

# Kanonička korelaciona analiza

# Kanonička korelaciona analiza

- Kombinacija multiple regresije i faktorske analize / analize glavnih komponenti
- U kontekstu valjanosti, odnosi se na prognostičku i kriterijumsku valjanost
  - Predviđanje nekog spoljnog kriterijuma
    - Predviđamo uspeh na fakultetu na osnovu ocena iz prve godine - da li su ocene iz svih predmeta jednako relevantne?
  - Ali se može koristiti i bez jasne ideje o prediktorskom i kriterijumskom skupu - kao provera povezanosti dve grupe varijabli
    - Kako se HEXACO crte preslikavaju na Velikih 5

# Analiza glavnih komponenti – podsećanje

- Šta je cilj AGK?
  - $n$  izvornih koreliranih varijabli  $\rightarrow$  manji broj nezavisnih linearnih kombinacija izvornih varijabli
  - svaka nova komponenta objašnjava najveću moguću proporciju rezidualne varijanse
- Osnovna pitanja u AGK?
  - Koliko je komponenti potrebno da bismo objasnili odnos datih varijabli (matricu korelacija)?
  - Kakva je priroda ovih komponenti?
  - Koliko dobro pretpostavljene komponente objašnjavaju date manifestne varijable?
  - Kolika je proporcija specifične varijanse i varijanse greške?

# Linearna regresija – podsećanje

- Šta je cilj MR?
  - n izvornih koreliranih prediktorskih varijabli -> linearna kombinacija koja najbolje predviđa kriterijumsku varijablu
- Osnovna pitanja u MR:
  - Da li postoji veza između nezavisnih varijabli i zavisne varijable?
  - Koliki procenat varijanse kriterijuma možemo objasniti pomoću prediktora?
  - Koliki je specifični doprinos pojedinih prediktora u predviđanju kriterijuma?
  - Kakva je struktura latentne dimenzije (linearnog kompozita)?

# Predviđanje složenog kriterijuma

- Pretpostavimo da na osnovu srednjoškolskog uspeha, poena na testu znanja i poena na TOI želimo da predvidimo uspešnost u studijama
- Šta čini nekog uspešnim studentom?
  - Prosečna ocena
  - Redovnost studiranja
  - Broj ostvarenih ESPB u datom trenutku
  - ...
- Imamo kompleksan kriterijum!

# Predviđanje složenog kriterijuma

Koje sve opcije imamo na raspolaganju u slučaju složenog kriterijuma (kada nemamo jednu, već nekoliko kriterijumskih varijabli)?

- Nekoliko multiplih regresija
- Sumarni skor za kriterijumski skup
- Računamo prvu glavnu komponentu za kriterijumski skup
- KKA

# Predviđanje složenog kriterijuma: nekoliko multiplih regresija

- Povećava se verovatnoća da će neka regresija biti značajna samo na osnovu slučaja
- Dobijene regresione funkcije nisu nezavisne jer su kriterijumske varijable takođe međusobno korelirane – varijansa koju one objašnjavaju nije aditivna

# Predviđanje složenog kriterijuma: sumarni skor kriterijumske varijable

Implicitno pretpostavljamo da su sve varijable jednako važne i učestvuju u linearnom kompozitu svom svojom varijansom

- Varijable ne koreliraju jednako sa sumarnim skorom, a pogotovo ne sa regresionom funkcijom koja je linearna kombinacija prediktora
- Varijable nisu nezavisne pa postoji veliko preopkrivanje među njima



# Predviđanje složenog kriterijuma: prva glavna komponenta za kriterijumski skup

- Može imati smisla ako su kriterijumske varijable relativno homogene po prirodi ili nam nije važno da pravimo distinkciju između njih
- Ali ako nisu – gubimo važne delove kriterijuma (koji ostaju van glavne komponente) koje naši prediktori potencijalno predviđaju (na različite načine)

# Predviđanje složenog kriterijuma

Kanonička korelaciona analiza predstavlja generalizaciju postupka multiple regresije na situaciju kada imamo **više nezavisnih** (prediktorskih) varijabli ( $x$ ) i **više zavisnih** (kriterijumskih) varijabli ( $y$ )

Rešava probleme prethodnih rešenja

- Samo jedna analiza (manji rizik od lažnih pozitiva)
- Uzima u obzir korelacije i unutar prediktorskog i unutar kriterijumskog skupa
- Može opisati nekoliko načina povezanosti između dva skupa

# Kanonička korelaciona analiza

- KKA predstavlja postupak **linearne transformacije** inicijalnih varijabli koje su međusobno korelirane u **linearne kompozite koji su međusobno nezavisni** (ortogonalni)
  - Slično kao AGK
- Ovo se radi **simultano u dva skupa** – prediktorskom (levom) i kriterijumskom (desnom)
- Ono što se **maksimizuje** jeste **korelacija** ova dva linearna kompozita - kao da radimo dve AGK, s tim što je uslov da maksimizujemo korelaciju između prve GK iz jednog skupa i prve GK iz drugog skupa
- Linearne kompozite nazivamo **kanoničkim komponentama, kanoničkim faktorima ili kanoničkim funkcijama**
- Kao kod FA, kanoničke komponente se mogu posmatrati kao latentne dimenzije

# Kanonička korelaciona analiza

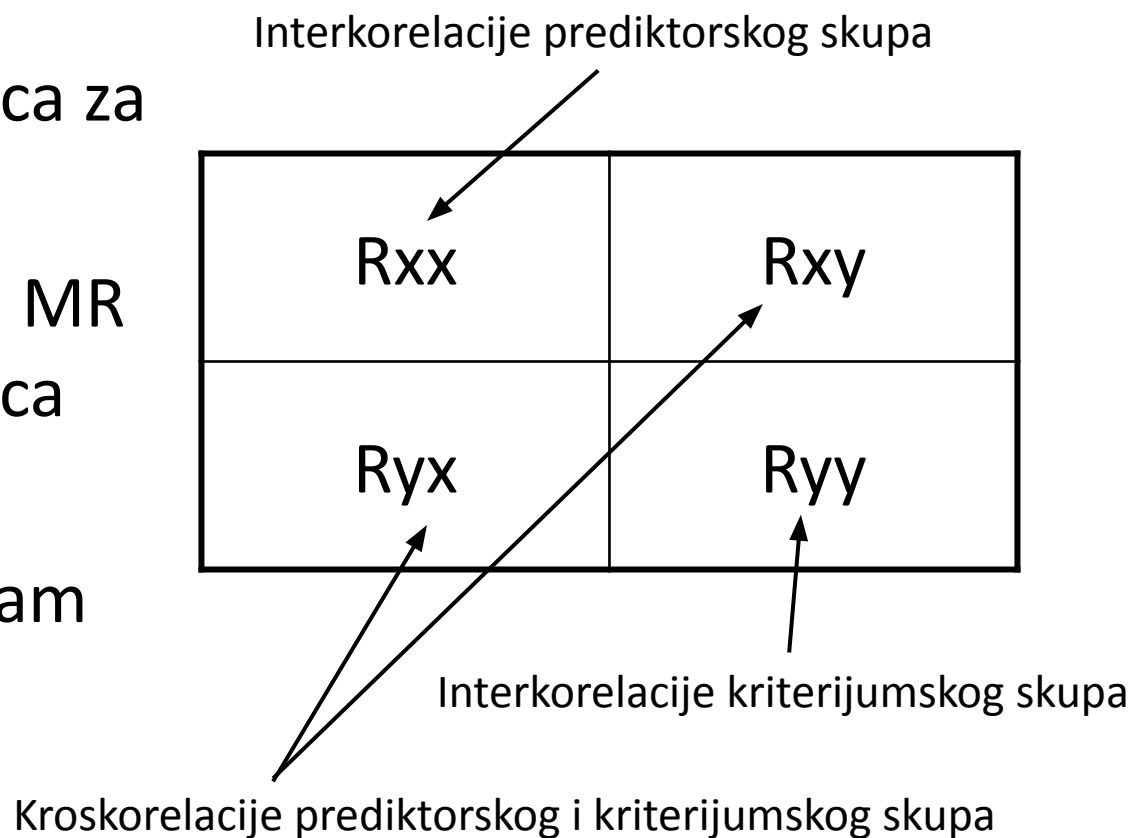
- Korelacije između kanoničkih komponenti nazivamo **kanoničkim korelacijama**
- Prva kanonička korelacija je najveća moguća Pirsonova korelacija između bilo koja dva linearna kompozita u jednom i drugom skupu varijabli - prostije, to je korelacija između prve komponente iz levog skupa i prve komponente iz desnog skupa
- Druga kanonička korelacija je sledeća najveća moguća linearna korelacija između ova dva skupa varijabli - analogno drugoj glavnoj komponenti u AGK
- Kanoničke korelacije su **uvek pozitivne** (imaju vrednost od 0 do 1)!

# Kanonička korelaciona analiza

- Koliko najviše kanoničkih korelacija možemo dobiti?
- Onoliko koliko ima varijabli u manjem skupu
  - Pošto se nakon toga „iscrpljuje“ varijansa manjeg skupa
- Dobijamo onoliko parova linearnih kompozita koliko ima varijabli u manjem skupu
- Ako manji skup čini samo 1 varijabla dobijamo identičnu korelaciju kao u multiploj regresiji
  - KKA je generalizacija MR!

# KKA ulazna matrica

- Šta je ulazna matrica za KKA?
- Kao i kod AGK i kod MR – u pitanju je matrica korelacija
- Za KKA, važne su nam četiri matrice



Rxx - interkorelacije prediktorskog skupa

	N	E	O	A	C
N	1	-.429	-.169	-.272	-.691
E	-.429	1	.508	.071	.468
O	-.169	.508	1	.078	.216
A	-.272	.071	.078	1	.254
C	-.691	.468	.216	.254	1

Rxy - kroskorelacije prediktorskog i kriterijumskog skupa

	impulsi	frustra	sadist
N	.580	.469	.440
E	-.205	-.233	-.283
O	-.090	-.171	-.233
A	-.456	-.420	-.422
C	-.583	-.336	-.394

Ryx - kroskorelacije kriterijumskog i prediktorskog skupa

	N	E	O	A	C
impulsi	.580	-.205	-.090	-.456	-.583
frustra	.469	-.233	-.171	-.420	-.336
sadist	.440	-.283	-.233	-.422	-.394

Ryy - interkorelacije kriterijumskog skupa

	impulsi	frustra	sadist
impulsi	1	.648	.689
frustra	.648	1	.765
sadist	.689	.765	1

# Kako dobijamo kanoničke komponente?

- Kanoničku korelacionu analizu možemo posmatrati kao dvostruku multiplu regresiju

$$\psi_1 y_1 + \psi_2 y_2 + \psi_n y_n = \beta_1 x_1 + \beta_2 x_2 + \beta_n x_n$$

$\psi_n$  - ponder za levi skup

$\beta_n$  - ponder za desni skup

$y_n$  - manifestna var. levog skupa

$x_n$  - manifestna var. desnog skupa

- Kako određujemo veličinu pondera  $\psi$  i  $\beta$ ?
  - Analogno multiploj regresiji
  - Koeficijenti će biti utoliko veći ukoliko su korelacije sa varijablama iz drugog seta (kroskorelacije) veće, a korelacije sa varijablama iz vlastitog seta (interkorelacije) manje



# Kako dobijamo kanoničke komponente?

- Slično kao u faktorskoj analizi, problem se matematički rešava preko svojstvenih vrednosti

$R_{xx}^{-1}$	$R_{xy}$
$R_{yx}$	$R_{yy}^{-1}$

- Kvadrat koeficijenta kanoničke korelacije - svojstvena vrednost matrice koja uzima u obzir matrice interkorelacija unutar oba skupa varijabli i kroskorelacije između skupova

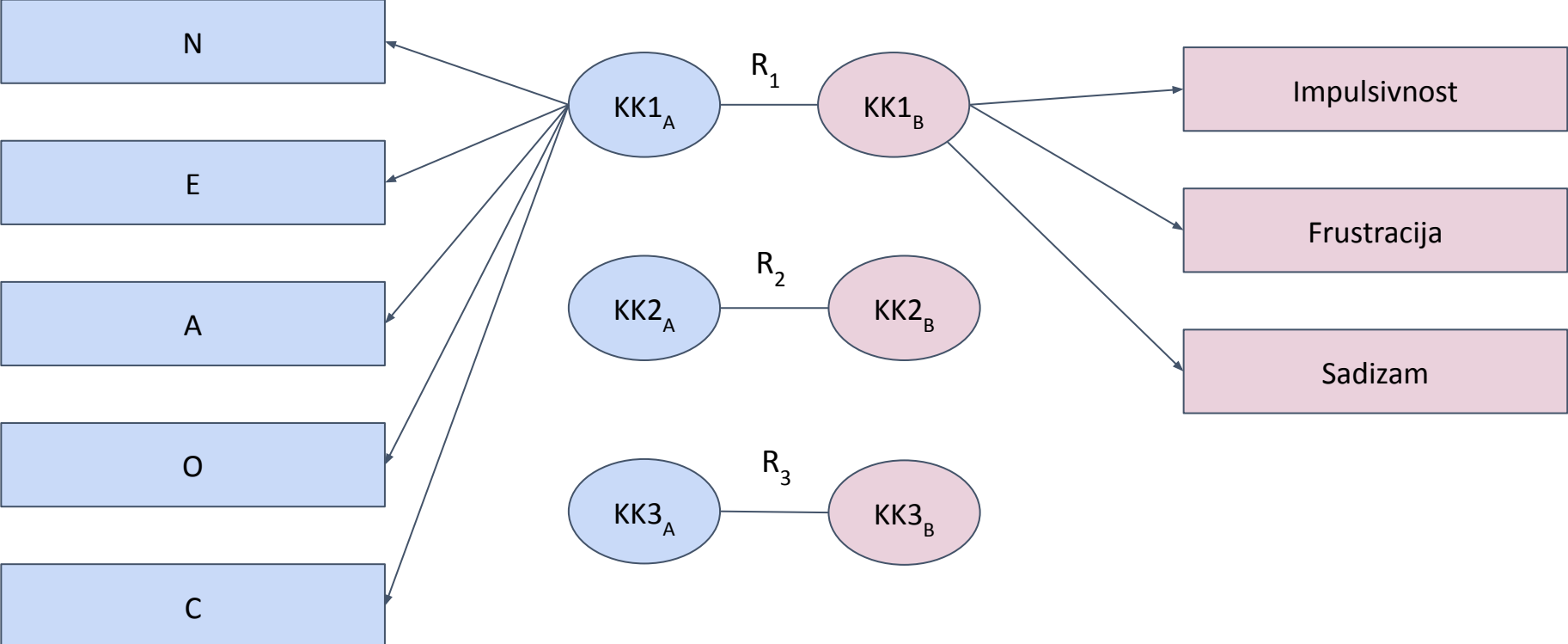
# Kako dobijamo kanoničke komponente?

- Posledica ovakvog računanja je:
  - Svaka kanonička korelacija je manja od prethodne, a veća od naredne
  - Kanonička komponenta je u korelaciji SAMO sa korespondentnom komponentom iz opozitnog skupa
    - ortogonalna je na sve druge komponente iz **oba** skupa

# KKK grafički

Set A- Velikih 5

Set B - Amoral

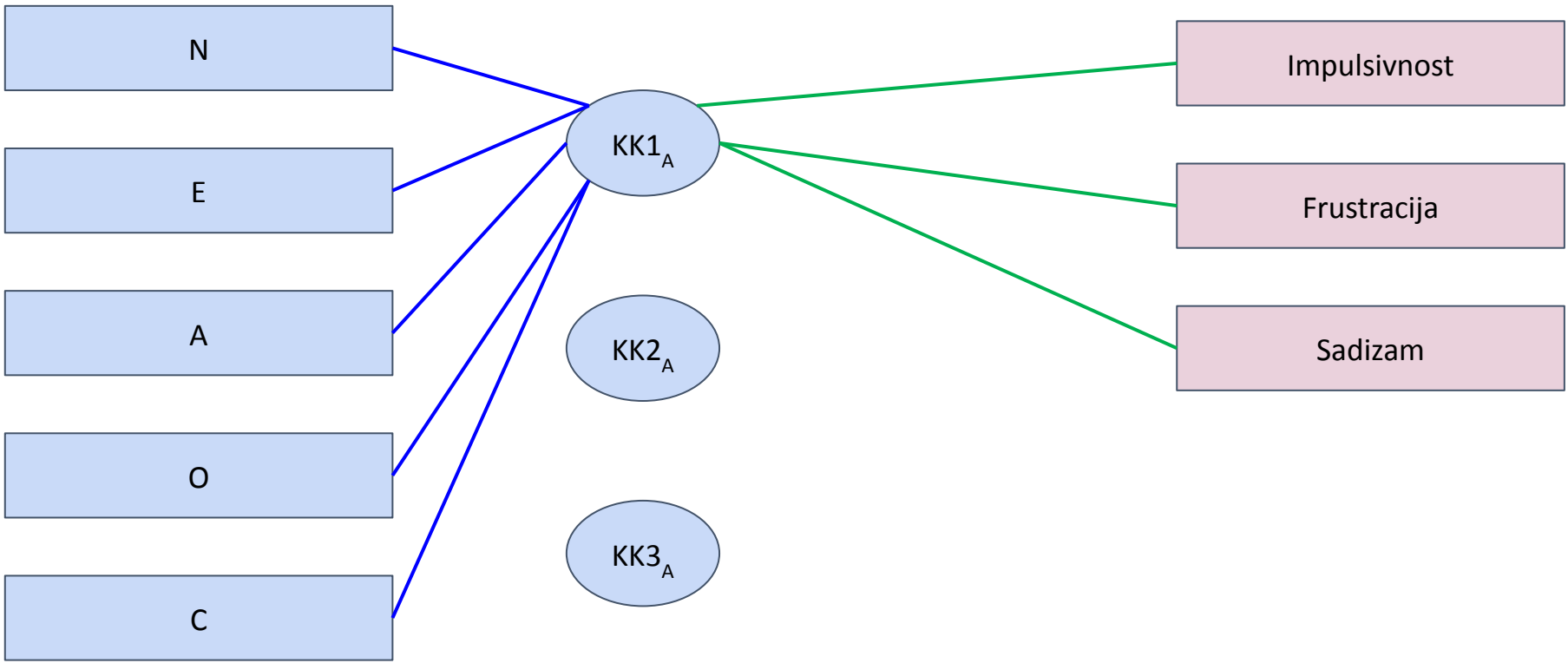


# Koeficijenti strukture i krosstrukture

Set A- Velikih 5

koeficijenti strukture - povezanost  
varijabli sa kanoničkim  
komponentama sopstvenog skupa

Set B - Amoral



koeficijenti krosstrukture - povezanost  
varijabli sa kanoničkim  
komponentama opozitnog skupa

# Osnovna pitanja u KKA

- Da li prediktorski skup predviđa kriterijumski?
  - Na koliko načina i u kom intenzitetu
- Koja je priroda povezanosti ova dva skupa?
- Koji procenat varijanse kriterijumskog skupa možemo objasniti na osnovu prediktorskog?

# Interkorelacije prediktorskog skupa

---

	N	E	O	A	C
N	1	-.429	-.169	-.272	-.691
E	-.429	1	.508	.071	.468
O	-.169	.508	1	.078	.216
A	-.272	.071	.078	1	.254
C	-.691	.468	.216	.254	1

---

# Interkorelacije kriterijumskog skupa

---

	impulsi	frustra	sadist
impulsi	1	.648	.689
frustra	.648	1	.765
sadist	.689	.765	1

---

# Kroskorelacije dva skupa varijabli

---

	impulsi	frustra	sadist
N	.580	.469	.440
E	-.205	-.233	-.283
O	-.090	-.171	-.233
A	-.456	-.420	-.422
C	-.583	-.336	-.394

---



# Ima li kanoničkih korelacija?

- Koliko maksimalno možemo dobiti kanoničkih korelacija za dva skupa?
  - Onoliko koliko ima varijabli u manjem skupu
- Ali da li će nam sve biti relevantne?
  - Ne nužno, neke korelacije mogu biti i neznačajne
- $H_0$  u KKA je da ne postoji ni jedna linearna kombinacija varijabli u dva seta varijabli čija je korelacija veća od 0

# Ima li kanoničkih korelacija?

Koeficijenti kanoničkih korelacija



Značajnost



	Rho	Lambda	Hi2	df	sig
1	.706	.426	825.342	15	.000
2	.346	.851	156.188	8	.000
3	.182	.967	32.757	3	.118

# Koja je priroda povezanosti?

- Za ovo nam je potrebno da gledamo odnose između izvornih varijabli i kanoničkih komponenti
- Slično kao u FA, postoji nekoliko matrica / koeficijenata koji su nam relevantni
  - **Koeficijent kanoničke funkcije** (matrica kanoničkih koeficijenata) - koliko izvorna varijabla doprinosi građenju kanoničke komponente
  - **Koeficijent kanoničke strukture** (matrica strukture) - korelacija izvorne varijable i kanoničke komponente
  - **Koeficijent kanoničke kros-strukture** (matrica kros-strukture)- korelacija izvorne varijable i kanoničke komponente suprotnog skupa
- Sve ove matrice / koeficijente dobijamo DVA puta – za levi (prediktorski) i desni (kriterijumski skup)

# Koeficijenti u KKA

<b>Koeficijent kanoničke funkcije</b>	<b>Koeficijent strukture</b>	<b>Koeficijent krosstrukture</b>
<ul style="list-style-type: none"><li>• <b>Doprinosi</b> izvornih varijabli izgradnji kanoničkih funkcija</li><li>• Analogni beta ponderima i faktorskim koeficijentima</li><li>• Njihova veličina zavisi od korelacije sa kanoničkom komponentom (poželjno što viša), ali i sa ostalim varijablama iz seta (poželjno što niža)</li><li>• Mogu imati vrednosti veće od 1</li></ul>	<ul style="list-style-type: none"><li>• <b>Korelacije</b> izvornih varijabli i kanoničkih komponenti</li><li>• Analogni koeficijentima strukture iz FA i regresionom faktoru iz MR</li><li>• Procenat deljene varijanse između izvorne varijable i kanoničke komponente - koeficijent strukture na kvadrat</li></ul>	<ul style="list-style-type: none"><li>• <b>Korelacije</b> izvornih varijabli i kanoničkih komponenti iz OPOZITNOG skupa</li><li>• Ne preslikavaju se direktno na koeficijente iz FA i MR</li><li>• Proizvod koeficijenta strukture i odgovarajuće kanoničke korelacije</li><li>• Procenat deljene varijanse između izvorne varijable i kanoničke komponente OPOZITNOG SKUPA - koeficijent krosstrukture na kvadrat</li></ul>

# Svi koeficijenti za levi skup

---

	$w_x1$	$f_x1$	$f1_{xy}$	$w_x2$	$f_x2$	$f2_{xy}$	$w_x3$	$f_x3$	$f3_{xy}$
N	-.471	-.836	-.590	.381	.078	.027	-1.097	-.270	-.049
E	-.125	.317	.224	-.515	-.449	-.155	-.285	-.449	-.082
O	.017	.160	.113	-.444	-.580	-.201	-.410	-.568	-.104
A	.444	.673	.475	-.424	-.336	-.116	-.045	-.028	-.005
C	.426	.809	.571	1.035	.328	.113	-.903	-.379	-.069

---

$w_x$  - kanonički koeficijenti

$f_x$  - kanonički faktori (koeficijenti strukture)

$f_{xy}$  - kanonički krosfaktori (koeficijenti krosstrukture)

# Kanonički koeficijenti za levi skup

	$w_x1$	$f_x1$	$f1_{xy}$	$w_x2$	$f_x2$	$f2_{xy}$	$w_x3$	$f_x3$	$f3_{xy}$
N	<b>-.471</b>	-.836	-.590	.381	.078	.027	-1.097	-.270	-.049
E	-.125	.317	.224	-.515	-.449	-.155	-.285	-.449	-.082
O	.017	.160	.113	-.444	-.580	-.201	-.410	-.568	-.104
A	<b>.444</b>	.673	.475	-.424	-.336	-.116	-.045	-.028	-.005
C	<b>.426</b>	.809	.571	1.035	.328	.113	-.903	-.379	-.069

# Koeficijenti strukture za levi skup

	$w_x1$	$f_x1$	$f1_{xy}$	$w_x2$	$f_x2$	$f2_{xy}$	$w_x3$	$f_x3$	$f3_{xy}$
N	-.471	<b>-.836</b>	-.590	.381	.078	.027	-1.097	-.270	-.049
E	-.125	.317	.224	-.515	-.449	-.155	-.285	-.449	-.082
O	.017	.160	.113	-.444	-.580	-.201	-.410	-.568	-.104
A	.444	<b>.673</b>	.475	-.424	-.336	-.116	-.045	-.028	-.005
C	.426	<b>.809</b>	.571	1.035	.328	.113	-.903	-.379	-.069

# Koeficijenti kros-strukture za levi skup

	$w_x1$	$f_x1$	$f1_{xy}$	$w_x2$	$f_x2$	$f2_{xy}$	$w_x3$	$f_x3$	$f3_{xy}$
N	-.471	-.836	<b>-.590</b>	.381	.078	.027	-1.097	-.270	-.049
E	-.125	.317	.224	-.515	-.449	-.155	-.285	-.449	-.082
O	.017	.160	.113	-.444	-.580	-.201	-.410	-.568	-.104
A	.444	.673	<b>.475</b>	-.424	-.336	-.116	-.045	-.028	-.005
C	.426	.809	<b>.571</b>	1.035	.328	.113	-.903	-.379	-.069

$w_y$  - kanonički koeficijenti

$f_y$  - kanonički faktori (koeficijenti strukture)

$f_{yx}$  - kanonički krosfaktori (koeficijenti krosstrukture)



# Svi koeficijenti za desni skup

---

	$w_y1$	$f_y1$	$f1_{yx}$	$w_y2$	$f_y2$	$f2_{yx}$	$w_y3$	$f_y3$	$f3_{yx}$
Impulsi	-.866	-.991	-.700	-1.132	-.128	-.044	-.101	.037	.007
Frustra	-.151	-.742	-.524	.766	.595	.206	-1.404	-.307	-.056
Sadist	-.039	-.752	-.531	.736	.542	.187	1.52	.376	.069

---

# Kanonički koeficijenti za desni skup

	$w_{y1}$	$f_{y1}$	$f1_{yx}$	$w_{y2}$	$f_{y2}$	$f2_{yx}$	$w_{y3}$	$f_{y3}$	$f3_{yx}$
Impulsi	<b>-.866</b>	<b>-.991</b>	<b>-.700</b>	-1.132	-.128	-.044	-.101	.037	.007
Frustra	-.151	<b>-.742</b>	<b>-.524</b>	.766	.595	.206	-1.404	-.307	-.056
Sadist	-.039	<b>-.752</b>	<b>-.531</b>	.736	.542	.187	1.52	.376	.069

# Koeficijenti strukture za desni skup

	$w_y1$	$f_y1$	$f1_{yx}$	$w_y2$	$f_y2$	$f2_{yx}$	$w_y3$	$f_y3$	$f3_{yx}$
Impulsi	-.866	<b>-.991</b>	-.700	-1.132	-.128	-.044	-.101	.037	.007
Frustra	-.151	<b>-.742</b>	-.524	.766	.595	.206	-1.404	-.307	-.056
Sadist	-.039	<b>-.752</b>	-.531	.736	.542	.187	1.52	.376	.069

# Koeficijenti kros-strukture za desni skup

	$w_{y1}$	$f_{y1}$	$f1_{yx}$	$w_{y2}$	$f_{y2}$	$f2_{yx}$	$w_{y3}$	$f_{y3}$	$f3_{yx}$
Impulsi	-.866	-.991	<b>-.700</b>	-1.132	-.128	-.044	-.101	.037	.007
Frustra	-.151	-.742	<b>-.524</b>	.766	.595	.206	-1.404	-.307	-.056
Sadist	-.039	-.752	<b>-.531</b>	.736	.542	.187	1.520	.376	.069

# Koja je priroda povezanosti?

- Kanoničke komponente posmatramo kao latentne varijable
- O sadržaju tih varijabli zaključujemo na osnovu kanoničkih koeficijenata i koeficijenata strukture
  - analogno regresiji (izvorne korelacije i beta ponderi)
- Ukoliko je **apsolutna vrednost** koeficijenta veća – varijabla je važnija za interpretaciju kanoničke komponente

# Koja je priroda povezanosti?

	$w_x 1$	$f_x 1$	$f1_{xy}$
N	<b>-.471</b>	<b>-.836</b>	-.590
E	-.125	.317	.224
O	.017	.160	.113
A	<b>.444</b>	<b>.673</b>	.475
C	<b>.426</b>	<b>.809</b>	.571

	$w_y 1$	$f_y 1$	$f1_{yx}$
Impulsi	<b>-.866</b>	<b>-.991</b>	<b>-.700</b>
Frustra	-.151	<b>-.742</b>	<b>-.524</b>
Sadist	-.039	<b>-.752</b>	<b>-.531</b>

	$w_x 1$	$f_x 1$	$f1_{xy}$
N	<b>-.471</b>	<b>-.836</b>	-.590
E	-.125	.317	.224
O	.017	.160	.113
A	<b>.444</b>	<b>.673</b>	.475
C	<b>.426</b>	<b>.809</b>	.571

Prvu kanoničku komponentu LEVO čine N-, A+ i C+

	$w_y 1$	$f_y 1$	$f1_{yx}$
Impulsi	<b>-.866</b>	<b>-.991</b>	<b>-.700</b>
Frustra	-.151	<b>-.742</b>	<b>-.524</b>
Sadist	-.039	<b>-.752</b>	<b>-.531</b>

Prvu kanoničku komponentu DESNO čini impulsivnost

Kako su povezane ove dve kanoničke komponente?

# Koja je priroda povezanosti?

- **VAŽNO:** Pošto je kanonička korelacija uvek pozitivna, smer veza se određuje na osnovu koeficijenata strukture
- Ako su koeficijenti strukture u dva skupa istog smera (-- ili ++) veze su pozitivne
- Ako su različitog smera veze su negativne
- Prilikom interpretacije može se promeniti predznak svim varijablama (ali se to mora uraditi u OBA seta)



# Koja je priroda povezanosti?

- Rekli smo da za interpretaciju gledamo kanoničke koeficijente i koeficijente strukture
- Međutim, može se desiti da se ovi koeficijenti međusobno bitno razlikuju
  - analogno razlici između izvorne korelacije i beta pondera u regresiji

	$w_y$	$f_y$	$f_{yx}$
Impulsi	<b>-.866</b>	<b>-.991</b>	<b>-.700</b>
Frustra	<b>-.151</b>	<b>-.742</b>	<b>-.524</b>
Sadist	<b>-.039</b>	<b>-.752</b>	<b>-.531</b>

# Tipovi prediktora u KKA

- Kao i kod MR možemo razlikovati 4 tipa prediktora
  - Dobar, loš prediktor, redundantnu varijablu i supresor
- Prepoznamo ih na osnovu kombinacije doprinosa i korelacije odnosno kanoničkog koeficijenta i koeficijenta strukture
  - **Dobar prediktor** – oba visoka
  - **Loš prediktor** – oba niska
  - **Redundantna varijabla** – visoka korelacija (koeficijent strukture), ali nizak doprinos (kanonički koeficijent); nije neophodna u modelu
  - **Supresor** - niska korelacija (koeficijent strukture), ali visok doprinos (kanonički koeficijent) ili suprotan smer korelacije i doprinosa (oba moraju biti različita od 0); može ukazivati na posredan uticaj varijable ili biti statistički artefakt

# Tipovi prediktora u KKA

---

	$w_y1$	$f_y1$	$f1_{yx}$	$w_y2$	$f_y2$	$f2_{yx}$	$w_y3$	$f_y3$	$f3_{yx}$
Impulsi	-.866	-.991	-.700	-1.132	-.128	-.044	-.101	.037	.007
Frustra	-.151	-.742	-.524	.766	.595	.206	-1.404	-.307	-.056
Sadist	-.039	-.752	-.531	.736	.542	.187	1.520	.376	.069

---

# Proporcija objašnjene varijanse

- a) Najčešće pitanje: koliko varijanse kriterijumskog skupa možemo objasniti prediktorskim skupom?
  - Odnosno koji procenat varijanse OPOZITNOG skupa objašnjava data kanonička komponenta
  
- b) Ali može nas zanimati i koji procenat varijanse SVOG skupa objašnjava koja komponenta
  
- c) Koliko je svaka manifestna varijabla objašnjena komponentama iz svog i opozitnog skupa

# Proporcija objašnjene varijanse

- Ako želimo da dobijemo proporciju objašnjene varijanse – šta bismo mogli da gledamo?
  - Matricu strukture, jer se u njoj nalaze korelacije
  - Ili matricu kros-strukture, koja takođe sadrži korelacije, ali sa kanoničkim komponentama iz opozitnog skupa
- **VAŽNO** – kanoničke komponente su ortogonalne
  - u istom skupu
  - i opozitnom, osim para koji maksimalno korelira
- Zbog toga su procenti varijanse koji kanoničke komponente objašnjavaju (i u svom i u opozitnom setu) **ADITIVNI**

# Proporcija objašnjene varijanse

- Šta dobijamo ako kvadriramo koeficijente kanoničke strukture i saberemo po redovima?
  - Proporciju varijanse date varijable koja je objašnjena kanoničkim komponentama iz SVOG skupa
  - **Analogno komunalitetu u FA**
- A ako kvadriramo koeficijente kanoničke strukture i *uprosečimo* po kolonama?
  - Proporciju varijanse u celom skupu varijabli koju objašnjava data kanonička komponenta (iz ISTOG seta)
  - Naziva se **VARIJANSOM** kanoničke komponente
  - Analogno varijansi / svojstvenoj vrednosti faktora
  - Ukazuje na reprezentativnost komponente za svoj skup

# Proporcija objašnjene varijanse

	$f_{x1}$	$f_{x2}$	$f_{x3}$	$f_{x1}^2$	$f_{x2}^2$	$f_{x3}^2$	objašnjena varijansa manifestne varijable
N	-.836	.078	-.270	.699	.006	.073	.778
E	.317	-.449	-.449	.100	.202	.202	.504
O	.160	-.580	-.568	.026	.336	.323	.685
A	.673	-.336	-.028	.453	.113	.001	.567
C	.809	.328	-.379	.654	.108	.144	.906

# Proporcija objašnjene varijanse - varijansa kanoničke komponente

	$f_x^1$	$f_x^2$	$f_x^3$	$f_x^1{}^2$	$f_x^2{}^2$	$f_x^3{}^2$
N	-.836	.078	-.270	.699	.006	.073
E	.317	-.449	-.449	.100	.202	.202
O	.160	-.580	-.568	.026	.336	.323
A	.673	-.336	-.028	.453	.113	.001
C	.809	.328	-.379	.654	.108	.144
varijansa kanoničke komponente:				.386	.153	.148



# Proporcija objašnjene varijanse

- Kolika je ukupna varijansa u jednom i drugom setu?
  - Jednaka je broju (standardizovanih) varijabli
- To znači da ukupna varijansa u dva skupa ne mora biti jednaka
  - Npr. u jednom skupu imamo 5 varijabli, a u drugom 3
- Zbog ovoga ne sabiramo, već uprosečavamo kvadrate koeficijenata strukture po redovima!
- Šta su implikacije ovoga?
- Koju proporciju varijanse skupova možemo objasniti ako zadržimo sve kanoničke komponente (ako su sve značajne)?
  - Možemo objasniti celu varijansu manjeg skupa, ali ne i većeg
- Generalno, proporcija objašnjene varijanse u dva skupa će se razlikovati

# Proporcija objašnjene varijanse

- Da li će prva kanonička komponenta uvek objašnjavati više varijanse u svom skupu od druge?
  - NE!
- Zato što se ne maksimizuje varijansa komponenti, već korelacija sa komponentom iz opozitnog skupa!
- Redosled komponenti po varijansi ne mora biti isti u dva skupa (npr. 2. komponenta objašnjava najviše varijanse u levom, a 3. u desnom skupu)

# Proporcija objašnjene varijanse

- Šta dobijamo ako kvadriramo koeficijente kanoničke kros-strukture i saberemo po redovima?
  - Proporciju varijanse date varijable koja je objašnjena kanoničkim komponentama iz OPOZITNOG skupa
  - Analogno multiplom  $R^2$  za datu varijablu
- A ako kvadriramo koeficijente kanoničke kros-strukture i uprosečimo po kolonama?
  - Proporciju varijanse u celom skupu varijabli koju objašnjava kanonička komponenta iz OPOZITNOG seta
  - Naziva se **REDUNDANSOM** kanoničke komponente ili **INDEKSOM REDUNDANSE** ili **PREPOKRIVANJEM**
  - Analogno multiplom  $R^2$  za ceo skup

# Redundansa (prepokrivanje)

- Kako dobijamo koeficijente kros-strukture?
  - Pomnožimo koeficijente strukture kanoničkom korelacijom
- Kako onda dobijamo redundansu?
  - Pomnožimo varijansu kanoničke komponente kanoničkom korelacijom
  - Odnosno kvadriramo i uprosečimo koeficijente kros-strukture
- Kao ni varijanse, ni redundanse u dva skupa neće biti jednake

# Redundansa (prepokrivanje)

	$f_x1$	Rho1	$f1_{xy}$	$f_x2$	Rho2	$f2_{xy}$	$f_x3$	Rho3	$f3_{xy}$
N	-.836	.706	-.590	.078	.346	.027	-.270	.128	-.049
E	.317	.706	.224	-.449	.346	-.155	-.449	.128	-.082
O	.160	.706	.113	-.580	.346	-.201	-.568	.128	-.104
A	.673	.706	.475	-.336	.346	-.116	-.028	.128	-.005
C	.809	.706	.571	.328	.346	.113	-.379	.128	-.069

$f_{xn}$  - koeficijenti strukture

$Rho_n$  - prva kanonička korelacija

$f_{nxy}$  - koeficijenti krosstrukture

# Proporcija objašnjene varijanse

- Šta predstavlja kvadrat kanoničke korelacije? Da li je to proporcija varijanse koju dele skupovi?
- Kanoničko  $\rho^2$  je proporcija varijanse koju dele dve kanoničke komponente
  - Ne celi skupovi!
- Da li onda možemo da saberemo sve značajne  $\rho^2$  i dobijemo procenat varijanse koji dele dva skupa?
  - NE!
- Zato što ni sve značajne komponente ne moraju obuhvatiti kompletnu varijansu svog skupa (u većem skupu) i ne moraju biti jednako „reprezentativne“ za svoje skupove
- Time bismo zanemarili varijansu komponenti u svojim skupovima (možda je korelacija visoka, ali varijanse komponenti koje koreliraju niske)

# Proporcija objašnjene varijanse

- Da bismo dobili procenat varijanse koji jedan skup objašnjava u drugom – treba da saberemo sve indekse redundanse!
- Ovo možemo uraditi za oba skupa i nećemo dobiti iste vrednosti
- Redundansa jednog skupa govori koliko je on objašnjen komponentama iz opozitnog skupa
- Veličina koeficijenta redundanse može da ukazuje na smer uticaja
  - Smatra se da smer uticaja ide iz skupa u kome je redundansa manja (jer je on manje objašnjen opozitnim skupom)
  - Ali za ovakvo njegovo tumačenje treba da bude zadovoljeno mnogo dodatnih uslova

# Proporcija objašnjene varijanse

	Bazična struktura ličnosti			Amoral	
	Varijansa	Prepokrivanje		Varijansa	Prepokrivanje
KK1-1	.386	.193	KK2-1	<b>.699</b>	<b>.349</b>
KK1-2	.153	.018	KK2-2	.221	.026
KK1-3	.148	.005	KK2-3	.079	.003

	$w_y1$	$f_y1$	$f1_{yx}$	$w_y2$	$f_y2$	$f2_{yx}$	$w_y3$	$f_y3$	$f3_{yx}$
Impulsi	-.866	<b>-.991</b>	<b>-.700</b>	-1.132	-.128	-.044	-.101	.037	.007
Frustra	-.151	<b>-.742</b>	<b>-.524</b>	.766	.595	.206	-1.404	-.307	-.056
Sadist	-.039	<b>-.752</b>	<b>-.531</b>	.736	.542	.187	1.52	.376	.069

varijansa KK2-1:  $(-.991^2 + -.742^2 + -.752^2) / 3 = .699$

prepokrivanje KK2-1:  $(-.700^2 + -.524^2 + -.531^2) / 3 = .349$



# Proporcija objašnjene varijanse

	Bazična struktura ličnosti			Amoral	
	Varijansa	Prepokrivanje		Varijansa	Prepokrivanje
KK1-1	.386	.193	KK2-1	<b>.699</b>	<b>.349</b>
KK1-2	.153	.018	KK2-2	.221	.026
KK1-3	.148	.005	KK2-3	.079	.003

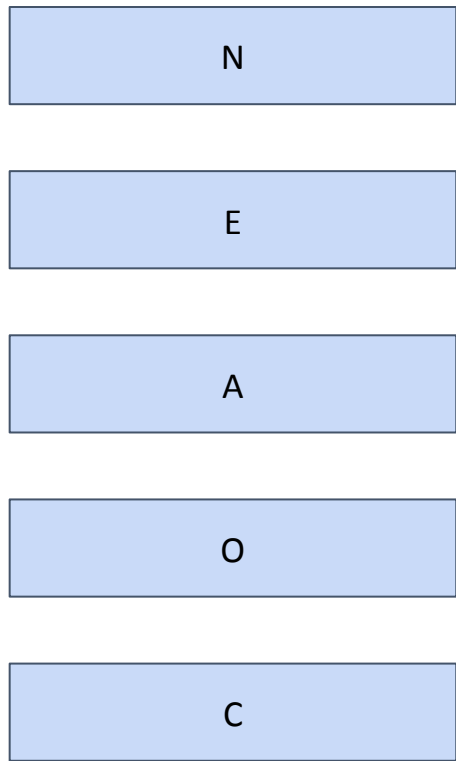
	Rho	Lambda	Hi2	df	sig
1	<b>.706</b>	.792	105.716	30	.000
2	.194	.936	3.163	18	.000
3	.182	.967	32.757	3	.000

$$.699 * .706^2 = .349$$

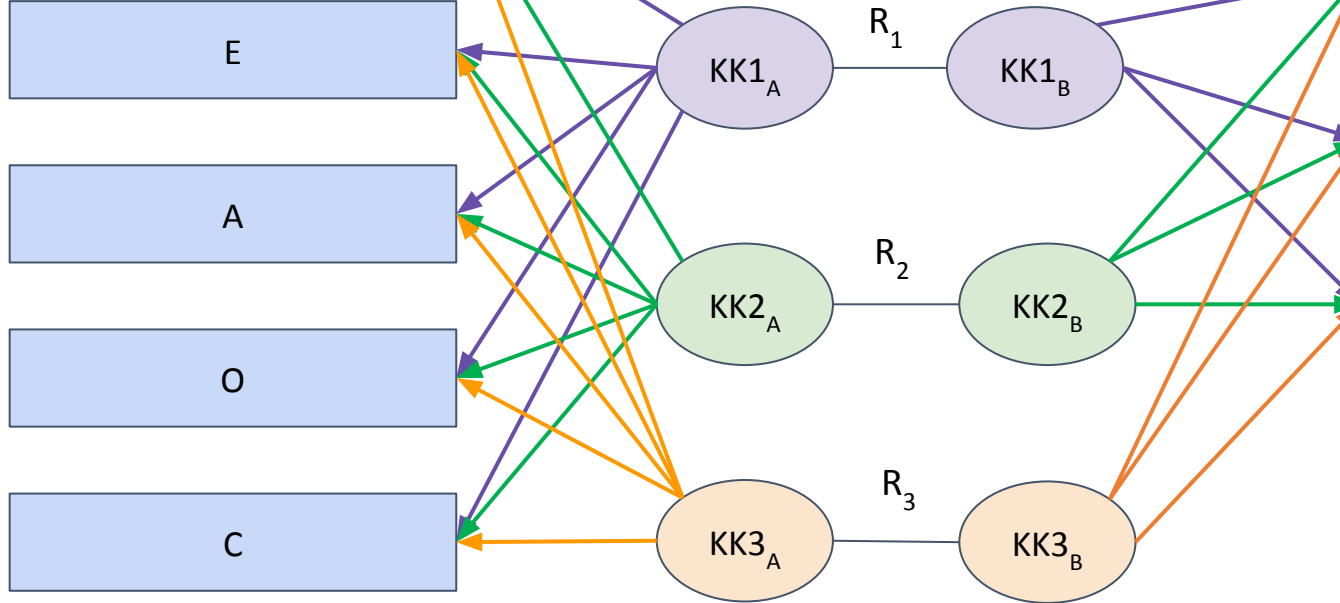
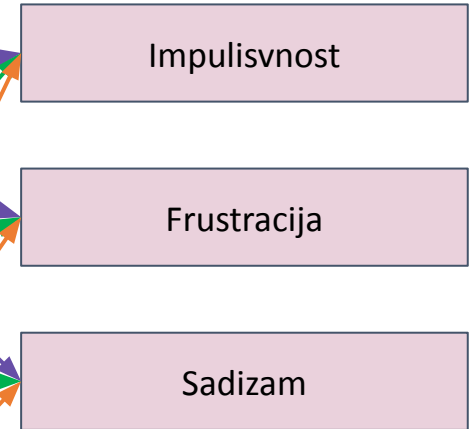
# Primer u SPSS-u

- Koliko osobine ličnosti predviđaju amoral?
  - Osobine ličnosti merene NEO-PI-R upitnikom
  - Amoral

## Set A- Velikih pet



## Set B - Self-koncept



$$\psi_1 y_1 + \psi_2 y_2 + \psi_n y_n = \beta_1 x_1 + \beta_2 x_2 + \beta_n x_n$$

$\psi_n$  - ponder za levi (B5) skup  
 $y_n$  - manifestna var. levog (B5) skupa

$\beta_n$  - ponder za desni (amoral) skup  
 $x_n$  - manifestna var. desnog (amoral) skupa

# Primer u SPSS-u

- U starije verzije SPSS nije direktno implementirana kanonička korelaciona analiza pa zato koristimo makroe
  - Počevši od SPSS 25, kanonika je implementirana u meniju *Correlations* → *Canonical*
  - Nažalost, u računarskoj učionici imamo SPSS 22 ili neki još stariji
- Izvorni SPSS-ov makro **CANCORR.SPS** ili makroi domaćih autora (Knežević i Momirović) **QCCR.SPS** i **KANONIKA.SPS**

# Primer u SPSS-u – pokretanje makroa

```
Include 'C:/kanonika.sps'.
```

```
kanonika
```

```
set 1 = neurot ekstra otvor dobroodus savest/
```

```
set2 = impulsi frustra sadizam/
```

```
inc = .05/.
```

- Potrebno je proveriti da li odgovarajući makro postoji i na datoj lokaciji
- Komanda inc je opciona, ako se izostavi koristi se p = .01

# Primer u SPSS-u

## Interkorelacije levog skupa varijabli

	neurot	ekstra	otvor	dobrodus	savest
neurot	1.000	-.429	-.169	-.272	-.691
ekstra	-.429	1.000	.508	.071	.468
otvor	-.169	.508	1.000	.078	.216
dobrodus	-.272	.071	.078	1.000	.254
savest	-.691	.468	.216	.254	1.000

## Interkorelacije desnog skupa varijabli

	impulsi	frustra	sadist
impulsi	1.000	.648	.689
frustra	.648	1.000	.765
sadist	.689	.765	1.000

## Kroskorelacije levog i desnog skupa varijabli

	impulsi	frustra	sadist
neurot	.580	.469	.440
ekstra	-.205	-.233	-.283
otvor	-.090	-.171	-.233
dobrodus	-.456	-.420	-.422
savest	-.583	-.336	-.394

# Primer u SPSS-u

```
Koeficijenti kanoničkih korelacija i njihova značajnost:
```

	Rho	Lambda	Hi2	df	sig
1	.706	.426	825.342	15.000	.000
2	.346	.851	156.188	8.000	.000
3	.182	.967	32.757	3.000	.000

- Koliko maksimalno kanoničkih korelacija možemo dobiti?
  - 3, jer toliko varijabli imamo u manjem skupu
- Sve tri kanoničke korelacije su značajne
- Najveća moguća korelacija između dva skupa je .706

# Primer u SPSS-u

- Kako znamo na koji način smo kreirali kanoničke komponente (linearne kompozite)?
  - Na osnovu kanoničkih koeficijenata
- Koje će biti dimenzije matrice kanoničkih koeficijenata u levom skupu, a koje u desnom?
  - U levom 5x3, u desnom 3x3



# Primer u SPSS-u

za prvu kanoničku komponentu

Kanonicki koeficijenti levog skupa varijabli

	1	2	3
N	-.471	.381	-1.097
E	-.125	-.515	-.285
O	.017	-.444	-.410
A	.444	-.424	-.045
C	.426	1.035	-.903

Kanonicki koeficijenti desnog skupa varijabli

	1	2	3
impulsi	-.866	-1.132	-.101
frustra	-.151	.766	-1.404
sadist	-.039	.736	1.520

- Najveći kanonički koeficijent u levom skupu ima Neuroticizam, zatim sledi Saradljivost, pa Savesnost
- U desnom skupu koeficijenti su manje ujednačeni, pa primetno najveći ima impulsivnost

# Primer u SPSS-u

- Kako interpretiramo prirodu kanoničkih komponenti?
  - (Dominantno) na osnovu matrice strukture
- Koje će biti dimenzije matrice strukture u levom skupu, a koje u desnom?
  - U levom 5x3, u desnom 3x3
- U ispisu, matrica strukture se zove matrica kanoničkih faktora (analogno regresionom faktoru)

# Primer u SPSS-u

za prvu kanoničku komponentu

Kanonicki faktori levog skupa varijabli

	1	2	3
N	-.836	.078	-.270
E	.317	-.449	-.449
O	.160	-.580	-.568
A	.673	-.336	-.028
C	.809	.328	-.379

Kanonicki faktori desnog skupa varijabli

	1	2	3
impulsi	-.991	-.128	.037
frustra	-.742	.595	-.307
sadist	-.752	.542	.376

- U prostoru ličnosti imamo nisko anksiozne, ali savesne i saradljive ljude
- Dok u prostoru amoralu, imamo niske koeficijente sva tri faceta
- Možemo isto tako reći da su generalno savesni, saradljivi i nisko neurotični ljudi i niski na amoralu

# Primer u SPSS-u

za drugu kanoničku komponentu

Kanonicki faktori levog skupa varijabli

	1	2	3
N	-.836	.078	-.270
E	.317	-.449	-.449
O	.160	-.580	-.568
A	.673	-.336	-.028
C	.809	.328	-.379

Kanonicki faktori desnog skupa varijabli

	1	2	3
impulsi	-.991	-.128	.037
frustra	-.742	.595	-.307
sadist	-.752	.542	.376

- U prostoru ličnosti govorimo pre svega o niskoj otvorenosti i introverziji, ali i niskoj saradljivosti, a visokoj savesnosti
- U prostoru amoralna, to je visoka frustriranost i sadizam

# Primer u SPSS-u

za treću kanoničku komponentu

Kanonicki faktori levog skupa varijabli

	1	2	3
N	-.836	.078	-.270
E	.317	-.449	-.449
O	.160	-.580	-.568
A	.673	-.336	-.028
C	.809	.328	-.379

Kanonicki faktori desnog skupa varijabli

	1	2	3
impulsi	-.991	-.128	.037
frustra	-.742	.595	-.307
sadist	-.752	.542	.376

- U prostoru ličnosti imamo sličan obrazac kao za drugu komponentu, ali sada imamo i nisku Savesnost
- U prostoru amorala, imamo visok sadizam, ali nisku frustraciju

# Primer u SPSS-u

- Kako znamo na koji način su kanoničke komponente povezane sa opozitnim skupom?
  - Na osnovu matrice kros-strukture
- Koje će biti dimenzije matrice kros-strukture u levom skupu, a koje u desnom?
  - U levom 6x3, u desnom 4x3
- U ispisu, matrica kros-strukture se zove matrica kanoničkih kros-faktora

# Primer u SPSS-u

Kanonički kros-faktori levog skupa varijabli			
	1	2	3
N	-.590	.027	-.049
E	.224	-.155	-.082
O	.113	-.201	-.104
A	.475	-.116	-.005
C	.571	.113	-.069

Kanonički kros-faktori desnog skupa varijabli			
	1	2	3
impulsi	-.700	-.044	.007
frustra	-.524	.206	-.056
sadist	-.531	.187	.069

- Vidimo da koeficijenti kros-strukture odražavaju relacije u matricama strukture
- Prosečna visina koeficijenata kros-strukture opada u svakoj sledećoj komponenti
- Zašto?
- Zato što su umnožak koeficijenta strukture i kanoničke korelacije, a kanoničke korelacije su sve manje i manje sa svakim sledećim parom

# Primer u SPSS-u

## Analiza prepokrivanja levog skupa

	Var.	Prepok.
CV1-1	.386	.193
CV1-2	.153	.018
CV1-3	.148	.005

## Analiza prepokrivanja desnog skupa

	Var.	Prepok.
CV2-1	.699	.349
CV2-2	.221	.026
CV2-3	.079	.003

Var. je varijansa, prepok. je redundansa

- Prva komponenta u prostoru ličnosti objašnjava 38.6% varijanse, dok je varijansa ličnosti objašnjena komponentom iz opozitnog skupa 19.3%
- Prva kanonička komponenta u prostoru amoralna objašnjava 69.9% varijanse, dok komponenta iz opozitnog skupa objašnjava 34.9% varijanse amoralna
- Ovo potencijalno ukazuje na to da smer uticaja ide iz ličnosti u amoral



# KKA uslovi za analizu

- Najmanje intervalni nivo merenja
- Multinormalna distribucija varijabli
- Linearne veze među varijablama (nasuprot nelinearnim)
- Linearna nezavisnost varijabli (varijable ne smeju biti linearni kompoziti drugih varijabli)
- Homoscedastet veza (varijansa greške ista na svim nivoima)
- Broj ispitanika višestruko veći od broja varijabli (bar 10-20 puta, čak do 60 puta)

# KKK dobre prakse

Kada se pravi prediktivna baterija testova, varijable (testovi) treba da budu izabrane tako da visoko koreliraju sa varijablama iz opozitnog skupa a nisko sa varijablama iz vlastitog skupa

# KKA - nedostaci

- KKA ne daje uvek rešenja koja zadovoljavaju principe jednostavne strukture (jedna varijabla može imati više visokih zasićenja/doprinosa)
  - Treba razmotriti sve varijable značajne za definisanje kanoničkih komponenti
- Maksimizuje se korelacija kanoničkih komponenti – čime se zanemaruje njihova varijansa, odnosno reprezentativnost za skup
  - Može se desiti da samo jedna varijabla ima visoke (skoro jedinične) projekcije na kanoničku komponentu - tada su i koeficijenti strukture i koeficijenti kros-strukture praktično jednaki izvornim korelacijama date varijable i ostalih varijabli iz jednog i drugog skupa
  - Dakle, ne dolazi ni do kakve generalizacije i sažimanja informacija
  - Pre interpretacije uvek treba proveriti da li u matrici kros-korelacija postoje parovi varijabli sa ekstremno visokim korelacijama

Hvala na pažnji!

Pitanja?