ODNOSI IZMEĐU DVE PROMENLJIVE

Ukrštene tabele i dijagram rasipanja

Predmet:

Alati za obradu podataka

Tema:

Vizuelizacija podataka

Od jedne ka dve promenljive

PROMENA ISTRAŽIVAČKOG PITANJA

JEDNA PROMENLJIVA

Koliko putovanja traje duže od 30 minuta?

Opisujemo raspodelu jedne karakteristike.

DVE PROMENLJIVE

Da li vreme putovanja zavisi od udaljenosti?

Za svaku opservaciju posmatramo par vrednosti.

Kod dve promenljive analiziramo kako se njihove vrednosti menjaju zajedno.

Izbor prikaza prema vrsti promenljivih

VRSTA PODATKA ODREĐUJE PRIKAZ

Promenljiva 1	Promenljiva 2	Preporučeni prikaz	Pitanje
Kategorijska	Kategorijska	Ukrštena tabela	Kako su kategorije povezane?
Kategorijska	Kategorijska	Grupisani stubovi	Kako se grupe porede?
Numerička	Numerička	Dijagram rasipanja	Da li se vrednosti menjaju zajedno?

Pre izbora grafikona odredite tip obe promenljive.

Ukrštena tabela

CROSS TABLE ILI CONTINGENCY TABLE

Ukrštena tabela prikazuje broj opservacija u svakoj kombinaciji kategorija dve promenljive.

Redovi

Kategorije prve promenljive, na primer vrsta investicije.

Kolone

Kategorije druge promenljive, na primer investitor.

Ćelija

Broj, iznos ili procenat za jednu kombinaciju reda i kolone.

Ukupno

Zbir reda i kolone olakšava proveru i poređenje grupa.

Podaci pre formiranja ukrštene tabele

JEDAN RED PREDSTAVLJA JEDNU KOMBINACIJU

Investitor	Vrsta investicije	Iznos
A	Akcije	50
A	Obveznice	30
A	Nekretnine	20
B	Akcije	20
B	Obveznice	50
B	Nekretnine	30
C	Akcije	35
C	Obveznice	20
C	Nekretnine	45

Tri kolone

Investitor identifikuje grupu. Vrsta investicije opisuje kategoriju. Iznos je vrednost koja se sabira.

Devet redova

Svaki investitor ima tri vrste ulaganja. Ukrštena tabela ih pretvara u preglednu matricu.

Ukrštena tabela ulaganja

MATRICA DVE KATEGORIJSKE PROMENLJIVE

Vrsta investicije	Investitor A	Investitor B	Investitor C	Ukupno
Akcije	50	20	35	105
Obveznice	30	50	20	100
Nekretnine	20	30	45	95
Ukupno	100	100	100	300

Ćelija 50 znači da je Investitor A u akcije rasporedio 50 jedinica.

Ukupne vrednosti po redovima

KOLIKO JE ULOŽENO U SVAKU VRSTU IMOVINE

Vrsta investicije	A	B	C	Ukupno reda
Akcije	50	20	35	105
Obveznice	30	50	20	100
Nekretnine	20	30	45	95

Računanje zbira

Akcije: 50 + 20 + 35 = 105

Obveznice: 30 + 50 + 20 = 100

Nekretnine: 20 + 30 + 45 = 95

Redni zbir odgovara jednoj vrsti investicije.

Procenti omogućavaju poštenije poređenje

PROCENAT UNUTAR SVAKE KOLONE

Vrsta investicije	Investitor A	Investitor B	Investitor C
Akcije	50%	20%	35%
Obveznice	30%	50%	20%
Nekretnine	20%	30%	45%
Ukupno	100%	100%	100%

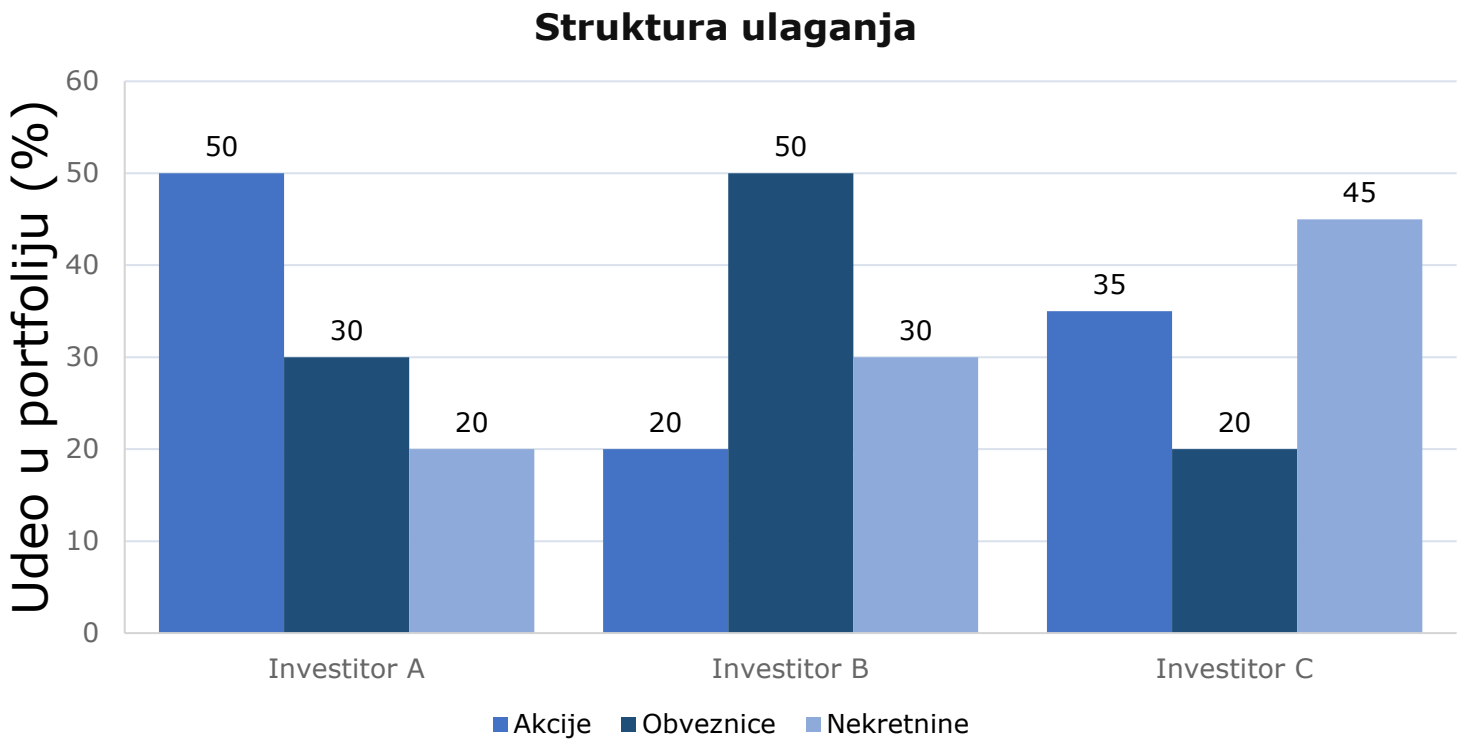
Procenat u koloni = vrednost ćelije ÷ zbir kolone

Primer: 50 ÷ 100 = 50%

Investitor B najveći deo portfolija drži u obveznicama.

Grupisani stubičasti dijagram

POREĐENJE KATEGORIJA IZMEĐU INVESTITORA



Unutar investitora

Investitor A najveći deo ulaže u akcije.

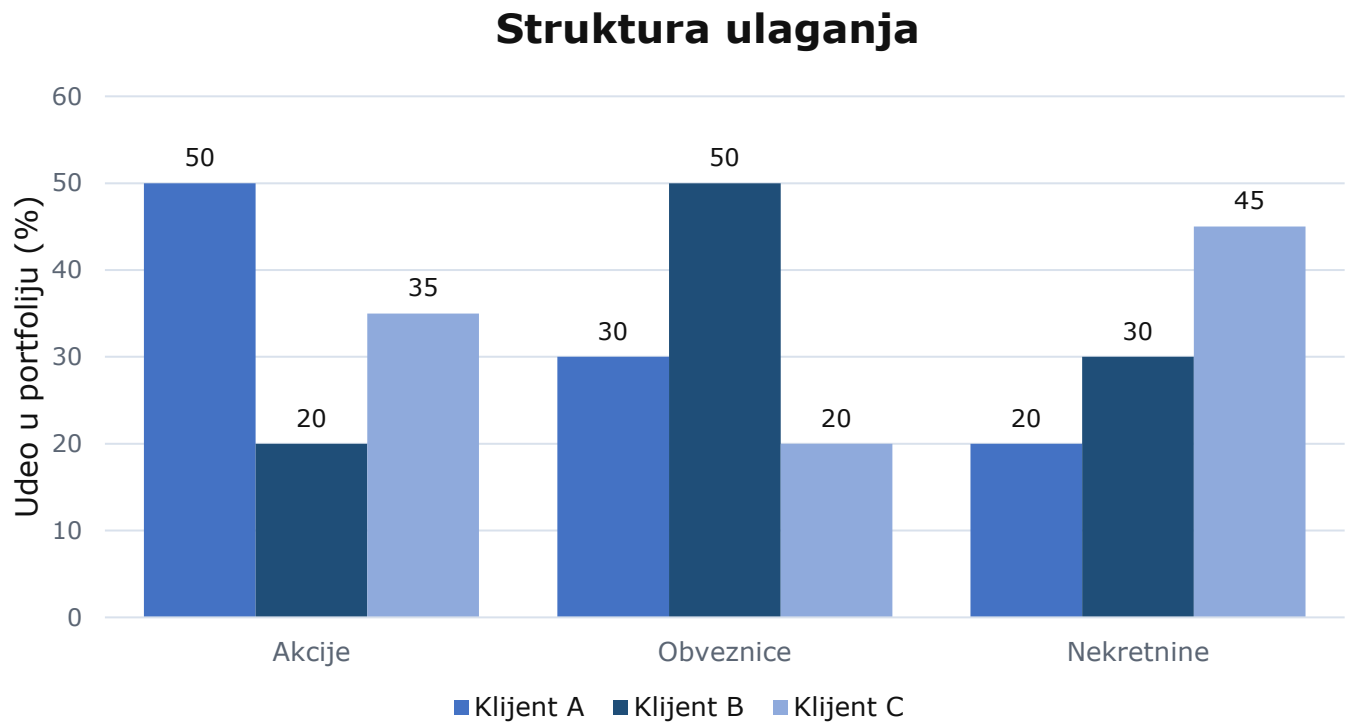
Između investitora

Investitor C ima najveći udeo nekretnina.

Stubovi iste grupe pripadaju istom investitoru.

Grupisani stubičasti dijagram

POREĐENJE KLIJENATA UNUTAR VRSTE INVESTICIJE



Akcije i obveznice

Klijent A ima najveći udeo akcija, dok Klijent B najveći udeo ulaže u obveznice.

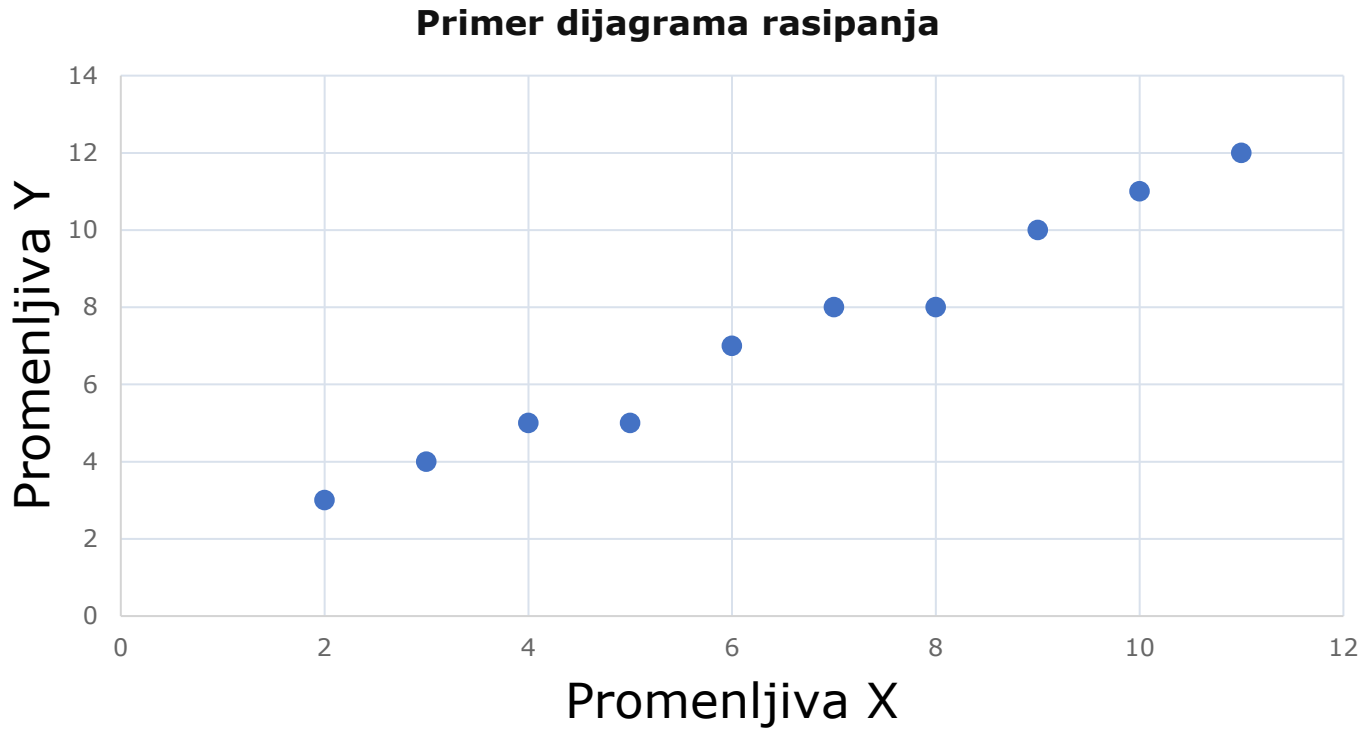
Nekretnine

Klijent C ima najveći udeo nekretnina.

Stubovi iste grupe pripadaju istoj vrsti investicije.

Dijagram rasipanja

SCATTER PLOT ZA DVE NUMERIČKE PROMENLJIVE



Jedna tačka

Jedna opservacija sa dve numeričke vrednosti.

Položaj tačke

X određuje položaj levo-desno, a Y položaj dole-gore.

Posmatramo obrazac svih tačaka, a ne jednu tačku izdvojeno.

Primer testa: dve numeričke promenljive

KOORDINATE JEDNE OPSERVACIJE

Student	Rezultat iz čitanja	Rezultat iz pisanja
1	470	450
2	520	540
3	610	590
Ana	550	300

Horizontalna osa X

Rezultat iz čitanja, od 200 do 800 poena.

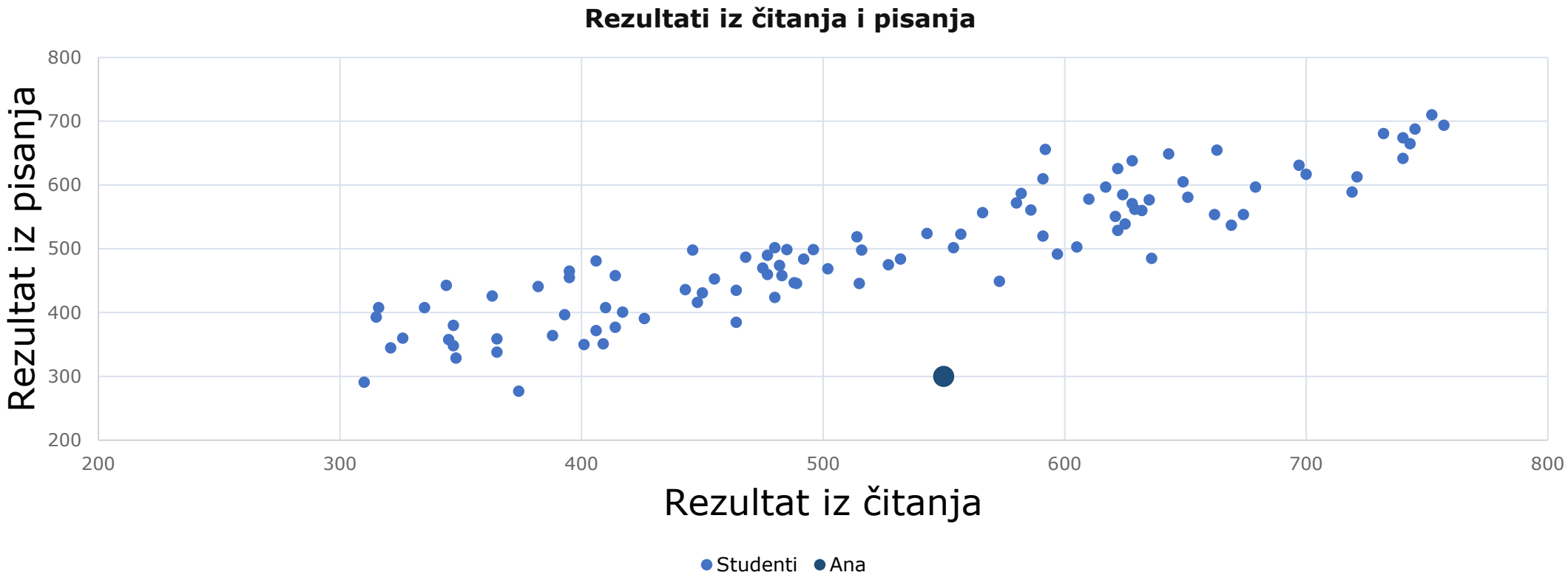
Vertikalna osa Y

Rezultat iz pisanja, od 200 do 800 poena.

Student 2 se prikazuje tačkom (520, 540).

Ilustrativni rezultati testa za 100 studenata

SVAKA TAČKA PREDSTAVLJA JEDNOG STUDENTA



Rastući oblak tačaka pokazuje pozitivnu povezanost. Ana je izdvojena tačka.

Opservacija i promenljive na dijagramu

JEDAN RED POSTAJE JEDNA TAČKA

Opservacija	X: rezultat iz čitanja	Y: rezultat iz pisanja
Ana	550	300

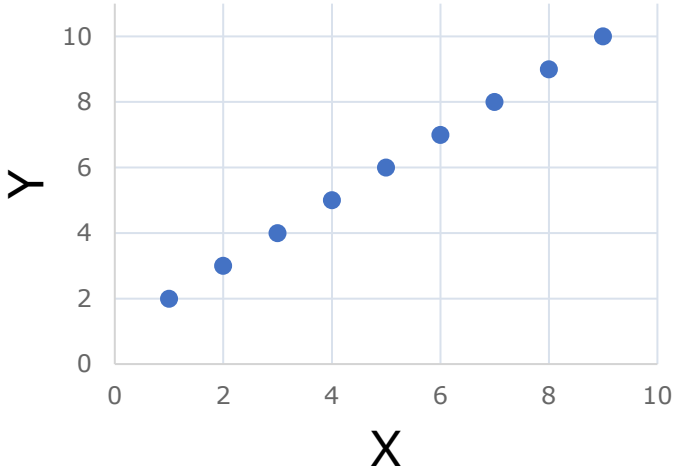


Ana je jedna opservacija. Rezultati iz čitanja i pisanja su dve promenljive.

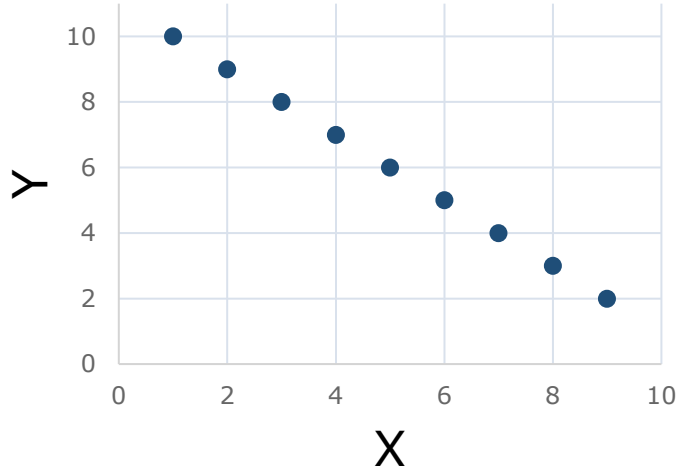
Smer veze

DA LI Y RASTE, OPADA ILI NEMA JASAN OBRAZAC KADA X RASTE

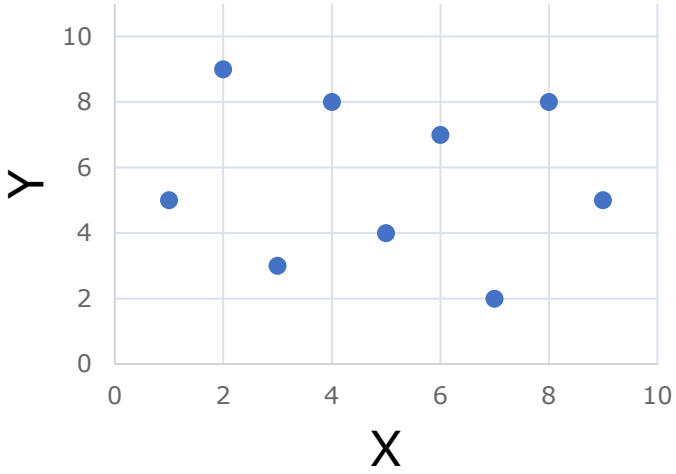
Pozitivna veza



Negativna veza



Bez jasne veze

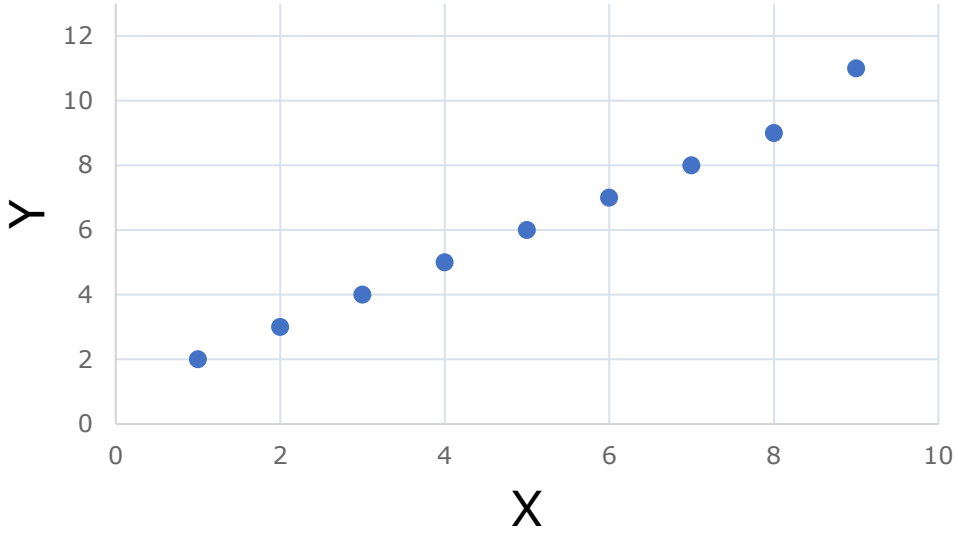


Smer opisuje opšti pravac oblaka tačaka.

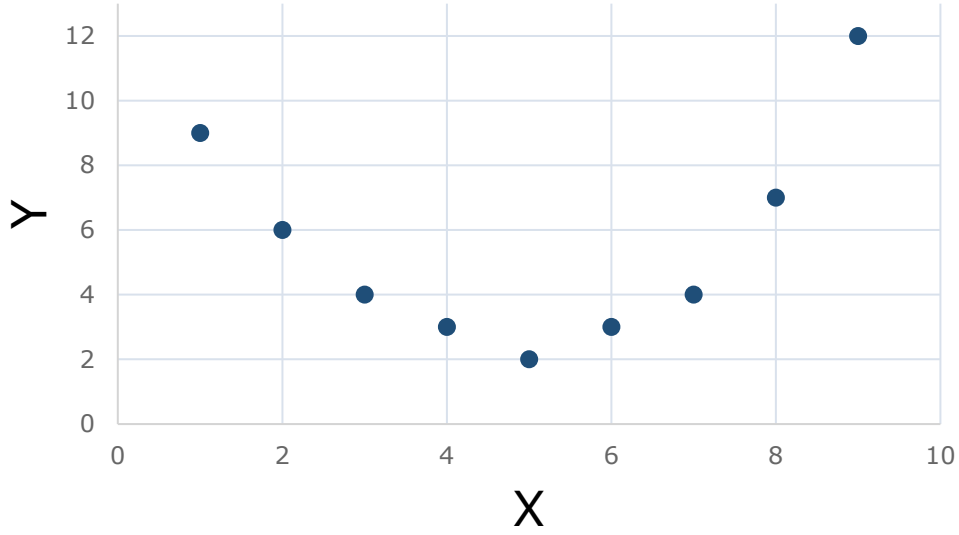
Jačina i oblik veze

BLIZINA TAČAKA OBRASCU

Jaka linearna veza



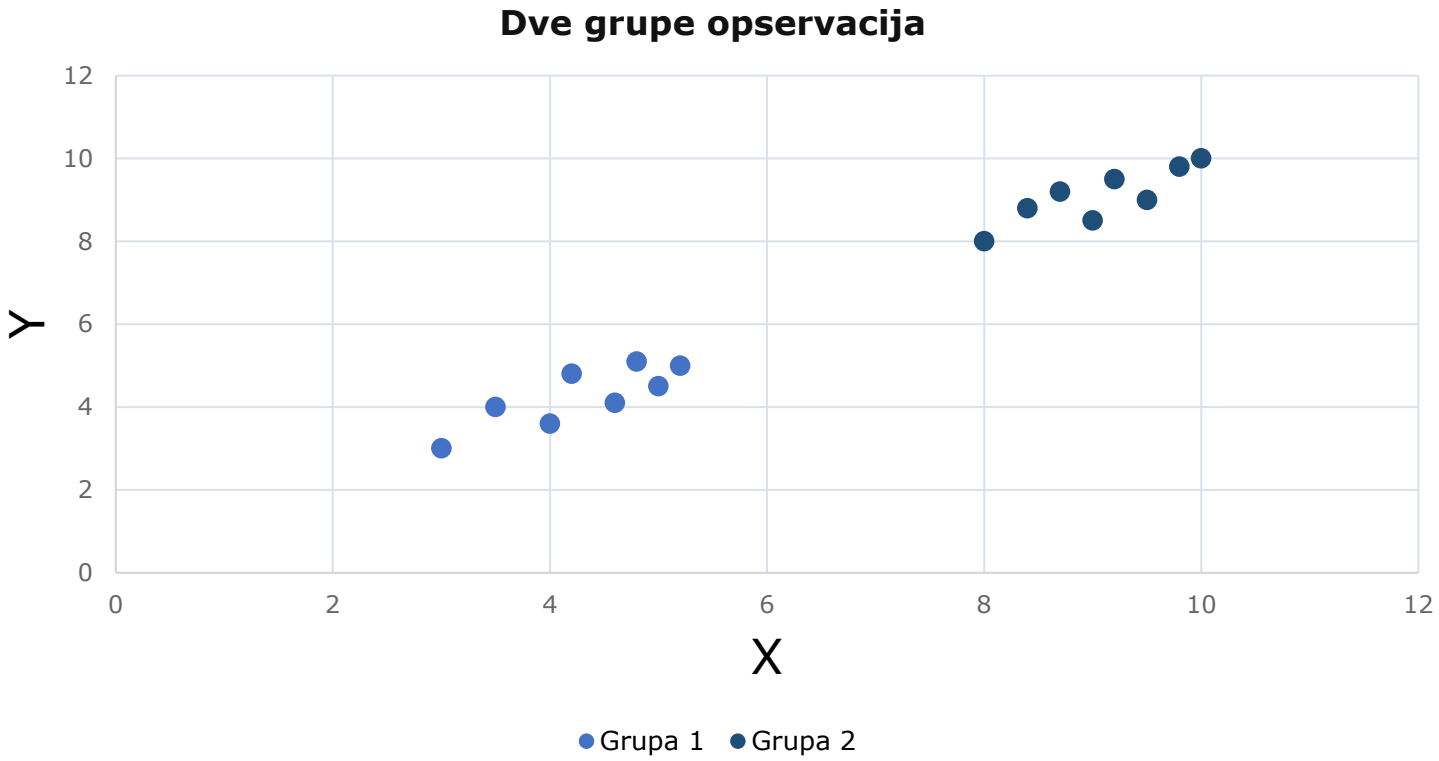
Nelinearan obrazac



Veza može biti jaka i kada nije pravolinijska.

Koncentracija i klasteri

ODVOJENE GRUPE U OBLAKU TAČAKA



Koncentracija

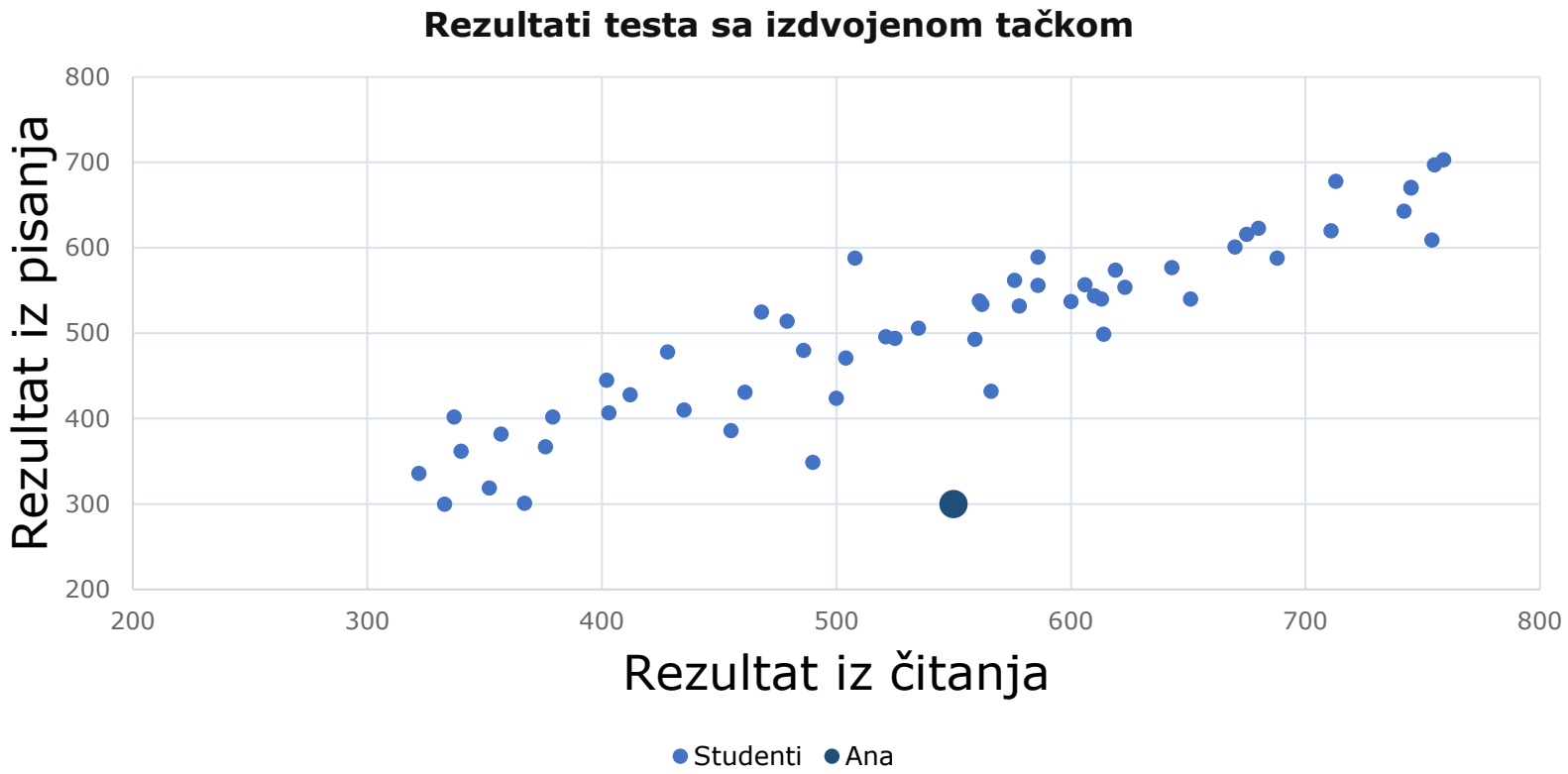
Područje sa mnogo tačaka pokazuje gde se nalazi veliki deo opservacija.

Klaster

Odvojena grupa može ukazati na različite podgrupe ili dodatnu promenljivu.

Izdvojena vrednost

OUTLIER ODSTUPA OD OBRASCA VEĆINE PODATAKA



Ana: (550, 300)

Rezultat iz čitanja je iznad sredine skupa, dok je rezultat iz pisanja znatno ispod obrasca.

Proveriti unos i kontekst. Izdvojenu tačku ne treba automatski brisati.

Povezanost ne dokazuje uzrok

KORELACIJA I UZROČNOST

Dijagram može pokazati da se dve promenljive menjaju zajedno. Sam prikaz ne dokazuje da jedna izaziva drugu.

Primer

Tokom toplijih dana mogu rasti prodaja sladoleda i broj poseta bazenu.

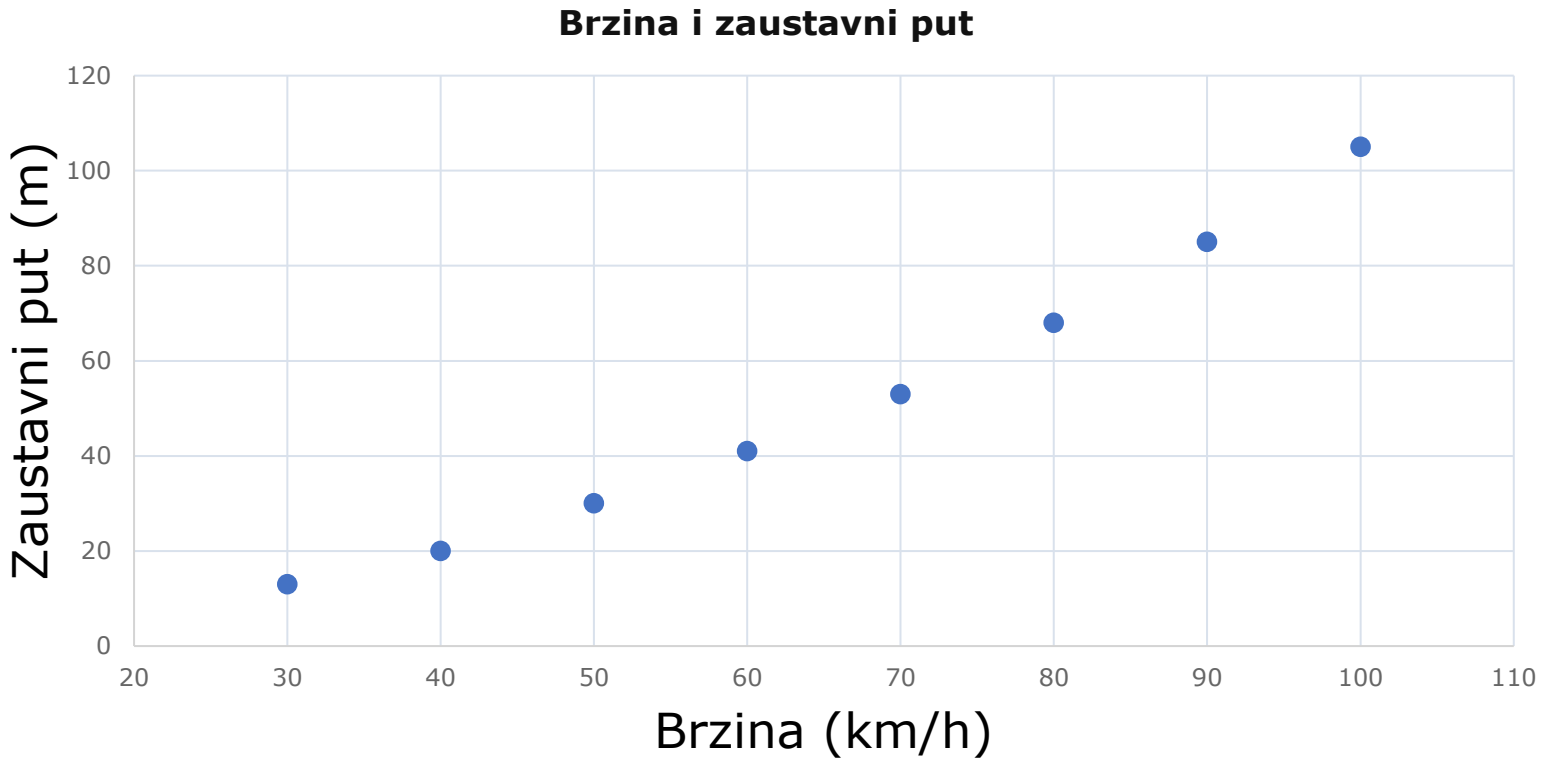
Treća promenljiva

Temperatura može uticati na obe pojave. Njihova povezanost zato ne dokazuje uzrok.

Za tvrdnju o uzroku potreban je odgovarajući plan istraživanja i kontrola drugih faktora.

Primer iz saobraćaja: brzina i zaustavni put

RASTUĆI I ZAKRIVLJEN OBRAZAC



Šta vidimo

Veća brzina prati duži zaustavni put. Obrazac postaje strmiji pri većim brzinama.

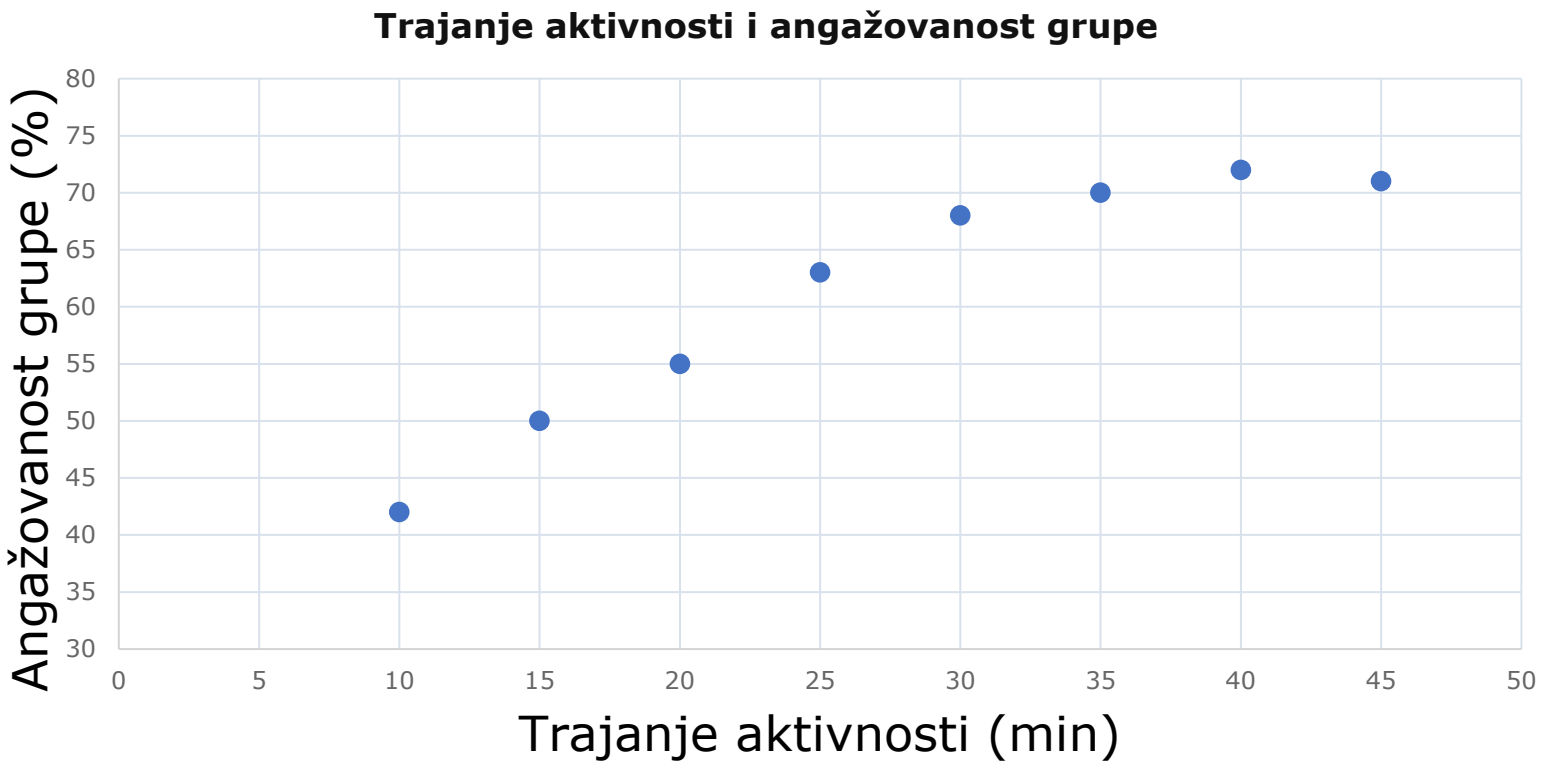
Moguća odluka

Podaci mogu podržati analizu ograničenja brzine i bezbedne udaljenosti.

Ilustrativni podaci

Primer za vaspitače: trajanje aktivnosti i angažovanost

RAST DO TAČKE ZASIĆENJA



Šta vidimo

Angažovanost raste do približno 40 minuta, a zatim se rast zaustavlja.

Granica zaključka

Prikaz se odnosi na grupne aktivnosti. Ne koristi se za rangiranje pojedinačne dece.

Ilustrativni podaci

Kako izabrati odgovarajući prikaz

OD PITANJA DO GRAFIKONA

Pitanje	Vrste promenljivih	Prikaz
Da li se kategorije pojavljuju zajedno?	Kategorijska + kategorijska	Ukrštena tabela
Kako se kategorije porede među grupama?	Kategorijska + kategorijska	Grupisani stubovi
Da li dve veličine rastu ili opadaju zajedno?	Numerička + numerička	Dijagram rasipanja
Gde su koncentracije, klasteri i izdvojene tačke?	Numerička + numerička	Dijagram rasipanja

Prvo pitanje je: koje vrste podataka imam?

Use case: analiza kašnjenja gradskih linija

OD EVIDENCIJE DO ODLUKE

Korak	Šta radimo	Rezultat
1	Beležimo liniju, period dana i status kašnjenja.	Uređena tabela podataka
2	Formiramo tabelu period dana × status kašnjenja.	Broj i procenat kašnjenja
3	Pravimo grupisani stubičasti dijagram.	Poređenje perioda
4	Izdvajamo linije i periode sa najvećim udelom kašnjenja.	Prioritet za detaljnu analizu
5	Dodajemo opterećenje i vreme vožnje.	Dijagram rasipanja i nova pitanja

Ukrštena tabela locira problem. Dijagram rasipanja zatim ispituje numeričke odnose.

Najvažnije za pamćenje

ČETIRI KLJUČNE IDEJE

1. Dve kategorijske promenljive prikazujemo ukrštenom tabelom. Grupisani stubovi olakšavaju poređenje.

2. Dve numeričke promenljive prikazujemo dijagramom rasipanja. Jedna tačka predstavlja jednu opservaciju.

3. Kod rasipanja posmatramo smer, jačinu, oblik, koncentraciju, klastere i izdvojene tačke.

4. **Vidljiva povezanost sama po sebi ne dokazuje uzročno-posledičnu vezu.**

Vrsta podataka određuje prikaz. Pitanje određuje tumačenje.

Pitanje za proveru znanja 1

IZABERITE JEDAN TAČAN ODGOVOR

1. Koji prikaz je najpogodniji za dve kategorijske promenljive?

- A) Histogram
- B) Ukrštena tabela
- C) Dijagram rasipanja
- D) Linijski dijagram

Tačan odgovor i objašnjenje nalaze se u Notes delu slajda.

Pitanje za proveru znanja 2

IZABERITE JEDAN TAČAN ODGOVOR

2. Šta predstavlja jedna tačka na dijagramu rasipanja?

- A) Jednu promenljivu
- B) Jednu kategoriju
- C) Jednu opservaciju sa dve vrednosti
- D) Ukupan broj opservacija

Jedna tačka povezuje X i Y vrednost iste opservacije.

Pitanje za proveru znanja 3

IZABERITE JEDAN TAČAN ODGOVOR

3. Šta pokazuje zbir jednog reda u ukrštenoj tabeli ulaganja?

- A) Ukupan iznos jednog investitora
- B) Ukupan iznos jedne vrste investicije
- C) Prosečnu vrednost svih ćelija
- D) Broj kolona u tabeli

Redni zbir objedinjuje vrednosti iste vrste investicije.

Pitanje za proveru znanja 4

IZABERITE JEDAN TAČAN ODGOVOR

4. Koji zaključak je ispravan kada tačke prate rastući obrazac?

- A) X sigurno izaziva Y
- B) Promenljive pokazuju pozitivnu povezanost
- C) Promenljive su kategorijske
- D) Sve tačke imaju istu vrednost

Obratite pažnju na razliku između povezanosti i uzroka.

Pitanje za proveru znanja 5

IZABERITE JEDAN TAČAN ODGOVOR

5. Student ima rezultat iz čitanja 620 i rezultat iz pisanja 580. Gde se nalazi njegova tačka?

- A) X = 580, Y = 620
- B) X = 620, Y = 580
- C) X = 620, Y = 620
- D) Položaj se ne može odrediti

Dodatno objasnite šta predstavlja X osa, a šta Y osa.