L_2$ log-likelihood vrednost v drugem modelu. Statistika L_q je porazdeljena po χ^2 (hi kvadrat).
- 10 Za shranjevanje in urejanje podatkovnih baz obstaja nekaj programskih paketov (SIR, CASA), vendar - kolikor je meni znano - noben od teh sistemov še ni tako izpopolnjen, da bi zadostil vsem potrebam uporabnikov in bi prevladoval pri uporabi. Zato večina raziskovalcev tudi za urejanje podatkovnih baz uporablja programske pakete, ki so sicer namenjeni predvsem statistični analizi, kot na primer SAS, BMDP, RATE, SPSS ali GLIM.
- 11 To pa ne pomeni, da moramo epizode, ki jih generira ena oseba, nujno obravnavati kot neodvisne enote. Vrnimo se k našemu primeru: posamezniki, ki so bolj podvrženi fluktuaciji, bodo nadprezentirani v AZD podatkih. Če imamo opravka s homogeno populacijo, kar pomeni, da se osebe ne razlikujejo po karakteristikah, ki so pomembne za proces fluktuacije, lahko epizode obravnavamo kot neodvisne. Če pa je populacija heterogena, lahko dobimo zavajajoče rezultate. V tem primeru moramo to upoštevati pri analizi. Eden od možnih načinov je, da poskušamo dezagregirati vzorec, upoštevajoč teoretsko pomembne spremenljivke; obravnavamo na primer ženske in moške posebej, če je spol odločujoč dejavnik pri fluktuaciji, ali obravnavamo posebej prvo zaposlitev, če se izkaže, da se ljudje obnašajo v prvi zaposlitvi drugače kot pri kasnejših zaposlitvah. Lahko pa ravnamo tudi drugače in analiziramo vzorec kot celoto, vendar v analizo vključimo kovariate; v tem zadnjem primeru slavnato spremenljivko, ki za vsako epizodo pove, ali gre za prvo zaposlitev ali ne. Na ta način vsaj delno upoštevamo preteklo zgodovino procesa, ki ga obravnavamo.
- 12 Predstavitev življenjepisnega koledarja sledi članku Freedman, Thornton, Camburn, Alwin in Young-DeMarco (1988). V tem članku avtorji poročajo o raziskavi, katere začetki segajo v leto 1962, ko so anketirali okrog 900 mater iz Detroita in okolice, ki so rodile julija 1961. Prvotnemu intervjuju je sledilo šest panelnih valov. Leta 1980 je bila študija razširjena na otroke, katerih rojstva so matere "pripeljala" v vzorec. Ti otroci so bili takrat stari 18 let. Pet let kasneje, torej leta 1985, je bila raziskava na otrocih ponovljena, le da so tokrat raziskovalci zbirali podatke v AZD obliki in sicer za obdobje od 15. leta starosti pa do časa raziskave, ko so respondenti izpolnili starost 23 let. Podatki so bili zbrani z življenjepisnim

- koledarjem. Posebna vrednost tega primera je, da je podatke za leto 1980 možno kontrolirati s primerjanjem rezultatov iz takratne cross-section ankete. Freedman et al. (1988) poročajo, da so podatki izjemno konsistentni, ter pripisujejo to nadvse skrbnim pripravam in aplikaciji življenjepisnega koledarja.
- 13 Za mlajše anketirance, ki so bili takrat, ko se opazovano obdobje začne, še zelo mladi, se pri vnaprej definiranim časovnem obdobju navadno določi drugačen kriterij. Njih se ne sprašuje po dogodkih od n.pr. leta 1981 naprej, ampak o obdobju od njihove izpolnitve določene starosti (n.pr. 16 let) pa do časa anketiranja.

Reference

- Allison, Paul D. 1984. **Event History Analysis. Regression for Longitudinal Event Data.** Series: Quantitative Applications in the Social Sciences. No. 07-046. Newbury Park, London, New Delhi: Sage Publications.
- Antončič, Vojko. 1992. "Analiza tokov v kadrovskem potencialu" V: Ivan Svetlik in Branka Preželj (ur.) **Zaposlovanje - perspektive, priložnosti, tveganja.** Ljubljana: Znanstveno in publicistično središče.
- Blossfeld, Hans-Peter, Alfred Hamerle in Karl Ulrich Mayer. 1989. "Event History Analysis. Statistical Theory and Application in the Social Sciences." Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bottomore, Tom. 1975. "Structure and History". V: Peter Blau (ur.). **Approaches to the Study of Social Structure.** New York: Free Press. pp. 159-171.
- Freedman, Deborah, Arland Thornton, Donald Camburn, Duane Alwin and Linda Young-DeMarco. 1988. "The Life History Calendar: A Technique for Collecting Retrospective Data." V: Clifford C. Clogg (ur.). **Sociological Methodology 1988.** San Francisco: Jossey-Bass. pp. 37-68.
- Hanefeld, Ute. 1987. **Das Sozio-ökonomische Panel. Grundlagen und Konzeption.** Frankfurt, New York: Campus Verlag.
- Hannan, Michael T. 1989. "Macrosociological Applications of Event History Analysis: State Transitions and Event Recurrences." **Quality and Quantity** 23.
- Hannan, Michael T. in John Freeman. 1989. **Organizational Ecology.** Cambridge MA, London: Harvard University Press.
- Kalbfleisch, John D. in Ross L. Prentice. 1980. **The Statistical Analysis of Failure Time Data.** New York: John Wiley and Sons.
- Kiefer, Nicholas M. 1988. "Economic Duration Data and Hazard Functions." **Journal of Economic Literature.** Vol. XXVI. June: 646- 679.
- McPherson, Miller J., Pamela A. Popielarz in Sonja Drobnič. 1992. "Social Networks and Organizational Dynamics." **American Sociological Review.** Vol. 57, No. 2 (April: 153-170).
- Olzak, Susan. 1989. "Analysis of Events in the Study of Collective Action." **Annual Review of Sociology** 1989. 15:119-142.
- Tuma, Nancy Brandon in Michael T. Hannan. 1984. **Social Dynamics. Models and Methods.** Academic Press.