

PRISTRANSKOST¹ TELEFONSKIH VZORCEV V SLOVENIJI

Povzetek: V Sloveniji ima le dobra polovica gospodinjstev telefonski priključek, zato je lahko sklepanje na osnovi telefonskih vzorcev precej zavajajoče. V članku je analizirana pristranskost vzorčnih ocen, ki nastaja zaradi neenakomerne porazdelitve telefonskih priključkov v populaciji. Z uporabo loglinearnih modelov ter Bayesovega informacijskega kriterija je pokazano, da je povezava med posedovanjem telefona in mnenjskimi spremenljivkami v glavnem navidezna. Spremenljivke, ki tako povezanost pojasnjujejo, so: izobrazba, tip krajevne skupnosti (vaška-mestna), dohodek in vernost. Problem je ilustriran z opredeljevanjem za kandidata Krambergerja ob predsedniških volitvah 1990, kjer je navidezen vpliv telefona izredno močan.

Abstract: *Imprecision of the estimates from telephone surveys in Slovenia: The telephone coverage in Slovenia is below 60%. Thus the telephone surveys can only be of limited value. The paper examines the imprecision which arises from unbalanced distribution of the telephone in target population. It is shown - using loglinear models and Bayesian information criteria - that, generally, the relationship between telephone and opinions is spurious. Explanatory variables are education, rurality, income and religion. As an example, voting preference from presidential election 1990 (candidate Mr. Kramberger) is analyzed, where the spurious link with the telephone was extremely powerful.*

1) Pristranskost pomeni v najožjem smislu le specifično lastnost cenilke pri določeni porazdelitvi, vendar bomo v nadaljevanju pojem razumeli v širšem smislu, kot odstopanje vzorčnih ocen od dejanskih vrednosti parametra.

1. UVOD

Problem. Podatki PTT Slovenije ter ocene na osnovi anket Slovensko javno mnenje 1989 in 1990 govorijo, da ima v Sloveniji telefonski priključek le dobra polovica gospodinjstev. Ocenjevanje parametrov na osnovi telefonskega vzorca, ki sistematično izpušča polovico ciljne populacije, je lahko precej nenatančno, posebej če se oba dela populacije (s telefonom in brez njega) razlikujeta. V nadaljevanju nas bo zanimalo, kakšne so razlike v mnenjskih spremenljivkah med anketiranci, ki imajo doma (v gospodinjstvu) telefon ter preostalimi, ki telefona nimajo. Zanimalo nas bo, kako značilne so te razlike in kako jih je mogoče pojasniti.

Cilji in hipoteze. Želimo torej osvetliti povezanost med socio-demografskimi spremenljivkami, mnenjskimi spremenljivkami ter posedovanjem telefona. Osnovna hipoteza je, da je povezanost telefona in mnenjskih spremenljivk navidezna in nastaja zaradi vpliva socio-demografskih spremenljivk. Hipoteza je intuitivno razmeroma jasna: če imajo npr. izobraženi anketiranci več telefonov in tudi izraziteje podpirajo določeno stališče, potem je očitno, da dobimo na osnovi telefonskega vzorca izkrivljeno oceno za mnenje celotne populacije.

Zožitev problema. V nadaljevanju bomo obravnavali samo tiste probleme, ki nastajajo zaradi neenakomerne porazdelitve telefonskih priključkov v populaciji (npr. anketiranci v urbanem okolju imajo razmeroma veliko telefonov). Vprašanja neslučajnega² izbora enot znotraj "telefonske" populacije (npr. telefonske številke, ki niso v imeniku, neslučajen izbor anketiranca znotraj gospodinjstva, itd.) na tem mestu ne bomo obravnavali. Enako velja tudi za ostale vire napak pri telefonskem anketiranju (npr. neodgovori, merske napake, itd.).

Podatki. Uporabili bomo podatke ankete Slovensko javno mnenje (SJM'90) 1990, izvedene spomladi 1990. Podatki temeljijo na običajnem vzorcu SJM (Blejec, 1970). Ciljna populacija so stalni prebivalci Slovenije v starosti 18-70 let. Velikost vzorca SJM '90 je 2048 enot. Nekatere mnenjske spremenljivke temeljijo - zaradi osipa pri pisemskem panelnem vprašalniku - na manjšem številu anketirancev ($n=1112$)³. Vprašalnik oziroma dokumentacija je na voljo na Raziskovalnem inštitutu Fakultete za sociologijo, politične vede in novinarstvo (FSPN) v Ljubljani.

Predhodne analize. Izčrpen pregled podatkov v SJM90 in SJM89 (MEDIA) kaže, da je večina mnenjskih in tudi sociodemografskih spremenljivk statistično značilno

2) Neslučajno pomeni v zgornjem kontekstu odstopanje od "epsem" vzorčnega načrta, kjer ima vsaka enota enako verjetnost, da se pojavi v vzorcu.

3) Iz pregleda podatkov je razvidno, da anketiranci s telefonom raje odgovarjajo na pisemski vprašalnik, čeprav drugih razlik - razen morda starosti - ni opaziti, kar je nekoliko presenetljivo. Pričakovali bi npr. manjši odziv med anketiranci z nižjo izobrazbo.

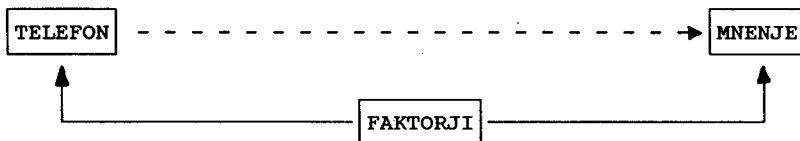
povezana s posedovanjem telefona. Nakazuje se, da imajo anketiranci brez telefona v gospodinjstvu nižji socioekonomski položaj, manj intenzivno spremljajo medije, politično so manj radikalni, bližje jim je načelo uravnilovke, močnejše so zaskrbljeni za socialno neenakost, pogosteje odgovarjajo "ne vem". Problem torej nesporno obstaja.

Metoda. Ker imamo opraviti z zapletenim prepletanjem večjega števila spremenljivk, z običajno bivariatno analizo problema ne moremo obvladati. Uporabili bomo loglinearne modele in iskali model, ki pojasnjuje povezanost med telefonom, socio-demografskimi spremenljivkami ter mnenjskimi spremenljivkami. Pri tem bomo uporabili Bayesov informacijski kriterij (BIC).

Računalniška obdelava. Osnovno tabeliranje podatkov in tudi ocenjevanje parametrov loglinearnih modelov za izbrane modele smo opravili s statističnim paketom SPSS-PC+ (verzija 6), iskanje modelov s paketom MICLOG (McDonald, 1990), pri interpretacijah pa smo uporabljali paket CHIP (Davis, 1990).

Model. Golo dejstvo, da anketiranec ima telefon, najbrž ne more izraziteje vplivati na njegovo mnenje. Najenostavnejši model zato predpostavlja, da tako na mnenje kot tudi na posedovanje telefona vplivajo določeni faktorji - sociodemografske spremenljivke:

Slika 1: Osnovni model



2. IZBIRA FAKTORJEV

Izbor faktorjev. Po vsebinskem preudarku, analizi bivariatnih povezav in naknadnem vstavljanju faktorjev, ki v model niso vključeni (poklic, zaposlitveni status, narodnost, članstvo v stranki), smo se odločili za šest dihotomnih spremenljivk: starost (A), izobrazba (E), ruralnost (R), dohodek (I), vernost (C) in spol (G). S spremenljivko telefon (T) pa bom v nadaljevanju označili dejstvo, ali ima anketiranec v gospodinjstvu telefon.

Tabela 1: Faktorji uporabljeni v analizi

spremenljivka	1	2	3
	pozitivna kat.	odstotek	D
A - starost	40 let in več	45.0	<u>2.9</u>
E - izobrazba	12 let izobr. in več ⁴	39.5	22.5
I - dohodek	visok ⁵	49.2	15.2
R - ruralnost	mesto ⁶	57.6	18.9
C - vernost	ni veren ⁷	22.9	20.0
G - spol	ženski	53.6	1.0
T - telefon	ima telefon ⁸	54.3	54.3

Stolpec 1 označuje pozitivno kategorijo pri dihotomizaciji, to je kategorijo, ki pozitivno vpliva na telefon, (npr. anketiranci stari nad 40 let imajo pogosteje telefon),

Stolpec 2 vsebuje delež pozitivne kategorije v vzorcu, npr. 45% anketirancev je staro 40 let in več,

Stolpec 3 vsebuje razlike deležev glede na telefon, npr.:

Tabela 2: Razlika deležev D - vpliv A (starosti) na spremenljivko T (telefon)

Starost	Odstotek anketirancev		skupaj
	s telefonom	brez telefona	
40 in več	55.9%	44.1%	100%
pod 40	53.0%	47.0%	100%
razlika D =	<u>2.9%</u>		

Dihotomizacija. Dihotomiziranje ima dobre in slabe strani. Omogoča vključitev večjega števila spremenljivk ter poenostavlja analizo in interpretacijo; izognemo se tudi ordinalnim spremenljivkam. Seveda pa izgubimo informacijo, ki se v podatkih že nahaja. Dihotomizaciji v prid govori tudi dejstvo, da se ordinalne kategorije pri vseh spremenljivkah, razen pri A (starost), obnašajo "linearno"⁹ glede na spremenljivko T,

4) Združene so kategorije: dokončana srednja, višja, visoka šola.

5) Kategorija vsebuje anketirance, ki so pozitivno odgovorili na eno od spodnjih dveh možnosti: Ali bi lahko rekli:

- da vam ničesar posebej ne primanjkuje, da se nič posebej ne omejujete?
- da ste bolj varčni in se omejujete v manj pomembnih stvareh, manj luksuza?

Ostali odgovori (omejitev pri obleki, hrani, ne vem...) tvorijo drugo kategorijo.

6) Združeni sta kategoriji mestna in primestna krajevna skupnost.

7) Kategorija vsebuje anketirance, ki so se eksplicitno opredelili kot ne-verujoči in tudi nikoli ne obiskujejo verskih obredov.

8) Kot rečeno gre za posedovanje telefona v gospodinjstvu.

9) Npr. z naraščanjem stopnje izobrazbe raste tudi delež telefona v posamezni kategoriji.

kot tudi glede na mnenjske spremenljivke. Le spremenljivko A (starost) je smiselno obravnavati tudi v treh kategorijah. Anketiranci v starosti od 30 do 50 let imajo namreč značilno večji odstotek telefona (57.6%) kot kategoriji mlajših oziroma starejših anketirancev (51.8% oziroma 52.0%).

3. BAYESOV INFORMACIJSKI KRITERIJ

Loglinearni modeli omogočajo zapis verjetnosti, da se enota nahaja v določeni celici, kot funkcijo parametrov loglinearnega modela, parametri pa izražajo povezanost (interakcijo) med posameznimi spremenljivkami oziroma njihovimi kategorijami. Dober loglinearni model nam pove, s katerimi povezavami lahko najboljše pojasnimo obstoječe verjetnosti oziroma frekvence. Pri izbiri modela je na voljo več kriterijev¹⁰. Kot rečeno, bomo uporabili Bayesov informacijski kriterij - **BIC** (Raftery, 1986), kar pomeni, da bomo iskali model, ki minimizira izraz:

$$G^2 - df \cdot \ln(n),$$

kjer je n velikost vzorca, df pa stopinje prostosti ustreznega modela. G^2 je definiran z običajnim izrazom

$$G^2 = 2 \cdot \sum_i (f_i \cdot \ln(f_i/e_i)),$$

kjer sumiramo po vseh celicah kontingenčne tabele in so f_i dejanske frekvence, e_i pa odgovarjajoče frekvence, ki izhajajo iz določenega loglinearnega modela¹¹. Uporaba BIC kriterija daje model, ki ima (glede na dane podatke) največjo posteriorno verjetnost, da je pravilen. Ko velikost vzorca n narašča proti neskončnosti, gre verjetnost, da je model pravilen, proti ena. Predpostavka BIC kriterija je, da imamo dovolj velike vzorce in dovolj blage¹² apriorne porazdelitve (Spiegelhalter, Smith, 1983). Operativen kriterij za vključitev določene interakcije v model pa je BIC vrednost določene interakcije

$$D(G^2)/df,$$

10) Npr.: Brown's screening, Akaike information criteria, forward selection, backward selection, Aitkin STP procedure (Agresti, 1990).

11) Običajni kriteriji iščejo model, kjer je vrednost G^2 najbližja stopinjam prostosti, kar izražamo tudi z stopnjo značilnosti p za G^2 , ki se porazdeljuje s hi kvadrat porazdelitvijo. S te plati je najustreznejši model, ki ima $p=0.5$.

12) Neinformativna, nedoločena apriorna porazdelitev (vague prior), ki jo uporabimo, kadar vnaprej o pojavu ne vemo veliko.

kjer $D(G^2)$ označuje spremembo v G^2 , ki nastane zaradi vključitve oziroma izključitve odgovarjajoče interakcije v model. Količnik mora biti večji od $\ln(n)$, v našem primeru je to 7.0 ($\ln 1112$) oziroma 7.6 ($\ln 2048$). Interakcije, ki imajo BIC vrednost manjšo od 7.0 oziroma 7.6, namreč ne pripomorejo k manjšanju BIC kriterija.

Uporaba drugačnega kriterija vodi k drugačnim rešitvam, običajno k bolj kompleksnim modelom, saj ostale metode ne vključujejo velikosti vzorca eksplicitno in so zato nagnjene k iskanju modelov, ki bolj ustrezajo vzorcu, kot pa populaciji. Posebej razvidno je to pri vzorcih z več deset tisoč enotami.

4. MODEL ZA FAKTORJE

V nadaljevanju bomo celoten sklop faktorjev vključili v model, saj nas njihove medsebojne povezave pri analizi vpliva na odvisno spremenljivko ne bodo zanimale (Upton, 1980:84). Kljub temu je potrebno za razumevanje problema njihovo povezanost nekoliko osvetliti. V primeru enostavnega modela, s katerim bi faktorje dobro pojasnjevali, pa je tak model faktorjev primerno uporabiti tudi namesto celotnega sklopa.

Za dober model dvojne interakcije (oziroma dvojne povzave) ne zadoščajo; izkaže se, da potrebujemo dve trojni interakciji:

IC/IG/IR/IE/IA/CG/CR/CE/CA/GR/REA/GEA,
 $G^2=40.0$, 40 df, $p<.472$.

Model je podan v običajni notaciji loglinearnih modelov. Razvidne so interakcije, ki so za pojasnjevanje dejanskih frekvenc najbolj pomembne. Spremenljivke so označene v tabeli 1. Oznaka npr. IC v modelu pomeni, da je povezava dohodek-vernost pomembna za pojasnjevanje problema. Z običajno strategijo bi izbrali zgornji model, saj so stopinje prostosti ravno enake vrednosti G^2 ($df=G^2=40$), in je p (stopnja značilnosti G^2) skoraj enaka 0.5, vendar BIC kriterij odstrani nekatere interakcije, ki so glede na velikost vzorca prešibke. Dobimo:

IE/CR/CE/REA/GEA,
 $G^2=57.9$, 47 df, $p<.133$.

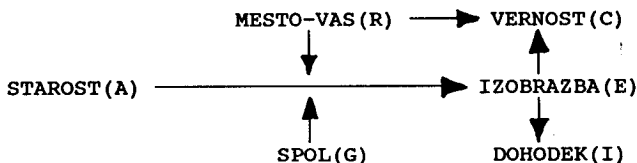
Za pojasnitev povezav med faktorji torej zadoščajo že interakcije IE, CR, CE, REA, GEA. BIC vrednosti zgornjega modela pa so naslednje:

IE(37.7/4.3), CR(72.2/6.2), CE(66.3/6.1), REA(18.4/4.1), GEA(8.6/2.0).

Prva številka v oklepaju pomeni BIC vrednost (ki je v našem primeru značilna¹³, kadar je večji od 7.6), druga pa je standardizirana vrednost odgovorjajočega parametra v loglinearnem modelu (ki je značilna, kadar je večja od 2). Očitno je, da BIC vrednost - na nekoliko drugačni skali - izraža posameznih interakcij.

Če povezavam dodamo vzročno vsebino (kar je vsebinski poseg, ki je lahko tudi arbitraren), lahko model prikažemo kot:

Slika 2: Model sociodemografskih faktorjev



Obe trojni interakciji (na sliki sta označeni kot puščici na povezavo AE) sta vezani na demografske spremenljivke:

- GEA (spol-izobrazba-starost): Mlajše ženske imajo višjo izobrazbo v primerjavi z moškimi, starejše pa precej nižjo.
- REA (ruralnost-izobrazba-starost): V mestu so razlike v izobrazbi med starejšo in mlajšo generacijo razmeroma majhne, v vaških krajevnih skupnostih pa so te razlike izredno velike - če poenostavimo, bi lahko rekli, da so mlajši izobraženi, starejši pa ne.

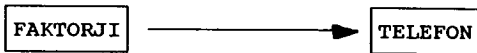
Zgornji interakciji sta precej močni in se ju je koristno zavedati pri vsaki družboslovni analizi na osnovi podatkov SJM (in slovenske populacije sploh), kjer nastopajo spremenljivke G, E, R, A.

¹³) Govorimo o statistični značilnosti. Seveda pa so navedene meje zelo grobe.

5. MODEL ZA FAKTORJE IN SPREMENLJIVKO TELEFON

Analiziramo povezavo med faktorji in telefonom:

Slika 3: Faktorji in telefon



Kot rečeno, je vsebinsko smiselno spremenljivke A,E,R,G,C in I fiksirati kot faktor (slika 3), saj nas njihova medsebojna povezanost ne zanima več, zato jih v celoti vnesemo v model kot zunanjo (eksogeno) interakcijo, ki je ne spreminjamo. Zanima nas samo povezava faktorjev s telefonom, zato je v model vključena eksogena interakcija AERGCi. Izkaže pa se, da daje tudi povezava med A,E,R,G,C,i iz modela v točki 4 skoraj povsem enake rezultate. Zadoščajo že dvojne interakcije

AERGCi/IT/CT/RT/ET/GT/AT,
 $G^2=64.4$, 55 df, $p<0.18$.

Če spol (G) in starost (A), ki nista značilni (BIC vrednosti manj kot 3), izpustimo, dobimo

ERCi/IT/CT/RT/ET,
 $G^2=69.6$, 57 df, $p<0.12$,

kjer lahko pomembnost dvojnih interakcij izrazimo z BIC vrednostmi:

ET(30.2), RT(24.3), IT (21.5), CT (14.1).

Ostale so torej samo povezave telefona (T) z izobrazbo (E), ruralnostjo (R), dohodkom (I) in vernostjo (C). Če vključimo spremenljivko A s tremi kategorijami, pa se pojavi v modelu namesto ET interakcija AET.

Vidimo torej, da je posedovanje telefona odvisno od vseh spremenljivk, ki nastopajo kot faktorji, razen od spola, G. Interakcije so enostavne, bivariatne. Interpretiramo jih lahko s tretjim stolpcem tabele 1. Vsaka mnenjska spremenljivka, ki je močnejše povezana s katerokoli od spremenljivk A,E,R,C,I, ki so v zgornjih modelih povezane s T, je tako potencialni kandidat za izkrivljeno oceno, kadar uporabljamo telefonski vzorec.

Če opravimo regresijo spremenljivke telefon (T) na faktorje, ki nastopajo v zgornjih modelih, so enote (anketiranci) glede na dejansko vrednost spremenljivke T klasificirane¹⁴ pravilno v 60-64% vseh 2048 enot, odvisno od modela. Pri tem so razlike med zgornjimi modeli izredno majhne. Podobne rezultate daje tudi odgovarjajoči logit model, saj se verjetnosti za pozitivne kategorije (tabela 1, kolona 2) gibljejo med 0.25 in 0.75. Faktorji torej razmeroma dobro determinirajo posedovanje telefona.

6. OSNOVNI MODEL

Osnovni model pričujoče analize (slika 1) vključuje še mnenjske spremenljivke. Pričakovati je, da bodo ocene na osnovi telefonskega vzorca napačne pri mnenjskih spremenljivkah, ki so povezane s faktorji A,E,C,I,R (starost, izobrazba, vernost, dohodek, ruralnost). Brez poglobljene analize, ki bi osvetlila osnovne determinante mnenjskih spremenljivk, je v splošnem seveda nemogoče pojasniti, kako telefon "vpliva" na odgovore na mnenjska vprašanja. Nakazuje se sicer, da bodo v telefonskem vzorcu premalo upoštevana mnenja socialno nižjih slojev, vendar bi to težko posplošili. Pokazali smo le, v kakšnih primerih je mnenjska spremenljivka lahko občutljiva na spremenljivko T (telefon). Tako je na osnovi zgornje analize očitno, da so npr. spremenljivke, katerih vrednosti se polarizirajo glede izobrazbe, močno občutljive na telefon, spremenljivke, ki se polarizirajo glede spola, pa ne¹⁵.

Seveda pa je **kljub temu mogoče**, da obstajajo mnenjske spremenljivke, kjer z navedenimi faktorji ne bi mogli pojasniti njihove povezanost s telefonom, vendar pri analizi konkretnih mnenjskih spremenljivk (npr. vsa volilna opredeljevanja v SJM '90) takega primera ni bilo. Glede na visoko stopnjo determiniranosti spremenljivke T (telefon) s faktorji ter glede na (običajno) močno povezanost mnenjskih spremenljivk z istimi faktorji pa obstoj takih spremenljivk tudi ni posebej verjeten. Skupaj z rezultati analize pri konkretnih mnenjskih spremenljivkah govori to v prid navidezni (spurious) povezavi telefon-mnenje, kar smo - kot osnovno hipotezo - opredelili že v uvodu.

14) Spremenljivki T priredimo vrednost ena, kadar je ocenjena regresijska vrednost pri določeni enoti večja ali enaka 0.5 in vrednost nič sicer.

15) Kot rečeno obstajajo v telefonskem anketiranju poleg telefonskega nepokritja tudi drugi problemi. Tako lahko npr. postopek izbire osebe znotraj gospodinjstva povzroči preveliko zastopanost žensk, s čimer lahko postanejo mnenjske spremenljivke pri telefonskem anketiranju občutljive tudi glede spola.

V splošnem lahko rečemo, da je vpliv faktorjev na telefon in na mnenjske spremenljivke različen:

- včasih je vpliv vsakega od faktorjev zanemarljiv,
- včasih se vplivi posameznih faktorjev med sabo izničijo,
- včasih se vplivi kumulirajo.

Čim močnejši je hkraten vpliv faktorjev na spremenljivko T (telefon) ter na mnenjsko spremenljivko, tem močnejša je tudi navidezna povezava med njima. Seveda že med tremi spremenljivkami obstaja - kot je pokazal že Lazarsfeld (1950) - vrsta najrazličnejših povezav in ne le interpretacija (Lazarsfeld: 135-167), ki jo navajamo v osnovnem modelu (slika 1) in tudi v spodnjem primeru. Vendar v naših podatkih ni nobenih znakov, ki bi navajali k drugačnemu tipu medsebojne povezanosti. V nadaljevanju si bomo ogledali tipičen primer spremenljivke, kjer se vplivi faktorjev kumulirajo, zaradi česar je v mnenjih med anketiranci s telefonom in brez njega precejšnja razlika.

7. PRIMER KRAMBERGER

Bivariatne analize. Ilustrirajmo zgornje ugotovitve na primeru spremenljivke K, ki bo pomenila¹⁶ opredeljevanje volilcev za predsedniškega kandidata Krambergerja v drugem pisemskem nadaljevanju SJM '90, marca 1990. Primer je več kot ilustrativen, saj je to spremenljivka, ki je na posedovanje telefona močno občutljiva. Znan je tudi volilni izid, kjer je Kramberger dobil skoraj petino vseh glasov, telefonske ankete v predvolilnem obdobju (SPEM, STIK, Varianta) pa mu v glavnem niso pripisovale več kot 8%. Tako velika razlika je seveda rezultat cele vrste razlogov, ki nas tu pravzaprav ne zanimajo. Zanima nas samo pristranskost, ki nastaja zaradi telefonskega nepokritja in za tak problem so podatki SJM '90 povsem primerni¹⁷. Vpliv telefona prikažemo na naslednji način:

16) Vprašanje V4: Na aprilskih volitvah za mesto predsednika slovenskega predsedstva bodo kandidirali štirje kandidati. Razvrščeni po abecednem redu so to: Marko Demšar, Ivan Kramberger, Milan Kučan in Jože Pučnik. Za koga bi glasovali, če bi bile volitve v tem hipu? (Odgovori: Demšar, Kramberger, Kučan, Pučnik, ne vem, ne bom volil).

17) Zaradi osipa v pisemski anketi in izločitve neopredeljenih anketirancev, razpolagamo samo z 1112 anketiranci.

Tabela 3: Spremenljivki T (Telefon) in K (Kramberger)

Telefon	Volilno opredeljevanje	
	%Kramberger	%ostalo
anketiranci brez telefona	20.8	79.2
anketiranci s telefonom	11.2	88.8
Razlika deležev	D = 9.6	

Iz tabele 3 je razviden statistično značilen vpliv T na K. Tudi enostavne bivariatne analize spremenljivke K in faktorjev kažejo precejšen vpliv faktorjev na spremenljivko K. Če jih izrazimo kot razlike deležev D (ki smo jih opredelili v tabeli 1), dobimo:

Tabela 4: Bivariatna analiza: faktorji - Kramberger

spremenljivka	D	spremenljivka	D
izobrazba (E)	11.3%,	vernost (C)	<u>12.3%</u> ,
dohodek (I)	8.1%,	ruralnost (R)	10.5%,
spol (G)	0.5%,	starost (A)	-3.2%

Videti je, kot da vernost (C) najbolj diskriminira opredeljevanje za Krambergerja, saj se med vernimi za Krambergerja opredeljuje 16.2%, med ostalimi anketiranci pa le 4.5% ($16.2 - 4.5 = 12.3$). O podobnih razmerjih glede pomembnosti posameznih faktorjev govorijo tudi Cramerjevi in kontingenčni koeficienti odgovarjajočih dvorazsežnih tabel. Faktorji torej delujejo na spremenljivko K podobno kot na spremenljivko T, zato je pričakovati močno navidezno povezanost, saj se vplivi kumulirajo. V nasprotno smer deluje le spremenljivka A (starost). Drugače kot pri telefonu (tabela 1), kjer imajo mlajši anketiranci manj telefonov, se namreč pri spremenljivki K mlajši anketiranci v večji meri opredeljujejo za Krambergerja (starost deluje negativno, $D = -3.2\%$). Seveda pa je razlika 3.2% na meji statistične značilnosti.

Model za spremenljivko K (Kramberger). Oglejmo si najprej model brez telefona:

AERGC1/KA/KE/KI/KC/KR,
 $G^2 = 56.8$, 56 df, $p < 0.45$,

kjer so vključene vse spremenljivke, razen spola, G. Vse so približno enako pomembne (BIC vrednost je 10), razen KE (Kramberger-izobrazba), ki ima BIC vrednost 17. **Rezultati so izredno poučni**, saj znatno odstopajo od videza, ki ga daje bivariatna analiza v tabeli 4. Model namreč kaže, da je E (izobrazba) bistveno

močnejši faktor za spremenljivko K (Kramberger) kot C (vernost). Vidimo tudi, da je spremenljivka K povezana z večino faktorjev modelu. Ker velja podobno tudi za T, je na videz močna povezanost TK razumljiva.

Osnovni model. Vključimo med faktorje še spremenljivko T. Če so vključeni v model eksogeni faktorji T,R,E,A,I,C, imamo

TREAIC/KR/KA/KE/KI/KC,
 $G^2=58.5$, 57 df, $p<0.42$.

Zveza TK (telefon-Kramberger) ni značilna in izpade iz modela, BIC vrednosti je 3. Kot je bilo pričakovati, samo posedovanje telefona ne vpliva na mnenjsko spremenljivko. Zveza KT iz tabele 3 je bila torej navidezna. Preostale dvojne interakcije so enako pomembne, nekoliko izstopa KE (Kramberger-izobrazba), ki ima BIC vrednost 14, ostali pa okoli 10. Če izločamo spremenljivke iz modela, potem pomembnost KT (Kramberger-telefon) postopoma narašča in se npr. pri spremenljivkah KTRE in modelu TRE/KT/KR/KE pojavi z BIC vrednostjo 10. **Kadar je v modelu premalo faktorjev, torej postane povezava KT značilna.**

8. INTERPRETACIJE

Parametri loglinearnega modela. Zgornje rezultate je mogoče natančneje interpretirati tudi s poglobljeno analizo parametrov loglinearnega modela. Vendar med interpretacijo BIC vrednosti ter interpretacijo parametrov loglinearnega modela v splošnem ni večjih razlik, kar smo pokazali tudi na primeru v točki 4. Tako tudi natančna analiza vrednosti parametrov za zgornje modele - kot tudi za odgovarjajoče saturirane modele - ne prinaša nič posebej novega, zato je tudi ne navajamo.

Razlike deležev. Podobna, vendar bolj nazorna, je interpretacija na osnovi razlik odstotnih deležev (Vehovar, 1989). Pri uporabi parcialnih razlik deležev, kjer kontroliramo vplive vseh faktorjev, se nazorno pokaže, kako navidezna je povezava KT (Kramberger-telefon): ko kontroliramo vse faktorje, postanejo razlike deležev v spremenljivki K (zaradi telefona) neznačilne. Še nazornejši je postopek standardizacije, ki podeli obema podvzorcema (anketiranci s telefonom in brez njega) enako socio-demografsko strukturo (to je, povprečno strukturo v vzorcu), zaradi česar razlike v opredeljevanju za Krambergerja skoraj v celoti izginejo.

Ponderiranje. Tudi z različnimi postopki ponderiranja (cenilka na osnovi ocenjenih deležev, "raking" cenilka, poststratifikacija¹⁸), ki jih opravimo na podvzorcju an-

ketirancev s telefoni, se kažejo podobni rezultati¹⁹. Ocena, ki jo dobimo s postopki ponderiranja, se močno približa vrednosti v celotnem vzorcu, kar potrjuje, da je bila povezava KT (Kramberger-telefon) navidezna. Zgornje metode je seveda mogoče uporabiti tudi pri korekciji rezultatov, ki temeljijo na telefonskih vzorcih, vendar pri tem nastopi vrsta praktičnih težav, ki nekoliko zmanjšujejo domet tovrstnim korekcijam.

9. SKLEP

Povzemimo še enkrat najpomembnejše ugotovitve:

- Povezava telefona z mnenjskim spremenljivkami je izredno močna. V SJM '90 obstaja več kot 100 (!) mnenjskih spremenljivk oziroma njihovih kategorij, pri katerih je razlika deležev - pri anketirancih s telefonom in brez njega - večja kot pri spremenljivki K (Kramberger), kjer je bila razlika deležev $D = 20.8\% - 11.2\% = 9.6\%$.
- Vpliv sociodemografskih spremenljivk na telefon je razmeroma jasen. Za posedovanje telefona je najpomembnejša spremenljivka izobrazba (E), nekoliko izstopa tudi ruralnost (R). Vernost (C) in dohodek (I) sta približno enako pomembna, vloga A (starost) pa je šibka in odvisna od dihotomizacije.
- Vsi rezultati govorijo v prid osnovni hipotezi, da je zveza mnenjskih spremenljivk s T (telefonom) navidezna. Seveda izraz "navidezna" (spurious) pomeni le to, da je povezava nastala zaradi določenih faktorjev, nikakor pa ne zanika samega obstoja povezave. Za ocene na osnovi telefonskih vzorcev so zato posebej občutljive mnenjske spremenljivke, ki so močnejše povezane z zgornjimi faktorji, predvsem z izobrazbo. Navidezna povezava obstaja vse dotlej, dokler v model ne vključimo dovolj faktorjev - socio-demografskih spremenljivk, ki so povezavo pravzaprav ustvarili. Navedeni faktorji pa predstavljajo tudi dobro izhodišče za odpravljanje pristranskosti, ki nastaja zaradi telefonskega nepokritja.

18) Poststratifikacija uporablja za izračun uteži populacijske vrednosti, "raking" cenilka le populacijske margine, cenilka na osnovi ocenjenih deležev pa temelji na vzorčni strukturi anketiranih enot.

19) Natančnejše navajanje rezultatov presega zastavljeni okvir pričujoče obravnave.

LITERATURA

Agresti A (1990): *Categorical data analysis*, New York, Wiley.

Blejec M (1970): *Slovensko javno mnenje*, Center za raziskavo javnega mnenja in množičnih komunikacij, VŠSPN, Ljubljana.

Davis J (1990): *CHIPpendale - Software for empirical model building for social sciences*, True Basic Ltd, Hannover, USA.

Lazarsfeld P (1950): *Problems of survey analysis*, Free Press.

McDonald K (1990): *MICLOG*, Computer package for loglinear analysis, Cambridge, UK.

Raftery A (1986): *A note on Bayesian Information Criteria*, *Journal of the Royal Statistical Association*, B48 (1986), 249-250.

Spiegelhalter D, Smith A (1983): *Bayes Factors for Linear and Loglinear Models with Vague Prior Information*, *Journal of the Royal Statistical Association*, B44 (1983), 377-387.

Upton G (1980): *The Analysis of Cross-tabulated Data*, New York, Wiley.

Vehovar V (1989): *Analiza kontingenčnih tabel na osnovi razlike strukturnih deležev*, Blejsko metodološko srečanje '89, Ljubljana, RI FSPN.