

Korelacija

Teorijske osnove

Začetnikom korelacijske i regresijske analize smatra se engleski antropolog Francis Galton koji je pod utjecajem rođaka Charlesa Darwina istraživao utjecaj naslijeđa na razvoj čovjekovih karakteristika.

Suradujući s Galtonom, Karl Pearson je razvio računski postupak za utvrđivanje povezanosti između dviju varijabli i nazvao ga *produkt -moment koeficijent korelacije*.

Pearsonov *produkt - moment koeficijent korelacije* (r) predstavlja mjeru međusobne linearne povezanosti rezultata dviju standardiziranih varijabli, a izračunava se formulom

$$r = \frac{\sum_{i=1}^n x_{ci}y_{ci}}{\sqrt{\sum_{i=1}^n x_{ci}^2} \cdot \sqrt{\sum_{i=1}^n y_{ci}^2}}$$

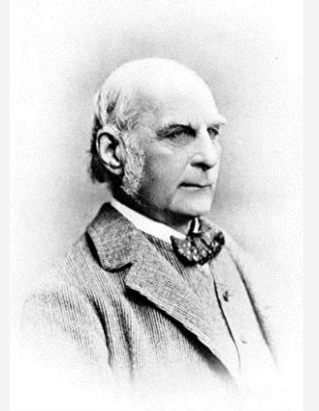
gdje je

$x_{ci} = x_i - \bar{x}$ (centrirani rezultat entiteta i u varijabli x)

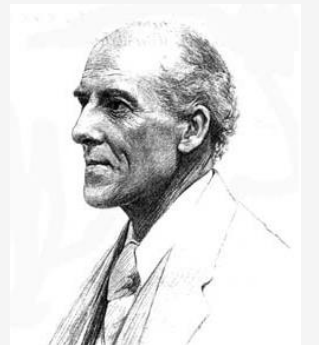
\bar{x} - aritmetička sredina varijable x

$y_{ci} = y_i - \bar{y}$ (centrirani rezultat entiteta i u varijabli y)

\bar{y} - aritmetička sredina varijable y



Francis Galton
(1822. – 1911.)



Karl Pearson
(1857. – 1936.)

Teorijske osnove

Pearsonov koeficijent korelacije može se izračunati i iz originalnih (necentriranih) rezultata pomoću formule

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}\right) \left(\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}\right)}}$$

gdje je

x_i - rezultat entiteta i u varijabli x

y_i - rezultat entiteta i u varijabli y

n - broj entiteta

Ako se rezultati ispitanika u varijablama x i y standardiziraju, onda formula poprima sljedeći oblik

$$r = \frac{\sum_{i=1}^n (z_{x_i} z_{y_i})}{n}$$

gdje je

z_{x_i} - standardizirani rezultat entiteta i u varijabli x

z_{y_i} - standardizirani rezultat entiteta i u varijabli y

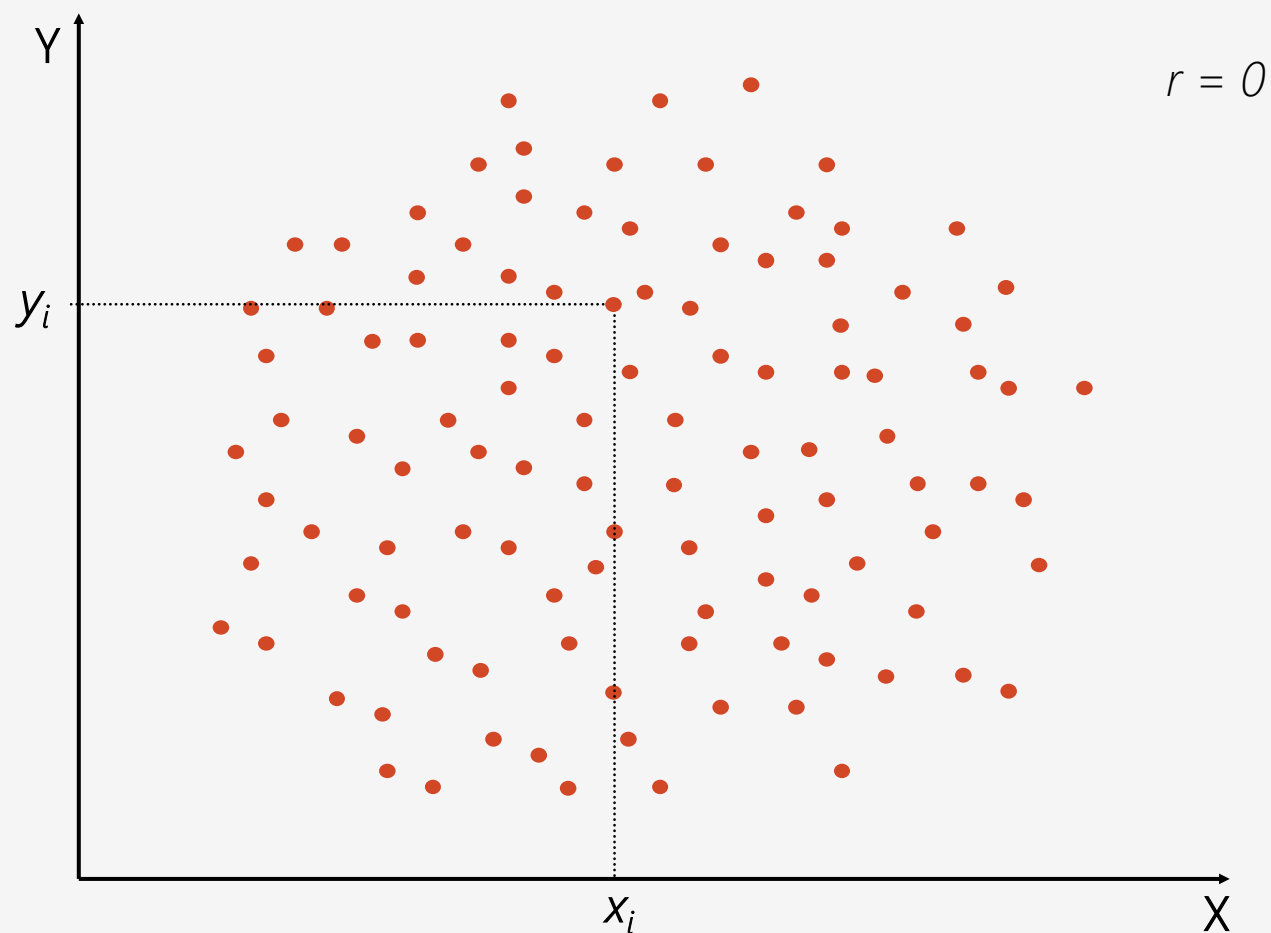
n - broj entiteta

Pearsonov koeficijent korelacije (izračunat bilo kojim postupkom) uvijek se nalazi u intervalu od -1 do 1 .

Teorijske osnove

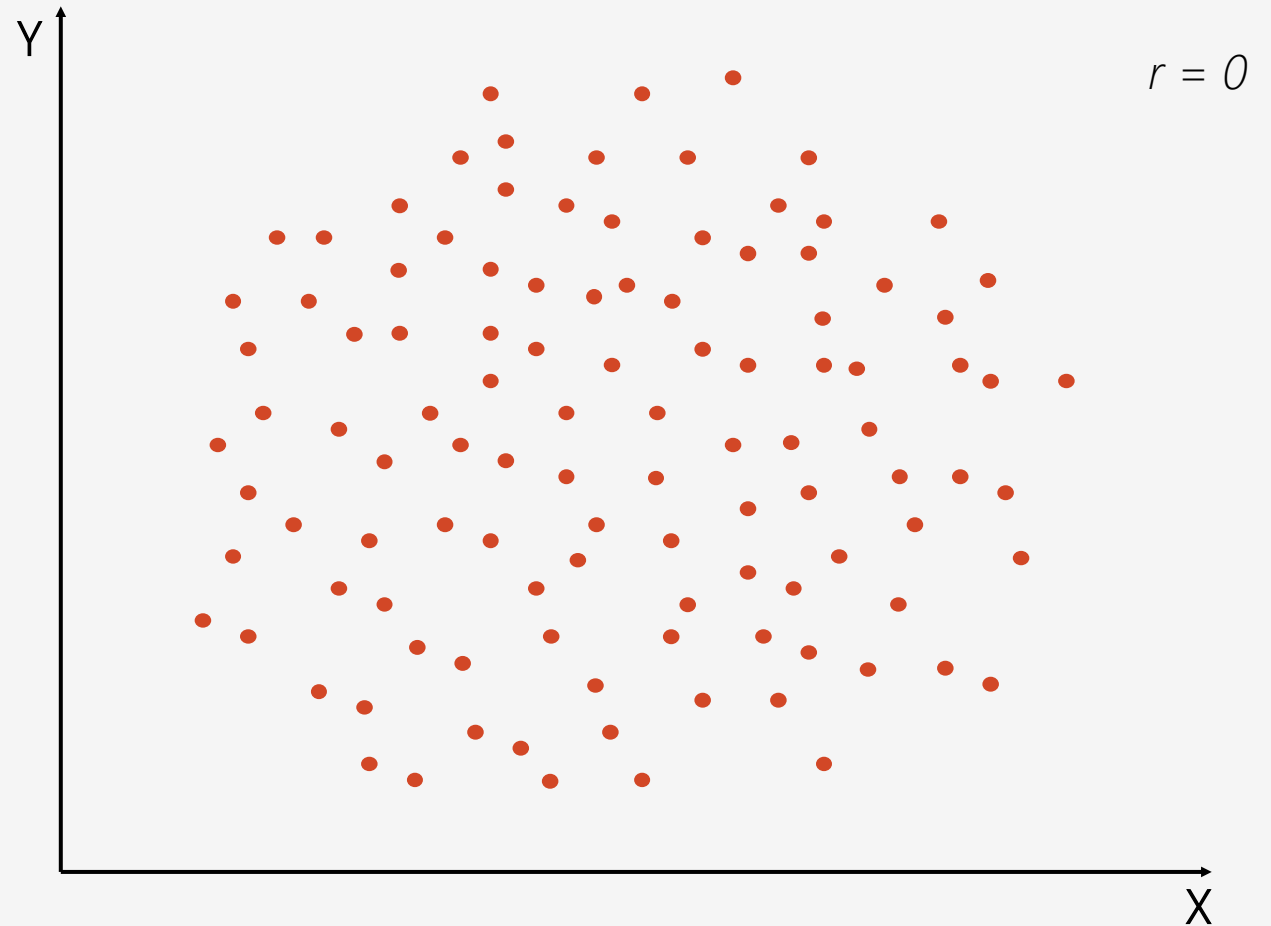
Dvodimenzionalni *korelacijski dijagram* je grafički način prikazivanja povezanosti rezultata dviju varijabli, a iscrtava se na način da se za svakog ispitanika odredi položaj u koordinatnom sustavu pri čemu se položaj na apscisi određuje sukladno rezultatu ispitanika u jednoj varijabli, a položaj na ordinati sukladno rezultatu u drugoj varijabli.

Oblik korelacijskog dijagrama zavisn je o smjeru i veličini povezanosti varijabli, odnosno o predznaku i veličini koeficijenta korelacije.



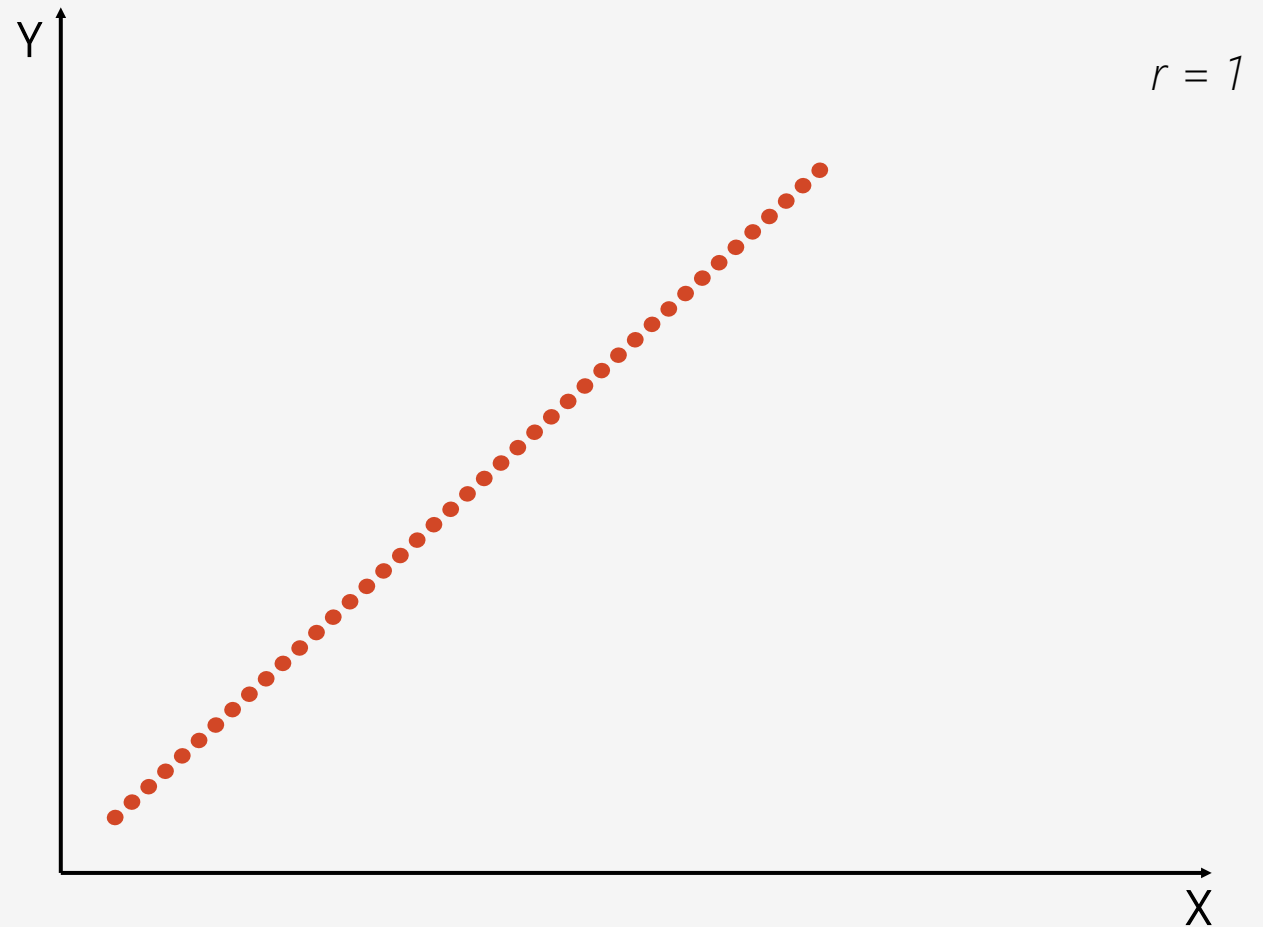
Teorijske osnove

Nulta korelacija ($r=0$) označava takav odnos između dviju varijabli u kojem svakom rezultatu u jednoj varijabli može odgovarati bilo koji rezultat u drugoj varijabli.



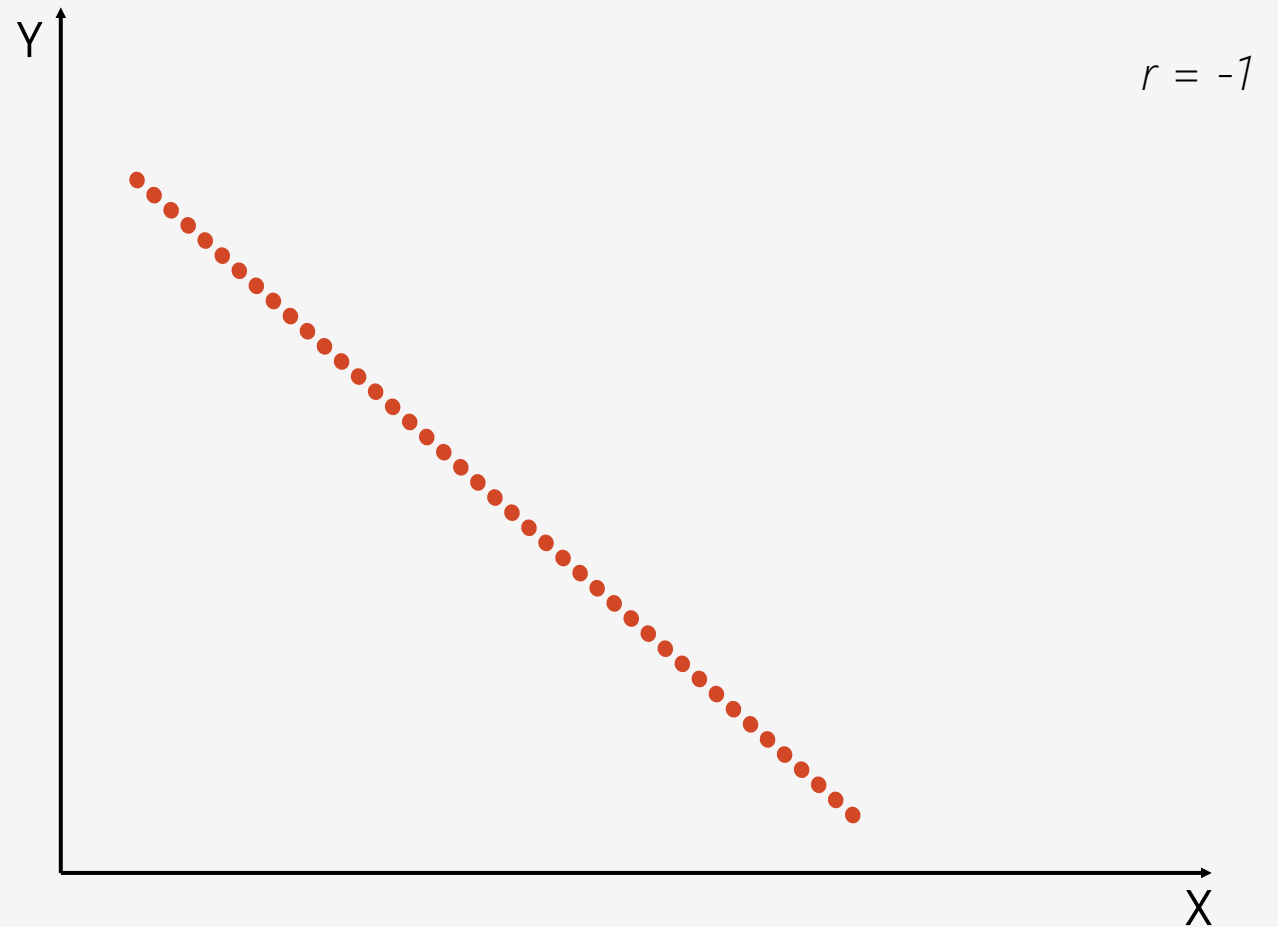
Teorijske osnove

Potpuna pozitivna korelacija ($r=1$) označava takav odnos između dviju varijabli u kojem svakom iznadprosječnom rezultatu u jednoj varijabli odgovara jednako iznadprosječan rezultat u drugoj varijabli, odnosno svakom ispodprosječnom rezultatu u jednoj varijabli odgovara jednako ispodprosječan rezultat u drugoj varijabli.



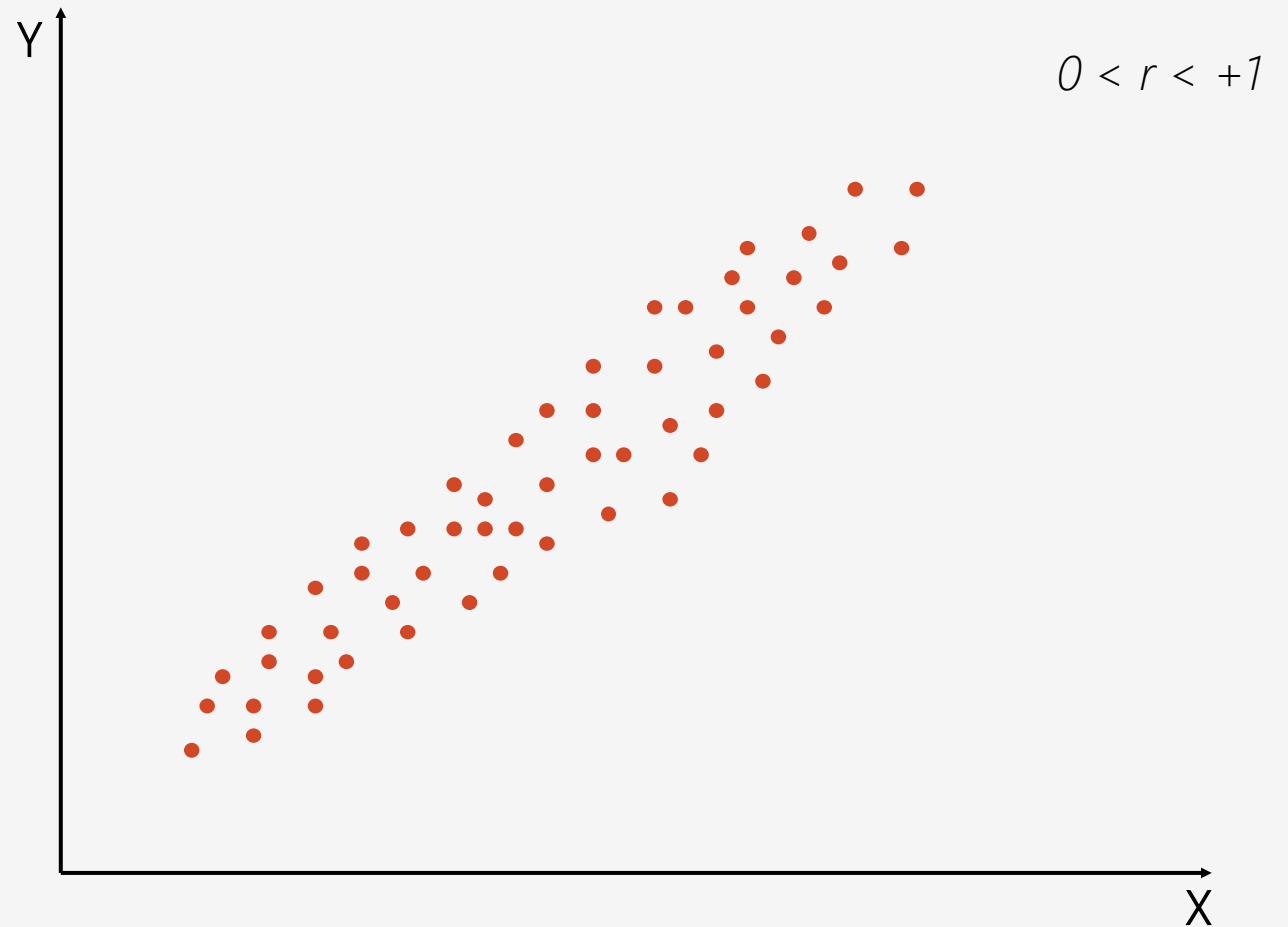
Teorijske osnove

Potpuna negativna korelacija ($r=-1$) označava takav odnos između dviju varijabli u kojem svakom iznadprosječnom rezultatu u jednoj varijabli odgovara jednako ispodprosječan rezultat u drugoj varijabli, odnosno svakom ispodprosječnom rezultatu u jednoj varijabli odgovara jednako iznadprosječan rezultat u drugoj varijabli.



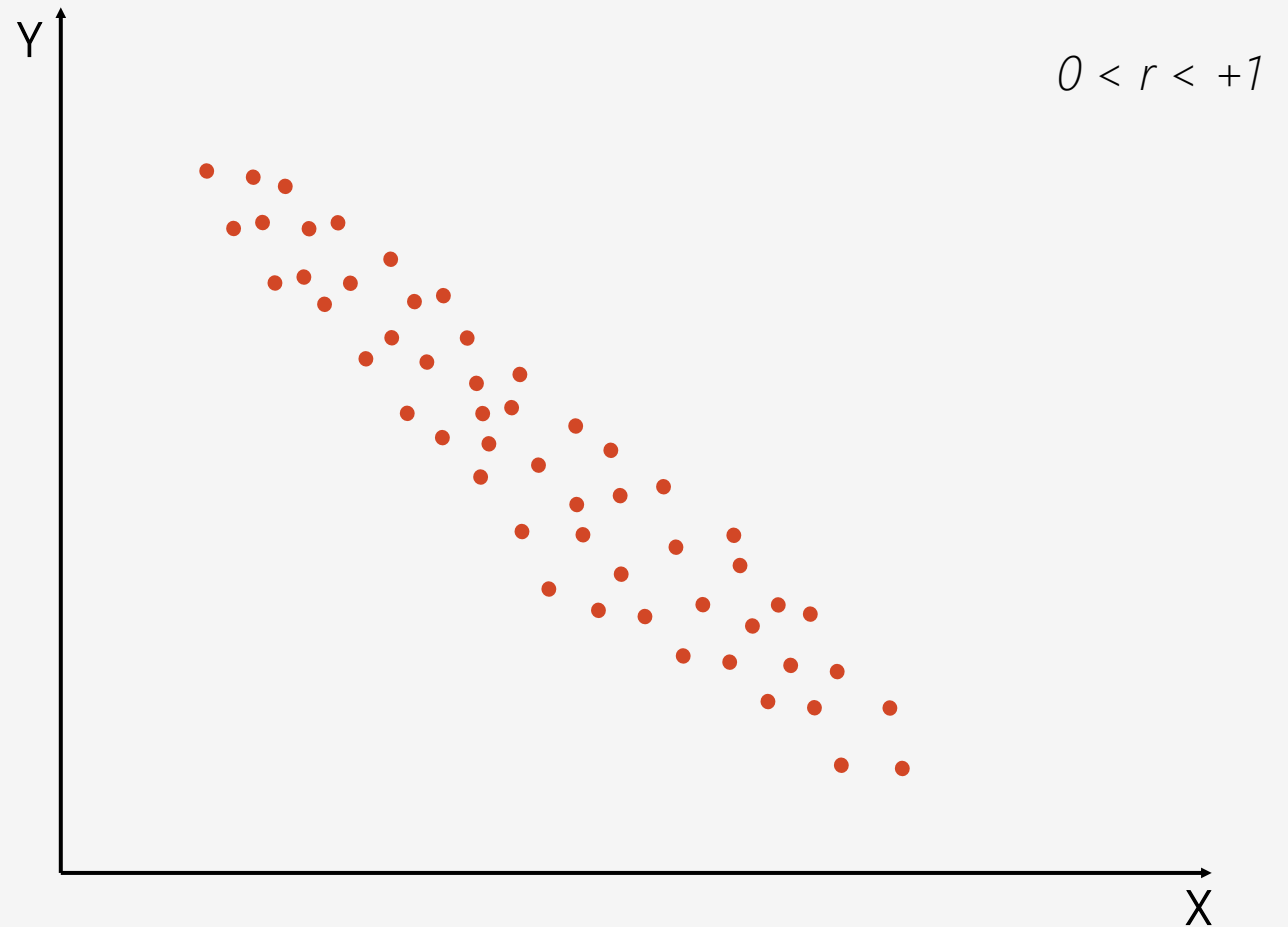
Teorijske osnove

Nepotpuna pozitivna korelacija ($0 < r < 1$) označava takav odnos između dviju varijabli u kojem svakom iznadprosječnom rezultatu u jednoj varijabli najvjerojatnije odgovara jednako iznadprosječan rezultat u drugoj varijabli, odnosno svakom ispodprosječnom rezultatu u jednoj varijabli najvjerojatnije odgovara jednako ispodprosječan rezultat u drugoj varijabli.



Teorijske osnove

Nepotpuna negativna korelacija ($0 > r > -1$) označava takav odnos između dviju varijabli u kojem svakom iznadprosječnom rezultatu u jednoj varijabli najvjerojatnije odgovara jednako ispodprosječan rezultat u drugoj varijabli, odnosno svakom ispodprosječnom rezultatu u jednoj varijabli najvjerojatnije odgovara jednako iznadprosječan rezultat u drugoj varijabli.



Teorijske osnove

Ako se povezanost između dviju varijabli utvrđuje na uzorku ispitanika, potrebno je testirati statističku značajnost koeficijenta korelacije odnosno utvrditi vjerojatnost da se korelacija nije dogodila slučajno. Pri testiranju statističke značajnosti koeficijenta korelacije moguće je postaviti sljedeće hipoteze

$H_0: r=0$ - korelacija nije statistički značajna uz pogrešku p

$H_1: r \neq 0$ - korelacija je statistički značajna uz pogrešku p

Statistička značajnost koeficijenta korelacije testira se putem t-distribucije pri čemu se kritična t vrijednost određuje na temelju pogreške statističkog zaključka p i broja stupnjeva slobode $df=n-2$. Vrijednost koja se uspoređuje s kritičnom t-vrijednosti izračunava se formulom

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

gdje je

t - vrijednost koja se distribuira prema t-distribuciji za $df=n-2$

r - koeficijent korelacije

n - broj entiteta

Iz prethodno navedene formule moguće je uočiti kako je vjerojatnost da se korelacija u uzorku dogodila slučajno iako u populaciji ne postoji manja što je broj entiteta u uzorku veći i što je apsolutna vrijednost izračunate korelacije veća.

Teorijske osnove

Rezultati korelacijske analize najčešće se prikazuju pomoću *korelacijske matrice*. U dijagonali korelacijske matrice su varijance varijabli, a izvandijagonalni elementi su korelacije svake varijable sa svakom. Korelacijska matrica je simetrična što znači da se koeficijent korelacije između svake dvije varijable nalazi i s gornje i s donje strane glavne dijagonale.

Formalni prikaz korelacijske matrice s četiri varijable

	v1	v2	v3	v4
v1	1	$r_{v1,v2}$	$r_{v1,v3}$	$r_{v1,v4}$
v2	$r_{v2,v1}$	1	$r_{v2,v3}$	$r_{v2,v4}$
v3	$r_{v3,v1}$	$r_{v3,v2}$	1	$r_{v3,v4}$
v4	$r_{v4,v1}$	$r_{v4,v2}$	$r_{v4,v3}$	1

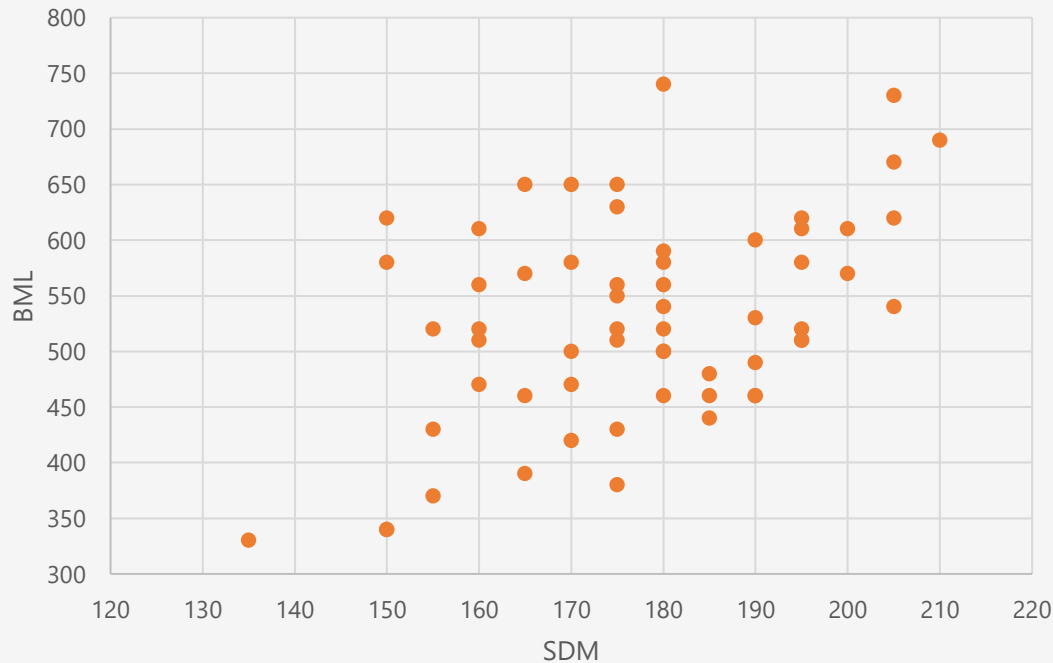
Primjer: U korelacijskoj matrici prikazani su Pearsonovi koeficijenti korelacije između četiri motorička testa. Označene korelacije su statistički značajne uz pogrešku $p=0,05$)

	ONT	OUZ	NEB	SKL
ONT	1	0,66*	-0,36*	-0,35*
OUZ	0,66*	1	-0,21	-0,62*
NEB	-0,36*	-0,21	1	0,15
SKL	-0,35*	-0,62*	0,15	1

Microsoft Excel



Zadatak 1: U datoteci *JUDO3F.xls* izračunajte koliki je Pearsonov koeficijent korelacije između varijabli *SDM* i *BML* te kreirajte korelacijski dijagram (*Scatter plot*).



- 1 Izračunavanje Pearsonovog koeficijenta korelacije vrši se pomoću funkcije **PEARSON**. Funkcija se unosi u označeno polje odabirom opcije *Function...* Putem traka *Array1* i *Array2* potrebno je definirati niz podataka prve, odnosno druge varijable.

