

Pregled izbranih kvantitativnih metod



Metode triangulacije



With the support of the
Erasmus+ Programme
of the European Union

prof. dr. Matej Makarovič
Matej.makarovic@fuds.si

Možnost za opravljanje obveznosti za Metode triangulacije

- Izdelava več krajših izdekov, povezanih z raziskovalnimi metodami: po dogovoru – interes sporočite na e-mail: matej.makarovic@fuds.si
- Izdelava raziskovalne naloge po zgledu znanstvenega članka v obsegu cca. 5000 besed, ki vključuje smiselno uporabo vsaj ene multivariatne statistične metode

Načrt predavanja

- Kdaj uporabiti kvantitativne metode
- Enote, spremenljivke, populacije in vzorci
- Univariatne statistike
 - Aritmetična sredina, mediana, modus, delež
 - Sklepanje iz vzorca na populacijo
- Bivariatne statistike
 - χ -kvadrat, Cramerjev V, korelacije, t-test, analiza variance
- Multivariatne statistike
 - Faktorska analiza, analiza klastrov, regresijska analiza, analiza poti in modeliranje strukturnih enačb

Kako izbrati metodo raziskovanja?

- Izbor metode mora izhajati iz raziskovalnega problema: izberemo takšno metodo, za katero sklepamo, da nam lahko pomaga odgovoriti na naša raziskovalna vprašanja in hipoteze
- Koliko ustrezna je metoda z vidika
 - Zanesljivosti
 - Veljavnosti
 - Praktičnosti

Zanesljivost kvantitativnih metod

- Kvantitativne metode so, če so pravilno izpeljane, najboljše jamstvo za zanesljivost, tj. da ponovitve raziskave pripeljejo do vedno enakih rezultatov
- Omogočajo nam zanesljivo posploševanje na podlagi vzorcev, ki so bistveno manjši od populacij
- Omogočajo nam, da z natančno opredeljeno verjetnostjo potrdimo ali zavrnemo naše hipoteze
- Sociologijo v zanesljivosti postavljajo na enako raven, kot velja za „trde“ znanosti

Veljavnost kvantitativnih metod

- Kvantitativne metode imajo lahko omejitve na ravni veljavnosti, tj. ali res merijo tisto, kar hočemo izmeriti
- Anketni podatki in uradne statistike ne omogočajo velike poglobljenosti: kompleksne dileme reducirajo na preprosta vprašanja in odgovore; ne nudijo veliko možnosti za izogibanje nesporazumov
- Številčno izraženi rezultati dajejo vtis objektivnosti, ki pa je lahko zaradi omejitev v veljavnosti sporna

Praktičnost kvantitativnih metod

- Izjemno praktične, če imamo na voljo že izdelane podatkovne baze in nam zadostuje sekundarna analiza
- Deloma praktične, če imamo opravka s populacijami, ki so zlahka dostopne in niso zahtevne za vzorčenje
- Nepraktične, če imamo opravka s težko dostopnimi populacijami, zahtevnimi za vzorčenje

Primeri sekundarnih baz podatkov

- Eurobarometer
- European Values Study
- European Social Survey
- World Values Survey
- International Social Survey Programme
- Slovensko javno mnenje
- Slovenski utrip
- Viri:
 - GESIS
 - Arhiv družboslovnih podatkov

Enote proučevanja

- Enote so tisto, kar v raziskovanju neposredno proučujemo
 - Posamezniki, organizacije, regije, države...
- Za kvantitativno raziskovanje se odločimo, ko imamo opravka z velikim številom enot
 - Kaj je dovolj veliko, je odvisno od konteksta in od statističnih metod, ki jih želimo uporabiti (30 ali več, nekaj sto, prek tisoč enot)

Populacija in vzorec

- Vse enote, o katerih želimo sklepati na podlagi našega raziskovanja, tvorijo populacijo (N)
- Nujno se moremo vnaprej odločiti, na katero populacijo se bo nanašalo naše sklepanje:
 - npr. populacija evropskih držav, populacija evropskih državljanov, populacija slovenskih podjetij, populacija slovenskih študentov, populacija študentov FUDŠ...
- Večinoma iz populacije izberemo vzorec, na katerem dejansko izvajamo našo analizo
- Pravilna izbira vzorca je mogoča le, če smo pred tem povsem jasno opredelili populacijo

Vzorci: kako jih izbrati

- Slučajno vzorčenje: vsaka enota ima enako verjetnost, da bo izbrana; praviloma je nujen seznam enot v populaciji
 - Preprosto slučajno vzorčenje
 - Sistematično slučajno vzorčenje (vsaka k-ta enota)
 - Stratificirano vzorčenje (podvzorci po kategorijah)
- Neslučajno vzorčenje: seznam enot v populaciji ni potreben
 - Priročno vzorčenje, snežena kepa ipd.
- Statistično sklepanje iz vzorca na populacijo je mogoče le pri slučajnem vzorčenju

Spremenljivke

- Enote imajo določene lastnosti, ki jih lahko izrazimo z vrednostmi: izrazimo jih kot spremenljivke
- Ista enota ima lahko na isti spremenljivki natanko eno vrednost

	Spremenljivka 1	Spremenljivka 2	Spremenljivka 3
ENOTA 1	Vrednost 1-1	Vrednost 1-2	Vrednost 1-3
ENOTA 2	Vrednost 2-1	Vrednost 2-2	Vrednost 2-3
ENOTA 3	Vrednost 3-1	Vrednost 3-2	Vrednost 3-3

Vrste spremenljivk glede na merske lestvice

- Nominalne
 - Lastnosti, ki jih ni mogoče kvantificirati
 - Spol („ženska“, „moški“); državljanstvo...
- Ordinalne
 - Lastnosti, ki jih lahko delno kvantificiramo
 - Stopnja izobrazbe („osnovna ali manj“, „poklicna“, „srednja“, „višja, visoka ali več“); šolske ocene...
 - Nekatere lahko obravnavamo (skoraj) kot intervalne
- Intervalne in razmernostne
 - Povsem kvantitativne lastnosti s teoretično neomejenim številom možnih številskih vrednosti;
 - Mesečni dohodek v evrih; starost matere ob rojstvu prvega otroka
- Dihotomne
 - Omogočajo, da se nominalna kategorija obnaša podobno kot intervalna spremenljivka
 - Le dve možni vrednosti (1 ali 0)

S čim računamo statistike

- Kalkulatorji le za res najbolj preproste izračune
- Excel in podobni tabelarni programi:
 - Uporabno za pripravo podatkovnih baz
 - Praktično za univariatne metode, deloma tudi bivariatne, vdelane formule, dobre grafične možnosti, vendar nepraktično ali neuporabno za multivariatne metode
- Specializirani statistični programi:
 - R: prosto dostopno, relativno zahtevno, splošna statistika, ogromno možnosti
 - SPSS, Stata: drago, bolj prilagojeno družboslovcem
 - PSPP: prosto dostopno, po zgledu SPSS, manj možnosti,

Univariatne statistike na vzorcu

- Opredelili smo populacijo
- Iz nje smo izbrali slučajni vzorec enot
- Za vsako enoto v vzorcu imamo vrednosti spremenljivk
- Kako lahko tak vzorec najbolje opišemo?

Srednja vrednost: aritmetična sredina ali povprečje

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$

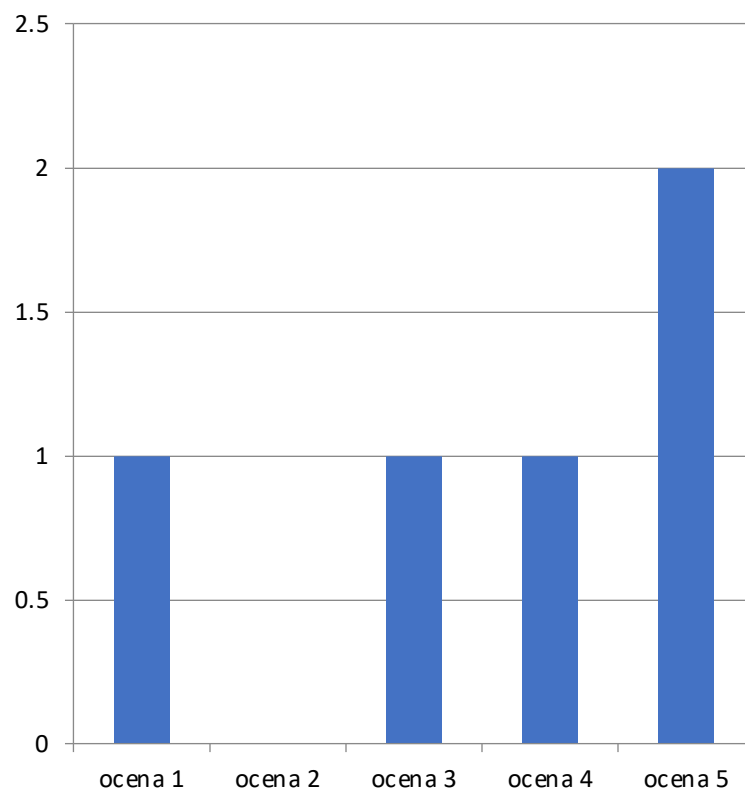
Ime	Ocena
Mia	4
Toni	3
Karin	5
Miha	5
Tina	1

Pogoj: zahteva intervalne spremenljivke

$$\bar{x} = \frac{4 + 3 + 5 + 5 + 1}{5} = 3,6$$

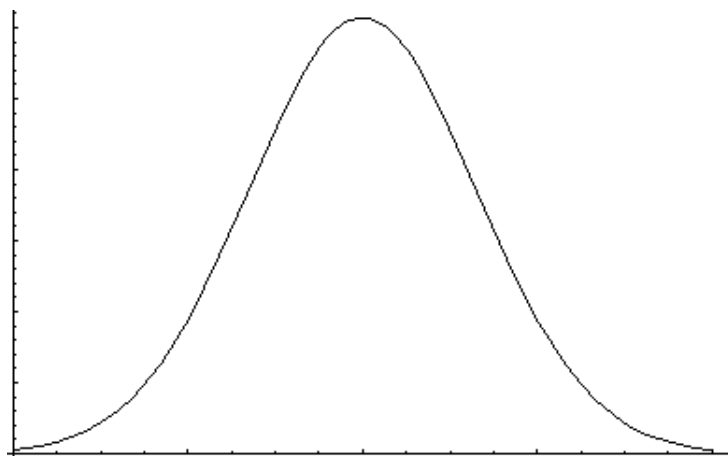
Frekvenčna porazdelitev

Ime	Ocena
Mia	4
Toni	3
Karin	5
Miha	5
Tina	1

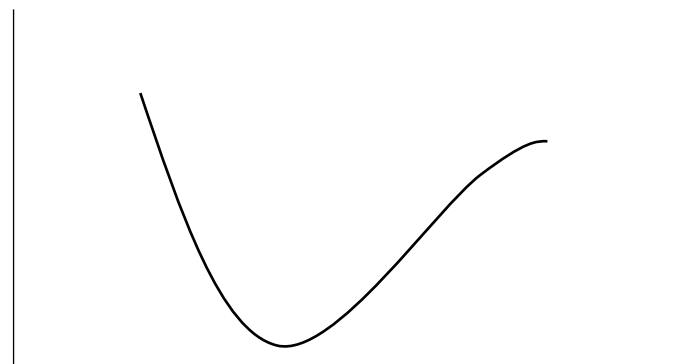
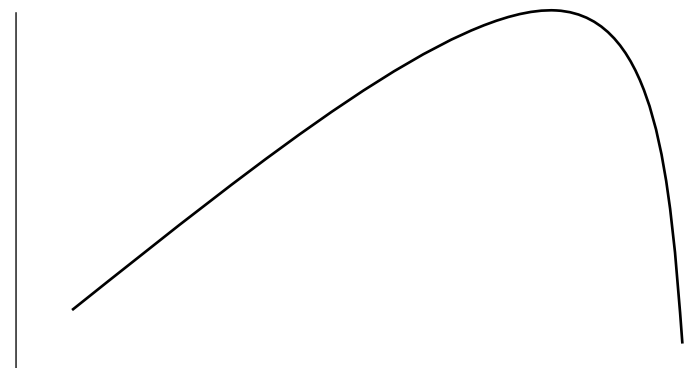


Histogram frekvenčne porazdelitve

Srednja vrednost ne pove vsega: različne porazdelitve



Normalna porazdelitev



Mediana in modus kot srednji vrednosti

- Mediana
 - vrednost, ki se uvrsti na sredino, ko enote rangiramo glede na vrednost
 - Pogoji: najmanj ordinalne spremenljivke
- Modus
 - vrednost, ki se najpogosteje pojavi
 - Uporaben pri vseh vrstah spremenljivk
- Pri povsem normalni porazdelitvi so aritmetična sredina, mediana in modus enaki

Primerjava srednjih vrednosti

- Aritmetična sredina = 3,6
- Mediana = 4
- Modus = 5
- Pri povsem normalni porazdelitvi bi bili aritmetična sredina, mediana in modus enaki

Ime	Ocena
Mia	4
Toni	3
Karin	5
Miha	5
Tina	1

Odstopanje od srednje vrednosti: standardni odklon

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Ime	Ocena
Mia	4
Toni	3
Karin	5
Miha	5
Tina	1

$$s = \sqrt{\frac{(4-3,6)^2 + (3-3,6)^2 + (5-3,6)^2 + (5-3,6)^2 + (1-3,6)^2}{5-1}} = 1,67$$

Standardni odklon in varianca

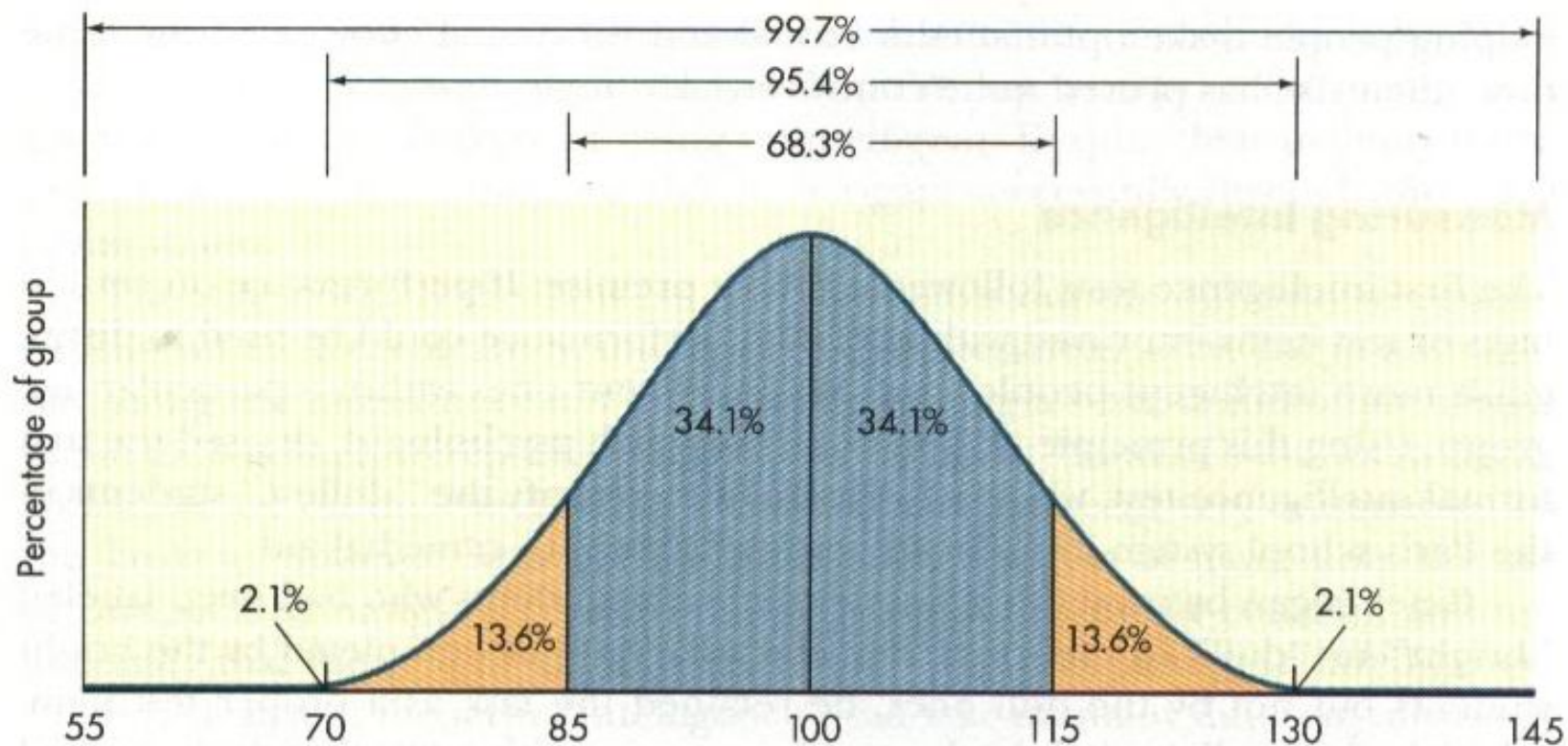
$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Standardni odklon

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Varianca

Normalna porazdelitev in standardni odklon



Možni opisi vzorca glede na merske lestvice spremenljivk

Tipi spremenljivk	Frekvenčna porazdelitev	Aritmetična sredina	Mediana	Modus	Standardni odklon, varianca
Nominalne	✓			✓	
Ordinalne	✓	Kadar so podobne intervalnim	Kadar so podobne intervalnim	✓	Kadar so podobne intervalnim
Intervalne	✓ Če je vrednosti veliko, oblikujemo kategorije	✓	✓	✓	✓

Sklepanje iz vzorca na populacijo

- Imamo izračunane statistike za vzorec – za potrebe naše raziskave pa nas zanima predvsem populacija
- V kolikšni meri statistike, ki smo jih izračunali za vzorec, veljajo za celotno populacijo?
- Na ravni vzorca imamo statistike $(\bar{x}, s, s^2)$, ki jih lahko izračunamo; na ravni populacije imamo ustrezne parametre (μ, σ, σ^2) , na katere lahko iz tega sklepamo, a jih ne moremo direktno izračunati

Matematika in verjetnost

- Pomaga nam centralni limitni izrek:
- pri dovolj velikih vzorcih velja, da je porazdelitev vzorčenja aritmetične sredine vzorca ($\bar{x}$) približno normalna z aritmetično sredino μ in standardnim odklonom $\sigma/\sqrt{n}$

Praktični učinek

- S pomočjo aritmetične sredine vzorca lahko ugotovimo, v katerem intervalu se z določeno verjetnostjo nahaja resnična aritmetična sredina populacije.

- Interval zaupanja: $\pm z \frac{s}{\sqrt{n}}$; $s = \hat{\sigma} \cong \sigma$

- Interval zaupanja za delež π :

$$\pm z \sqrt{\frac{\pi(1-\pi)}{n}}$$

Koeficienti zaupanja in vrednosti z

Koeficient zaupanja (verjetnost, da interval zaupanja za statistiko vzorca dejansko vsebuje vrednost parametra populacije)	Ustrezajoča vrednost z
90%	1,65
95%	1,96
98%	2,33
99%	2,58

1. Areas under the Normal Distribution

The table gives the cumulative probability
up to the standardised normal value z
i.e.

$$P[Z < z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2) dz$$

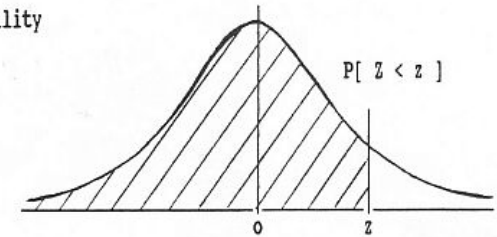


Tabela z-vrednosti

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

Testiranje hipotez

- Hipoteze je z logičnega vidika lažje ovreči kot potrditi
- Izhodišče: postavimo ničelno hipotezo H_0
 - Npr. „povprečni oceni v dveh vzorcih sta med seboj enaki“; „deleža odličnih ocen v obeh vzorcih sta enaka“
- Nato postavimo alternativno (raziskovalno) hipotezo H_a , ki je nasprotna ničelni
- Testiramo ničelno hipotezo

Test ničelne hipoteze za enakost deležev v dveh vzorcih

- Ničelna hipoteza $H_0: \pi_1 = \pi_2$
- Raziskovalna hipoteza $H_a: \pi_1 \neq \pi_2$
- Testiramo ničelno hipotezo

$$Z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}}$$

- Vrednost verjetnosti vezano na z pomnožimo z 2, če preverjamo obe možnosti, tj. da je π_1 večji ali manjši od π_2), sicer to ni potrebno
- Nato poiščemo ustrezno vrednost za signifikanco.

Interpretacija signifikance

- Signifikanca pomeni verjetnost, da bi dobili rezultate, kot smo jih dobili, če bi bila ničelna hipoteza resnična.
- Kolikšno verjetnost lahko še toleriramo, je odvisno od naših raziskovalnih potreb.
- Pogosto uveljavljena: 0,05 oziroma 5% verjetnosti kot zgornja meja, pri kateri je še mogoče zavrniti ničelno hipotezo in na tej podlagi sklepati, da je raziskovalna hipoteza resnična.

Bivariatne statistike

- Univariatne statistike se nanašajo na eno spremenljivko: povprečja, standardni odkloni, mediane, modusi, frekvence nam omogočajo *opisovanje* vzorca in prek tega *opisovanje* populacije
- S primerjanjem lastnosti različnih vzorcev že stopamo na področje bivariatnih statistik
- Bivariatne statistike se nanašajo na razmerje med dvema spremenljivkama: omogočajo nam že osnovno sklepanje o zakonitostih v odnosi med družbenimi pojavi.

Vzročno-posledična razmerja

- Največkrat gre za sklepanje o vzročno-posledičnih zvezah med pojavi ali vsaj o tem, da sta pojava nekako povezana
- Neodvisna spremenljivka → odvisna spremenljivka
- Statistika nam ne more odgovoriti na vprašanje, kaj je vzrok in kaj posledica: za to potrebujemo teorijo (ali vsaj zdravo pamet)

Dve nominalni spremenljivki: χ^2 (hi-kvadrat)

- χ^2 uporabljamo, ko imamo opravka z dvema nominalnima spremenljivkama
- Oblikujemo tabelo, v kateri pokažemo, kako se vrednosti odvisne spremenljivke razlikujejo glede na vrednosti neodvisne spremenljivke
- Izračuna se iz kvadratov razlik med pričakovanimi in dejanskimi vrednostmi v tabeli

Primer tabele

	Moški	Ženske	Skupaj
Zadovoljni	50	40	90
Delno zadovoljni	200	180	380
Nezadovoljni	50	40	90

	Moški	Ženske	Skupaj
Zadovoljni	90	0	90
Delno zadovoljni	130	250	380
Nezadovoljni	80	10	90

Interpretacija rezultatov

- Preverimo, ali je vzorec dovolj velik za statistično sklepanje: vsaj 80 odstotkov celic mora imeti pričakovano frekvenco višjo od 5
- Preverimo signifikanco: če ni višja od npr. 0,05 lahko zavrnilo ničelno hipotezo in sklepamo o pravilnosti raziskovalne hipoteze, tj. da povezanost med neodvisno spremenljivko in odvisno spremenljivko v populaciji obstaja
- Preverimo vrednost katerega od koeficientov, ki nam na primerljiv način pokaže intenzivnost povezanosti, npr. Cramerjev V ($0 \leq V \leq 1$)
 - $V = 0 \Rightarrow$ *ni povezanosti*
 - $V = 1 \Rightarrow$ *popolna povezanost*
 - Vzročno-posledična smer ni pomembna

Dve ordinalni spremenljivki: Kendallov tau-b in gama

- $-1 \leq \tau_B \leq 1$
- $-1 \leq \gamma \leq 1$
- Interpretacija:
 - Smer vzročnosti ni pomembna
 - -1 = popolna negativna povezanost
 - 0 = ni povezanosti
 - 1 = popolna pozitivna povezanost
- Prav tako moramo upoštevati signifikanco

Primer

Kakšna je stopnja vaše izobrazbe? * Starost po kategorijah Crosstabulation

Count		Starost po kategorijah			Total
		18-34 let	35-54 let	55 in več	
Kakšna je stopnja vaše izobrazbe?	Osnovna ali manj	17	62	138	217
	Poklicna	70	129	77	276
	Srednja	122	93	71	286
	Višja, visoka ali več	56	57	23	136
Total		265	341	309	915

$$\tau_B = -0,309$$

$$\gamma = -0,432$$

Dve intervalni spremenljivki: korelacijski koeficienti

- Za povezanost med intervalnima spremenljivkama je najbolj široko uporabljen Pearsonov korelacijski koeficient

$$r = \sum_{i=1}^n \frac{(xi - \bar{x})(yi - \bar{y})}{(n - 1)s_x s_y}$$

- $-1 \leq r \leq 1$; kontroliramo tudi signifikanco
- Interpretacija: smer vzročnosti ni pomembna
 - -1 = popolna negativna povezanost
 - 0 = ni povezanosti
 - 1 = popolna pozitivna povezanost
- Alternativa: Spearmanov korelacijski koeficient

Nominalna z 2 kategorijama in intervalna spremenljivka: T-test

- Kako pripadnost eni od dveh kategorij vpliva na lastnost, ki jo lahko merimo intervalno (npr. spol in število točk na maturi)
- Kontroliramo ali sta varianci v obeh skupinah enaki (F-test)
- Glede na to izberemo ustrezno vrednost t-statistike in signifikance
- Ničelno hipotezo pravilo zavrnemo če:
 - $T \geq 2$
 - $P \leq 0,05$

Ali so razlike med več skupinami statistično pomembne: Analiza variance

- Kako pripadnost posamezni kategoriji vpliva na lastnost, ki jo lahko merimo intervalno (npr. regija in število točk na maturi)
- F-test nam pokaže, ali so razlike v varianci statistično pomembne
- Dodatni testi nam omogočajo primerjavo med posameznimi pari kategorij (LSD, Scheffe, Duncan...)

Omejitev bivariatnih statistik

- Lahko nam kažejo navidezno vzročno povezanost med pojavi (npr. pogostost uporabe interneta in pričakovana življenjska doba)
- Realnost je bistveno bolj zapletena: v model moramo vključiti več kot dve spremenljivki
- To omogočajo multivariatne metode

Primeri pogosto uporabljenih multivariatnih metod

- Faktorska analiza
- Analiza klastrov
- Regresijska analiza in njene izpeljave

Faktorska analiza

- Problem: imamo veliko intervalnih spremenljivk, med katerimi so mnoge v medsebojnih korelacijah: morda jih več meri isto ali podobno stvar
- Iz večjega števila dejansko izmerjenih (tj. manifestnih) spremenljivk želimo dobiti latentne spremenljivke v ozadju, ki pojasnjujejo korelacije med našimi manifestnimi spremenljivkami
 - Eksploratorni pristop: analiziramo naše podatke brez vnaprejšnjih predpostavk
 - Konfirmatorični pristop: imamo teoretski model, ki ga želimo preveriti

Glavni koraki v faktorski analizi

- Imamo standardizirane manifestne spremenljivke ($\bar{x} = 0; s = 1$)
- Oblikujemo korelacijsko matriko med njimi
- Izberemo metodo ekstrakcije faktorjev (glavne komponente, najmanjši kvadrati...)
- Odločitev o številu faktorjev, ki jih bomo upoštevali (eigenvalue > 1) – pomembno koliko variance pojasnjuje posamezni faktor in koliko vsi skupaj
- Odločitev o rotaciji, ki ustvari preglednejšo sliko
 - Pravokotna (ortogonalna) rotacija, kjer ne bo korelacije med faktorji: npr. varimax
 - Nepravokotna rotacija: npr. oblimin
- Interpretacija, poimenovanje pridobljenih faktorjev
- Za vse enote v vzorcu lahko nato shranimo vrednosti na posameznih faktorjih in jih lahko uporabljamo v nadaljnjih analizah

Rotated Component Matriã

	Component				
	1	2	3	4	5
svoboda	,702	,110			
odgovornost	,792				
enakost	,785			,183	
tekmovalnost			,259	,368	,312
ãlovekove pravice	,754	,192		,230	
red	,721	,130			,201
varnost	,678	,238			
kapitalizem			,570	,530	-,142
demokracija	,371	,486		,517	
socializem		,165		,292	,740
liberalnost	,217	,271		,710	,265
konservativnost	-,123		,649	,293	,156
narod	,450	,364	,192	-,303	,348
država	,236	,702	,119		,139
religija		,394	,701	-,103	-,108
patriotizem	,118	,541	,194		,216
odprtost v svet	,451	,502			
evropejstvo		,764		,179	
multikulturalnost	,190	,555		,255	,136
partizanstvo					,787
domobranstvo			,728		

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 8 iterations.

Oblikovanje indeksov

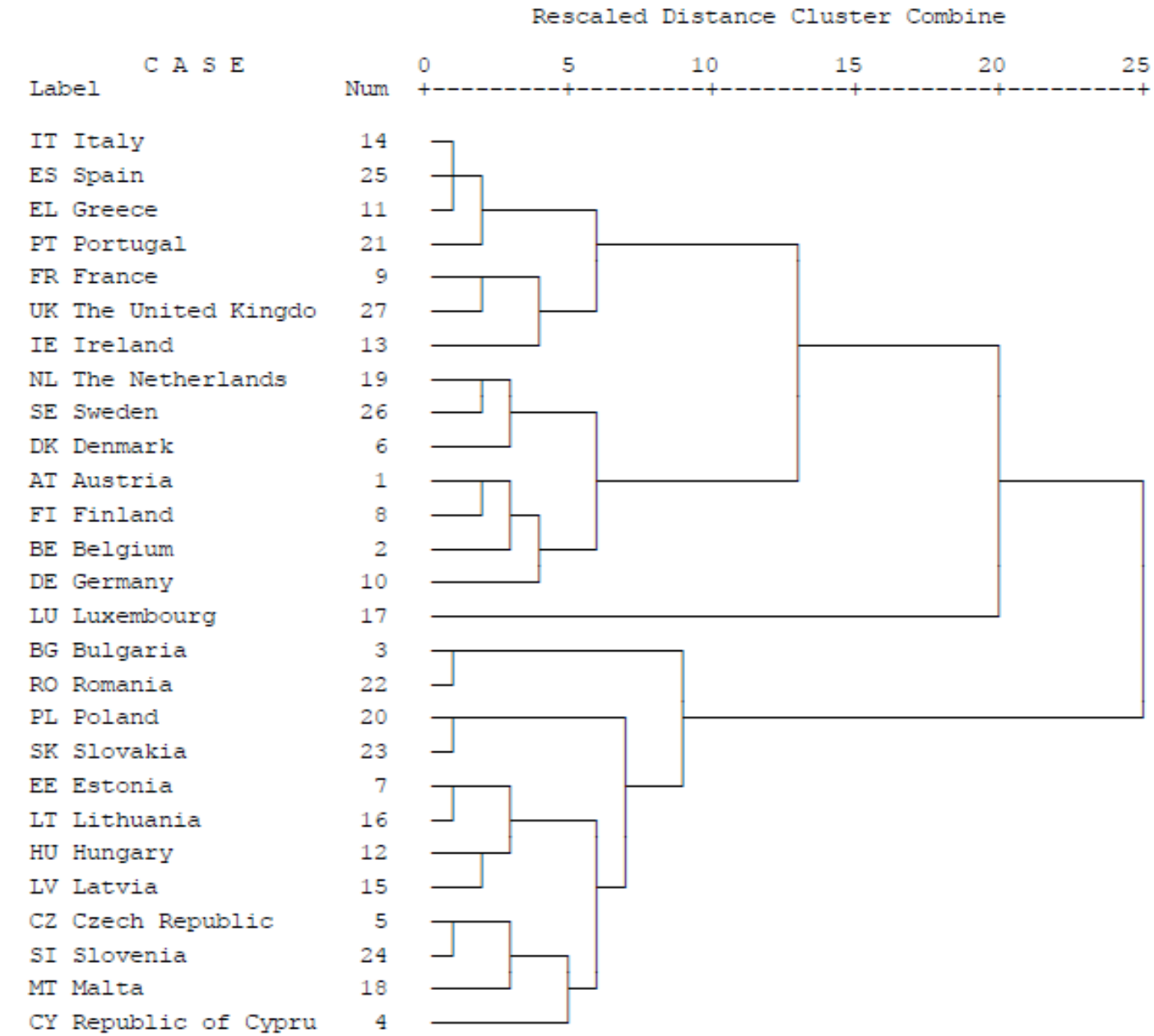
- Več medsebojno povezanih spremenljivk, ki prispevajo k merjenju istega koncepta, lahko uporabimo za tvojenje indeksa: npr. kot indiktor določenega teoretskega koncepta
- Nujnost: indeks mora biti enodimenziionalen
- Test: faktorska analiza ali Cronbachov alfa

Analiza klastrov

- Problem: imamo manjše število enot, ki jih opisuje veliko število spremenljivk – katere enote bi lahko kategorizirali glede na to, katera je kateri bližja (npr. evropske države)
- Uporabimo hierarhično analizo klastrov
- Pregleden rezultat je dendrogram, ki nam pokaže, katere enote so si med seboj blizu
- Pogoji: spremenljivke so intervalne in standarizirane

***** H I E R A R C H I C A L C L U S T E R A N A L Y S I S *****

Dendrogram using Ward Method



Druge možnosti uporabe analize klastrov

- Nehierarhična klaster analiza:
 - Veliko število enot
 - Število klastrov opredelimo vnaprej
- Hierarhični klastri spremenljivk namesto klastrov enot

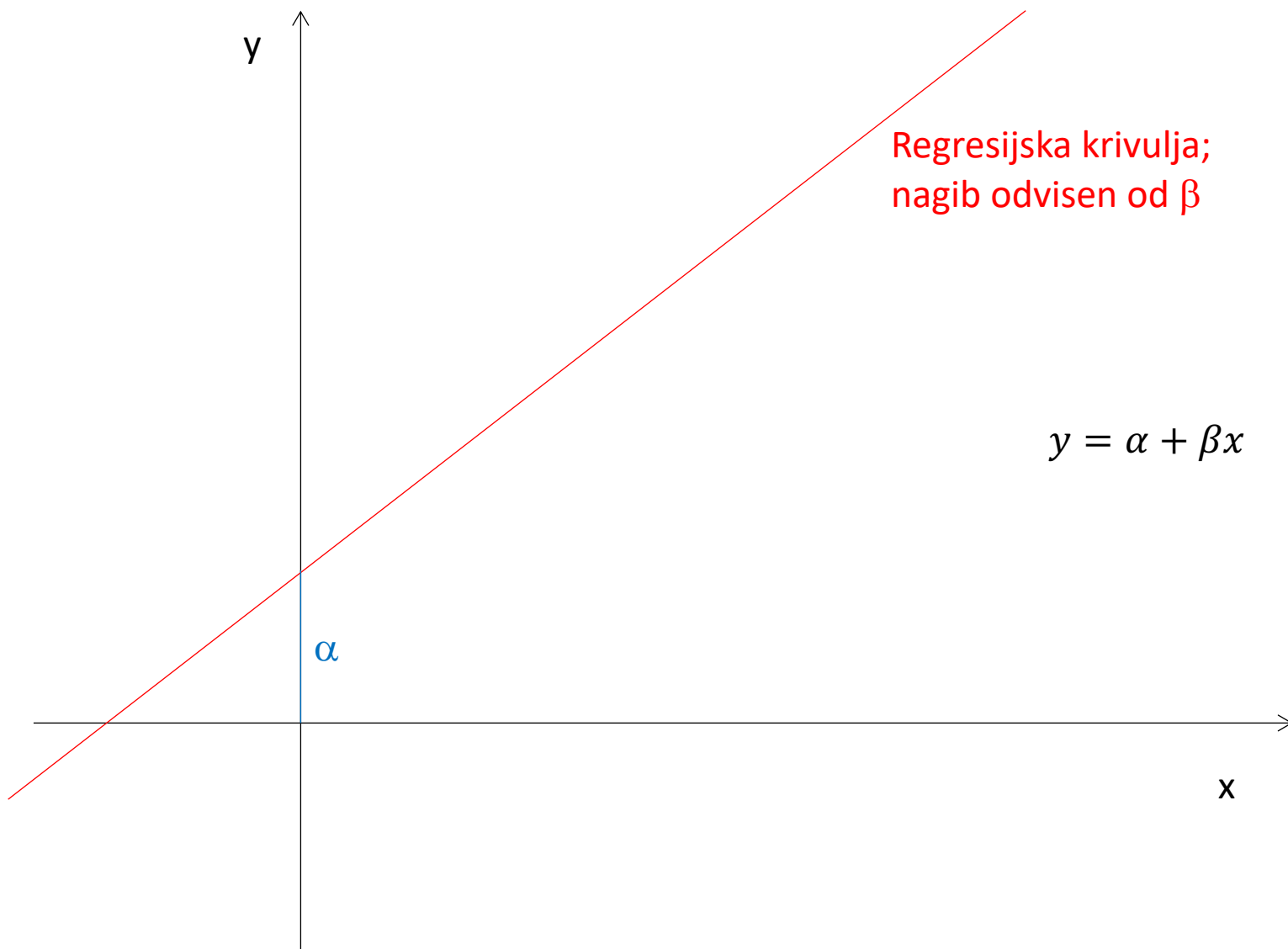
Regresijska analiza

- Problem: imamo veliko število neodvisnih spremenljivk, za katera domnevamo, da vplivajo na odvisno spremenljivko

- Osnovna oblika: linearna regresijska enačba

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots + \beta_n x_n$$

- α je intercept, β_1 do β_n so regresijski koeficienti



Pomembne vrednosti

- B_i – standardizirana vrednost β_i nam pove, kolikšen je relativen učinek posamezne neodvisne spremenljivke x_i na odvisno spremenljivko y
- Vsak koeficient je opremljen tudi s podatkom o signifikanci, ki nam pokaže ali dejansko ima statistično relevanten učinek
- R^2 determinacijski koeficient nam pove kolikšen delež variance y pojasnjuje naš celoten regresijski model

Nekaj omejitev in rešitev glede vrst spremenljivk

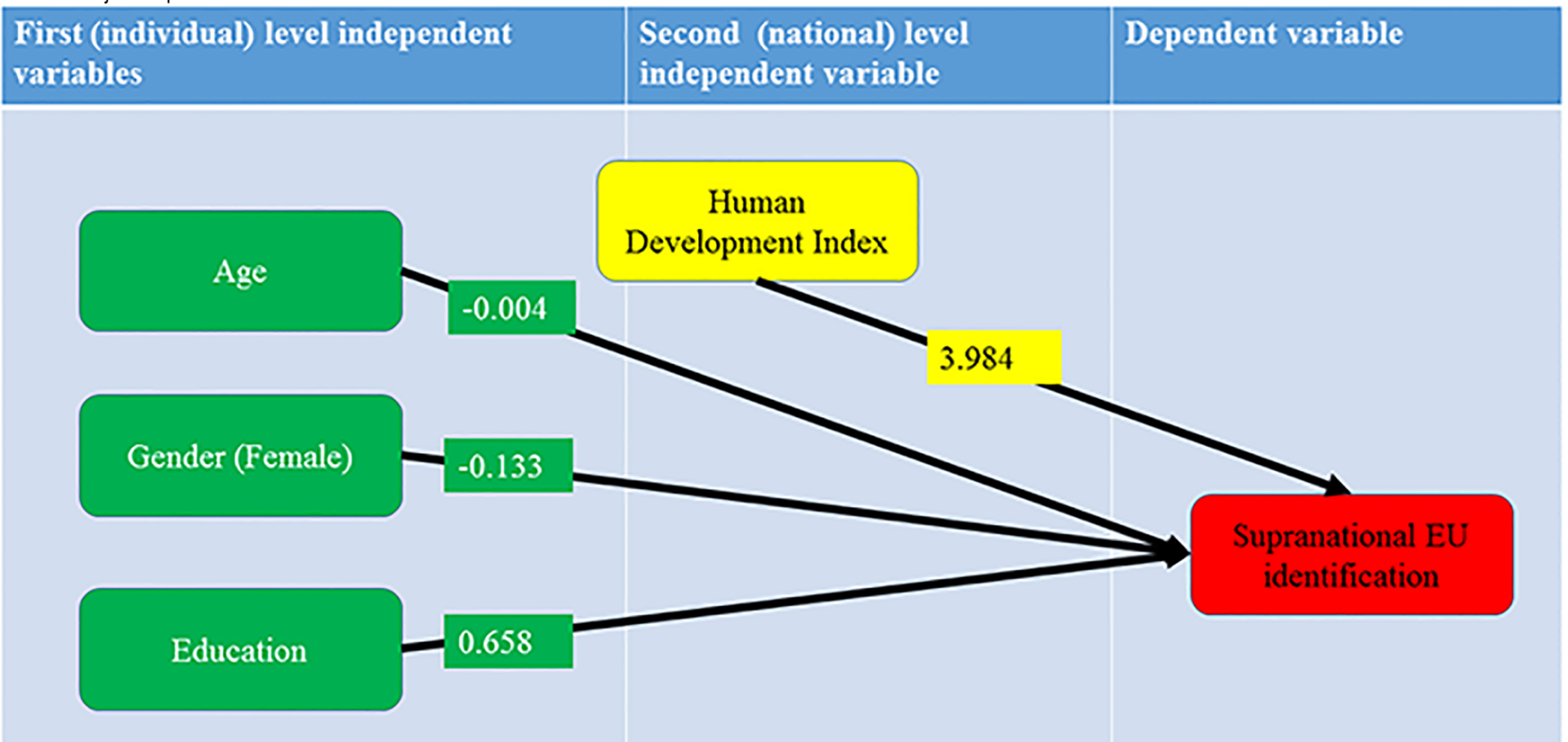
- Linearna regresijska analiza v striktnem smislu praviloma zahteva intervalne spremenljivke
- Druge različice regresijske analize omogočajo tudi vključevanje drugih vrst spremenljivk, praviloma tako, da jih spremenimo v dihotomne
- Morebitna dihotomna odvisna spremenljivka ne omogoča linearne, ampak zahteva npr. *binarno logistično regresijo*
- Tako odvisno spremenljivko interpretiramo kot verjetnost, ki se giblje med 0 in 1

Večstopenjska hierarhična regresija

- Več ravni analize
 - Individualna enota (prva stopnja analize) je vgnezdena v širši enoti (druga stopnja analize)
 - Ta širša enota je vgnezdena v še širši enoti analize (tretja stopnja analize)
 - Itd.

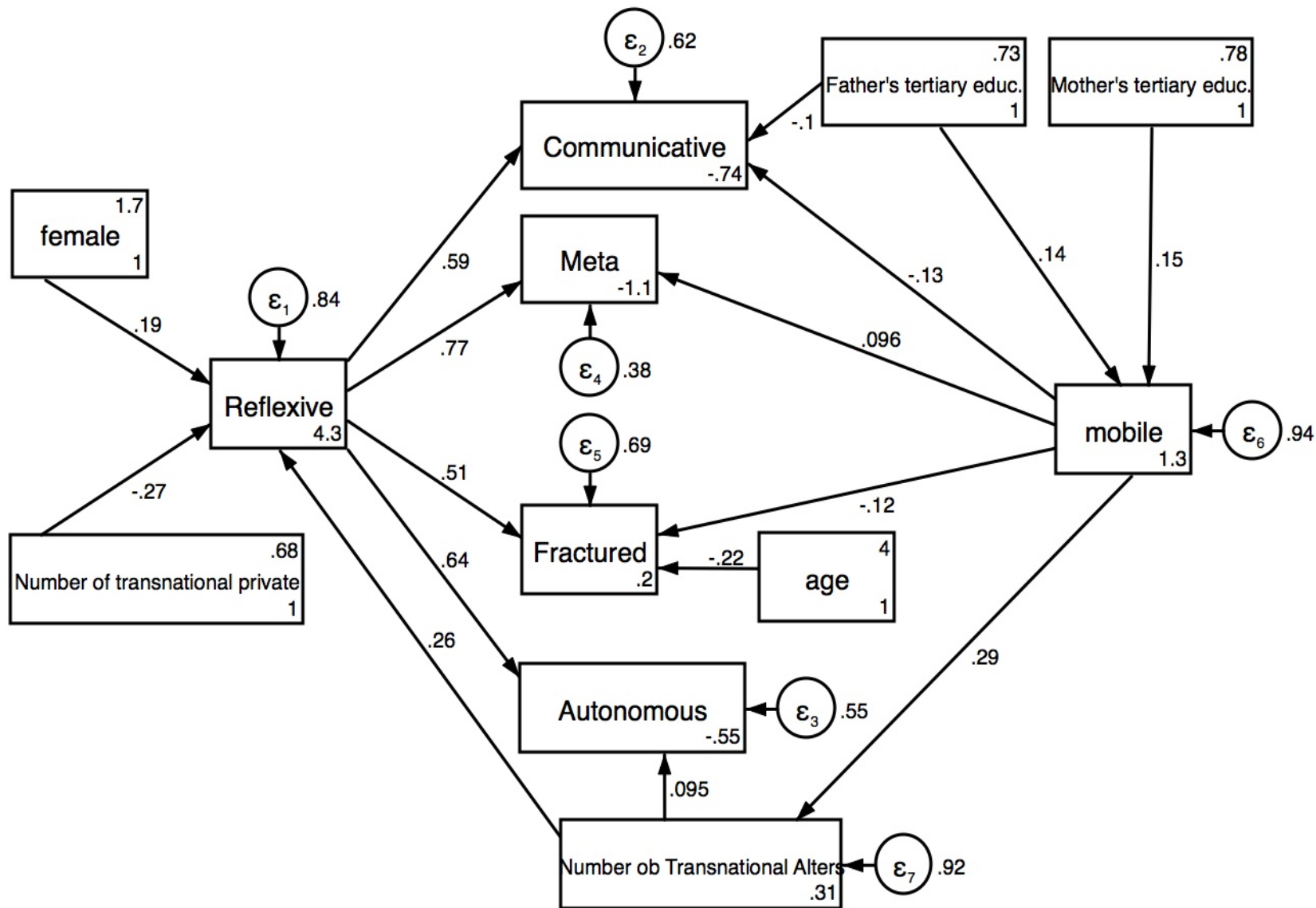
Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	0,519	0,069	7,536	24	<0,001
HDI, γ_{01}	3,984	1,526	2,610	24	0,015
For AGE slope, β_1					
INTRCPT2, γ_{10}	-0,004	0,002	-1,663	22821	0,096
For EDUCATER slope, β_2					
INTRCPT2, γ_{20}	0,658	0,045	14,742	22821	<0,001
For FEMALE slope, β_3					
INTRCPT2, γ_{30}	-0,133	0,028	-4,745	22821	<0,001

doi:10.1371/journal.pone.0146584.t005



Analiza poti

- Razširitev regresijskih modelov z
 - Več odvisnimi spremenljivkami
 - Eno ali več intervenirajočimi spremenljivkami



Modeliranje strukturnih enačb

- Razširitev analize poti:
- Vključitev več stopenj (kot večstopenjska hierarhična regresija)
- Vključitev latentnih spremenljivk (analogno faktorjem oziroma komponentam v faktorski analizi)

Ideje za nadaljnji študij

- Košmelj, Blaženka in Jože Rován (2007): *Statistično sklepanje*. Ljubljana. EF.
- Manoilov, Marjan (2010): *Osnove metodologije in statistike*. Nova Gorica: Zavod Novoznanje. <http://www.novoznanje.eu/e-gradiva/metodologija-in-statistika-statisticne-analize-obdelave/>
- Norušis, Marija (2006): *IBM SPSS Statistics 19 Guide to Data Analysis*. Addison Wesley.
- Stata User's Guide, <http://www.stata.com/manuals13/u.pdf>
- Raudenbush, S.W. and Byrk, A.S. (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications, London, 2002
- Priročniki iz statistično-metodološke zbirke Sage za posamezne zahtevnejše metode...