REGRESIONA I KORELACIONA ANALIZA

Reč *regresija* dospela je u statistiku kada je 1855.godine Fransis Galton objavio publikaciju u kojoj je analizirao visinu sinova u zavisnosti od visine očeva. Zaključak ove studije bio je da sinovi ekstremno visokih očeva nisu toliko visoki, dakle *regresiraju.*

Promena jednog obeležja statističkog skupa često utiče na promenu drugih obeležja zbog međusobne povezanosti. Povezanost između obeležja može se razlikovati i po smeru i po jačini povezanosti. Najjača ili najuža veza između obeležja je *funkcionalna veza,* tj. Takva veza da svakoj vrednosti jednog obeležja odgovara tačno određena vrednost drugog. Labavija veza između obeležja, koja su podložna manjim ili većim odstupanjima, naziva se *korelativnom (ili stohastičkom) vezom.*

Obično se jedna slučajno promenljiva identifikuje kao nezavisna (x), a druga kao zavisno slučajno promenljiva (y).

Skup statističkih metoda kojima se proučavaju uzajamne veze statističkih obeležja i pojava (smer, jačina, oblik) naziva se *teorijom korelacije*, a osnovni pokazatelji korelacionih veza su *jednačina regresije* i *koeficijent korelacije*.

Ispitivanje zavisnosti u statističkoj analizi ima dva osnovna pravca:

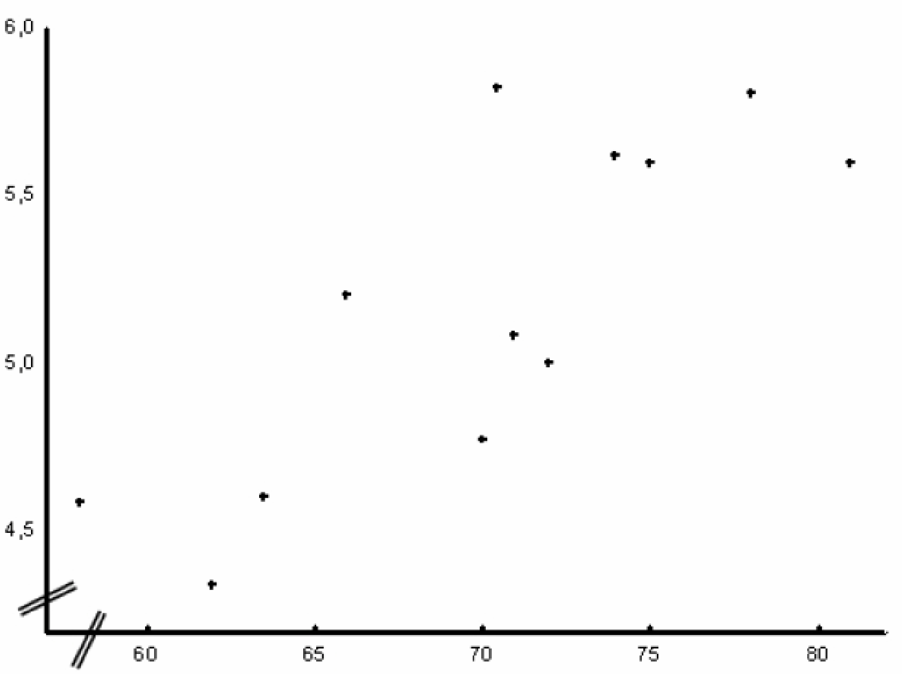
1. *oblik* zavisnosti koji ispituje **regresiona analiza**
2. *jačinu* zavisnosti koju određuje **korealicona analiza**

U medicinskim istraživanjima najčešće se sreće lineaarni model regresione i korelacione analize, pa će se naša razmatranja odnositi na taj model.

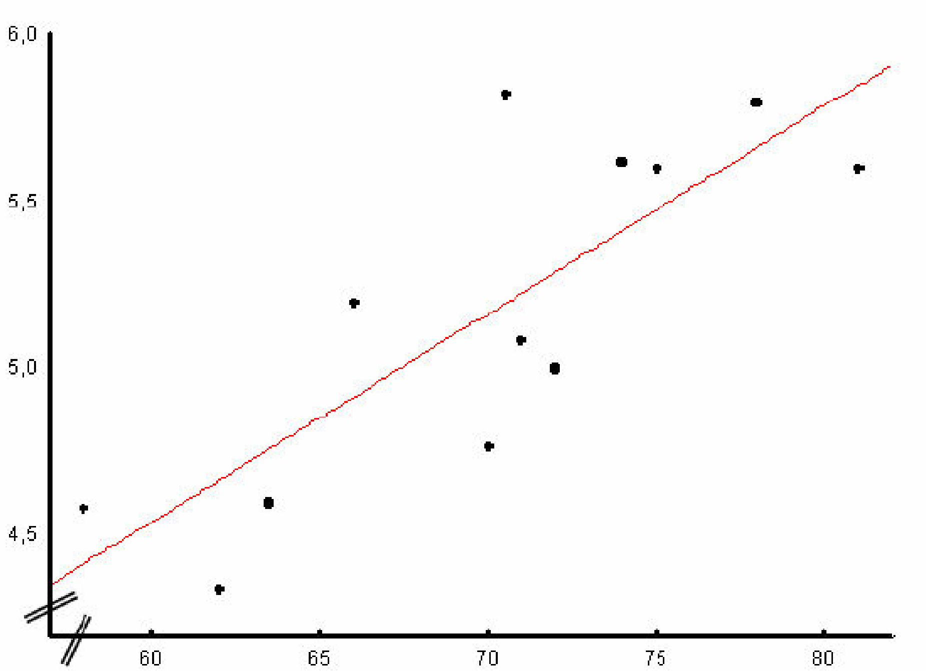
REGRESIONA ANALIZA

Regresiona analiza pokazuje *oblik* povezanosti između dve promenljive pomoću *regresione linije*.

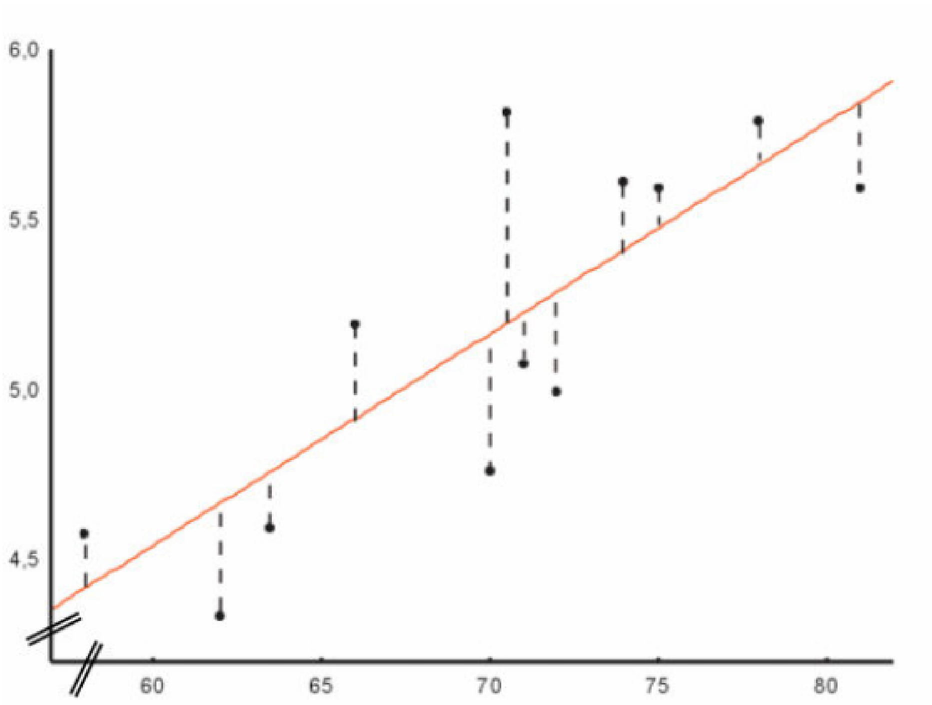
Odnos promenljive (y) prema promenljivoj (x) može biti različit, i zato je prvi korak ka otkrivanju oblika povezanosti ucrtati *dijagram rasturanja* ili *dijagram disperzije* između dva obeležja.



Da bi smo kvantifikovali približnu linearnu vezu između te dve veličine, možemo konstruisati pravac koji najbolje opisuje podatke. Intuitivno bi to učinili tako da je približno jednak broj tačaka iznad pravca i ispod njega.



Postoji egzaktan matematički način kojim se prikazuje najbolje prilagođen pravac linearne veze. Određuje se iz uslova da je zbir kvadrata vertikalnih udaljenosti tačaka od od pravca najmanja – **metoda najmanjih kvadrata.** Tako određen pravac povezanosti između dve varijable prikazuje se *regresionom linijom*.



Regresiona linija izražava se *jednačinom regresije*:

$$\mathbf{y = a + b \cdot x,}$$

gde je:

y – zavisno promenljiva,

x - nezavisno promenljiva,

a – regresiona konstanta,

b – koeficijent regresije.

Zavisno promenljiva y je nepoznata promenljiva koja se izračunava na osnovu vrednosti nezavisne promenljive x koja je poznata.

Regresiona konstanta (a) i koeficijent regresije (b) određeni pomoću *metoda najmanjih kvadrata* imaju formule:

$$a = \bar{y} - b \cdot \bar{x}$$

$$b = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{n \cdot \sum x^2 - \left(\sum x\right)^2}$$

Sa n je u jednačini označenukupan broj parova koji po nekim autorima ne bi smeo da bude manji od 12 (n>12) da bi se dobila reprezentativna regresiona prava i pravi oblik međuzavisnosti među pojavama.
Parametar **a– regresiona konstanta** određuje *„nivo“* regresione prave. To je vrednost Y za x = 0 i predstavlja tačku u kojoj regresiona linija seče Y-osu. Drugim rečima, to je početna vrednost zavisne Y kada još uvek nije počela da deluje nezavisna X. Osobine parametra a su:

1. ako je a =0, regresiona prava prolazi kroz koordinatni početak. To znači da ako obeležja ne mogu da imaju negativne vrednosti polaze od nultog „nivoa“
2. ako je a>0, regresiona prava seče ordinatnu osu iznad koordinatnog početka
3. ako je a<0, regresiona prava seče ordinatnu osu ispod koordinatnog početka

Parametar **b – koeficijent regresije** određuje *nagib* regresione prave. U matematičkom smislu on predstavlja tangens ugla koga regresiona prava zaklapa sa X-osom. Osobine parametra b su:

1. ako je b = 0, regresiona prava je paralelna sa X-osom. Toznači da obeležje Y ima uvek istu vrednost i da ne zavisi od obeležja X.
2. ako je b>1, regresiona prava se udaljava od X-ose i približava Y-osi
3. ako je b<1, regresiona prava je bliža X-osi a udaljava se od Y-ose

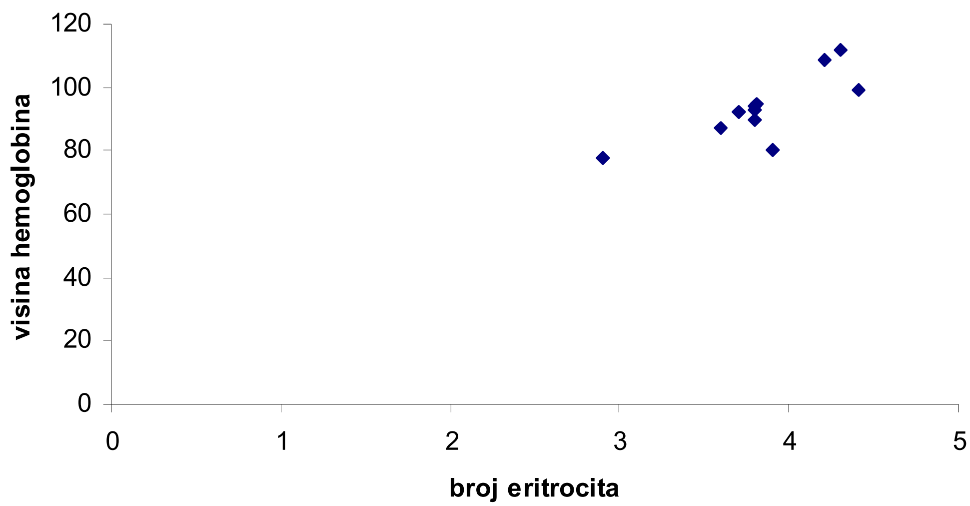
Pomoću linije regresije može se vršiti *interpolacija,* tj. određivanje vrednosti Y za bilo koju vrednost X.

Primer 1. Dat je broj eritrocita i visina hemoglogobina u krvi 12 ispitanika:

N	Broj eritrocita	Visina hemoglobina
1	4,21	108,4
2	4,3	112
3	3,6	87,3
4	4,41	99
5	3,8	93
6	3,7	92,3
7	3,8	90
8	3,8	94
9	3,81	95
10	3,7	92,3
11	2,9	7,96
12	3,9	80
Σ	44,72	1122,9

Konstruši regresionu liniju.

Kao prvi korak podatke treba ubaciti u dijagram rasturanja, da bi se ocenilo postojanje korelacije i oblik zavisnosti. Ucrtane tačke najbolje pokazuju (aproksimiraju) oblik prave linije, kao i porast u pozitivnom smeru. To znači da sa porastom broja eritrocita raste i koločina hemoglobina u krvi.



Za izračunavanje jednačine regresije najpre treba formirati radnu tabelu, koja u našem primeru izgleda:

N	x	y	x^2	y^2	xy
1	4,21	108,4	17,72	11750,56	456,36
2	4,30	112,0	18,49	12544,0	481,60
3	3,60	87,3	12,96	7621,29	314,28
4	4,41	99,0	16,81	980,0	405,90
5	3,80	93,0	14,44	8649,0	353,40
6	3,70	92,3	13,96	8519,19	341,51
7	3,80	90,0	14,44	8100,0	342,00
8	3,80	94,0	14,44	8836,0	357,20
9	3,81	95,0	14,52	9025,0	361,95
10	3,70	92,3	13,69	8519,29	341,51
11	2,90	77,96	8,41	6336,16	230,84
12	3,90	80,0	9,00	6400,0	240,00
Σ	44,72	1122,9	168,61	106101,59	4226,55

$$\bar{x} = \frac{\sum x}{n} = \frac{44,72}{12} = 3,73 \qquad \text{i} \qquad \bar{y} = \frac{\sum y}{n} = \frac{1122,9}{12} = 93,58$$

$$b = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{n \cdot \sum x^2 - \left(\sum x\right)^2} = \frac{12 \cdot 4226,5 - 44,72 \cdot 1122,9}{12 \cdot 168,61 - 44,72^2} = 22,44$$

$$a = \bar{y} - b \cdot \bar{x} = 93,58 - 21,44 \cdot 3,73 = 13,61$$

Sada imamo sve parametre za izračunavanje jednačine regresije:

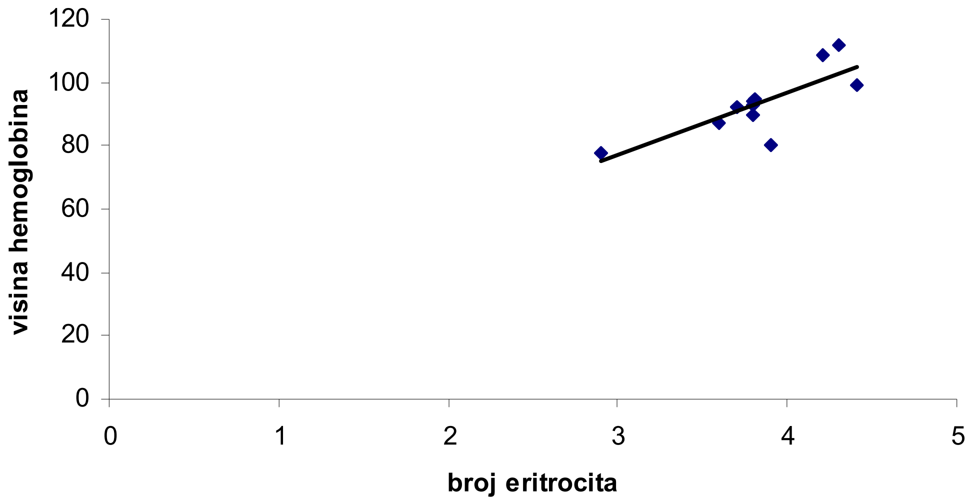
$$y_c = 13,61 + 21,44 \cdot x$$

Da bi smo konstruisali regresionu liniju, potrebno je odrediti bar dve koordinatne tačke. Uzećemo najmanju i najveću vrednost za nezavisno promenljivu (x).

Za x = 2,9 y_c = 13,61 + 21,44 x 2,9 = 75,79

Za x = 4,41 y_c = 13,61 + 21,44 x 4,41 = 108,16

gde je y_c ocena prosečne vrednosti za vrednost nezavisno promenljive.



Kao što se iz slike može videti, prava linija je blizu svih tačaka i zbir kvadrata odstupanja je manji nego za bilo koju drugu pravu liniju, tj. zbir kvadrata odstupanja je minimalan.

KORELACIONA ANALIZA

Korelaciona analiza pokazuje *stepen* zavisnosti između promenljivih, odnosno korelacijom se meri jačina već utvrđene povezanosti između dve promenljive.

Stepen intenziteta povezanosti između promenljivih, koje su u linearnom odnosu meri se:

- Kovarijansom kao apsolutnom merom intenziteta korelacije i
- Koeficijentom proste lenearne korelacije, kao relativnom merom intenziteta korelacione veze.

Kovarijansa predstavlja u suštini zajedničku meru varijabilnosti, jedne i druge varijabile, pa se matematički može da predstavi kao zbir varijansi jedne i druge varijable:

$$C_{xy} = SD_x^2 + SD_y^2, \quad \text{tj. } C_{xy} = \frac{\sum\left(x-\bar{x}\right)^2}{n} + \frac{\sum\left(y-\bar{y}\right)^2}{n}$$

Odakle se dobija radna formula za kovarijansu:

$$C_{xy} = \frac{\sum xy}{n} - \bar{x}\bar{y}$$

gde je n veličina uzorka, odnosno, broj koreliranih parova vrednosti.

Međutim, kovarijansa kao apsolutna mera stepena povezanosti nije pogodna za procenu, pa se pristupa izračunavanju relativne mere tj. izračunava se: koeficijent proste linearne korelacije.

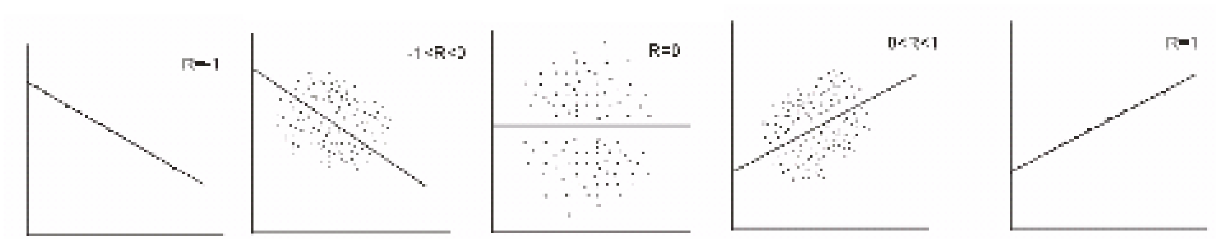
Koeficijent proste linearne korelacije ili **Pearson-ov koeficijent** predstavlja kovarijansu izraženu u jedinicama standardnih devijacija obeju varijabli.

Izračunava se kao količnik između kovarijanse i proizvoda standardnih devijacija jedne i druge varijable, pa je njegova formula:

$$r_{xy} = \frac{C_{xy}}{SD_x \cdot SD_y} = \frac{\text{kovarijansa}}{\text{proizvod standardnih devijacija x i y}}$$

Koeficijent proste linearne korelacije pokazuje stepen zavisnosti između promenljivih I on određuje veličinu disperzije (rasturanja) podataka oko regresione linije.

Ako varijable nisu povezane disperzija oko regresione linije je velika. Sa povećanjem linearne povezanosti, disperzija se smanjuje I grafik postaje sve spljošteniji. Ako između dve promenljive postoji apsolutno slaganje svi podaci leže na regresionoj liniji.



A B C D E

Grafik. A.savršena negativna korelacija; B.negativna korelacija; C. nema korelacije; D.pozitivna korelacija; E.savršena pozitivna korelacija

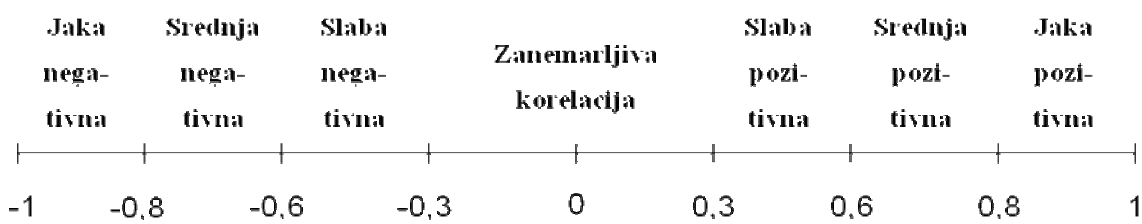
Koeficijent korelacije ima vrednost koja se kreće u rasponu od -1 do +1.

Ako varijable nisu povezane, r je jednak nuli. Kada većim vrednostima nezavisno promenljive x, odgovaraju i veće vrednosti zavisno promenljive y i obrnuto: opadanjem vrednosti nezavisne x, opadaju i vrednosti zavisne y - onda je to pozitivna korelacija ($r>0$). Obrnuto, kada većim vrednostima nezavisno promenljive x, odgovaraju manje vrednosti zavisno promenljive y , odnosno opadanjem vrednosti nezavisne x rastu vrednosti zavisne y - onda je to negativna korelacija ($r<0$).

Važi opšte pravilo: što je vrednost koeficijenta proste linearne korelacije bliža jedinici, to je međuzavisnost među posmatranim pojavama jača.

Koeficijent korelacije nikada nema vrednosti 1 ili -1, jer to bi značuilo da između pojava postoji matematička, a ne statistička veza.

Skala za tumačenje koeficijenata korelacije:



Vrednost koeficijenta korelacije može da se ocenjuje i preko specijalnih tablica za granične vrednosti r_{xy}

Tablice prikazuju kolika mora da bude najmanja vrednost r_{xy}, da bi se za određeni broj stepena slobode i za odgovarajući prag značajnosti (0,05 ili 0,01) mogao da smatra statistički signifikantnim. Broj stepeni slobode se izračunava po formuli:

S.S. = n-2   gde je: n = broj parova.

Primer 2. Utvrditi da li postoji korelacija između broja eritrocita i visine hemoglobina već datih podataka za 12 ispitanika.

Ho: Između broja eritrocita i visine hemoglobina 12 ispitanika ne postoji korelacija

Ha: Između broja eritrocita i visine hemoglobina 12 ispitanika ne postoji korelacija

Sledeći korak je izračunavanje koeficijenta proste linearne korelacije.

Formula za njegovo izračunavanje je:

$$r_{xy} = \frac{C_{xy}}{SD_x \cdot SD_y}$$

pa je potrebno izračunati prvo Cxy i SDx i SDy.

$$C_{xy} = \frac{\sum xy}{n} - \bar{x}\bar{y}$$

Kako smo već napravili radnu tabelu i izračunali $\bar{x}$ i $\bar{y}$, možemo odmah krenuti na izračunavanje kovarijanse.

$$C_{xy} = \frac{4226,56}{12} - 3,727 \cdot 93,575 = 3,46$$

$$SD_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{168,61}{12} - 3,727^2} = 0,4$$

$$SD_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{106101,59}{12} - 93,575^2} = 9,25$$

Sada imamo sve elemente za izračunavanje koficijenta, pa je:

$$r_{xy} = \frac{C_{xy}}{SD_x \cdot SD_y} = \frac{3,46}{0,4 \cdot 9,25} = \frac{3,46}{3,7} = 0,94$$

Za naš primer vrednost koeficijenta korelacije je 0,94 što znači: Između broja eritrocita i vrednosti hemoglobina postoji jaka pozitivna (direktna) korelacija.

Za naš primer S.S. = n – 2 = 12-2 = 10.

Za stepen slobode 10 i p=0,05 granična vrednost r_{xy}=0,576.

$r_{xy} = 0,94 > r_{xy(10;\ 0,05)} = 0,576$ i $p < 0,05$

Kako je dobijena r_{xy} vrednost od 0,94 veća od granične tablične vrednosti,

r_{xy} =0,576, za broj stepeni slobode 10 i prag značajnosti od p=0,05, to odbacujemo nultu i prihvatamo alternativnu hipotezu sa greškom p<0,05 i sigurnošću P>95% tvrdimo da postoji jaka pozitivna korelacija između broja eritrocita i količine hemoglobina 12 ispitanika.

Za stepen slobode 10 i p=0,01 granična vrednost r_{xy}=0,708.

$r_{xy} = 0,94 > r_{xy(10;\ 0,01)} = 0,708$ i $p < 0,01$

Kako je dobijena r_{xy} vrednost od 0,94 veća od granične tablične vrednosti,

r_{xy} =0,708, za stepen slobode 10 i prag značajnosti od p=0,01, odbacujemo nultu i prihvatamo alternativnu hipotezu sa greškom p<0,01 i sigurnošću P>99% tvrdimo da postoji jaka pozitivna korelacija između broja eritrocita i količine hemoglobina 12 ispitanika.

Postoji još jedan problem. Naime, koeficijent proste linearne korelacije se obično izračunava iz uzorka. Postavlja se pitanje njegove signifikantnosti za celu populaciju, odnosno da li uzorak iz koga je izračunat koeficijent dovoljno reprezentativan za donošenje nepristrasne ocene koeficijenta osnovnog skupa.

Dok se to ne utvrdi, dobijena vrednost koeficijenta na osnovu uzorka predstavlja samo hipotezu o vrednosti istog koeficijenta osnovnog skupa.

Problem je rešen na sledeći način:

Testira se hipoteza da li je izračunati prost koeficijent linearne korelacije iz uzorka (r_{xy}) i precizna ocena prostog koeficijenta linearne korelacije osnovnog skupa (Rxy).

Ako odgovarajućim testom odbacimo nultu hipotezu, prihvatamo izračunatu vrednost koeficijenta korelacije iz uzorka kao pravu ocenu koeficijenta u osnovnom skupu. Drugim rečima uzorak je reprezentativan, pa dobijeni rezultat može da se uopšti.

Testiranje koeficijenta proste linearne korelacije se zasniva na Studentovom rasporedu za n-2 stepena slobode, a dobijena t-vrednost se tumači na isti način kao i kod klasičnog Studentovog t-testa.

Test je matematički definisan formulom:

$$t = \frac{r_{xy}}{\sqrt{\frac{1-r_{xy}^2}{n-2}}} \quad \text{ili} \qquad t = r_{xy}\frac{\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

gde je: r_{xy} - dobijena vrednost iz uzorka,

n - velicina uzorka (broj parova).

Bronj stepena slobode se izračunava po obrascu: S.S. = n-2.

Dobijena t vrednost se tumači na isti način kao i kod klasičnog Studentovog t testa.

Primer: Testirajmo dobijenu vrednost, $r_{xy} = 0{,}94$ za 12 osoba kod kojih je tražena veza izmedu broja eritrocita i vrednosti hemoglobina.

Ho: Rxy (osnovnog skupa) = 0

Ha: Rxy (osnovnog skupa) $\neq$ 0

$$t = \frac{0{,}94}{\sqrt{\frac{1-0{,}94^2}{12-2}}} = \frac{0{,}94}{0{,}1078} = 8{,}712$$

$t = 8{,}712 > t_{(10 \text{ i } 0{,}05)} = 2{,}23$ i $p < 0{,}05$

Kako je dobijena t vrednost od 8,712 veća od granične tablične vrednosti,

t=2,23, za stepenislobode 10 i prag značajnosti od p=0,05, to odbacujemo nultu i prihvatamo alternativnu hipotezu sa greškom p<0,05 i sigurnošću P>95% zaključujemo: između broja eritrocita i vrednosti hemoglobina postoji visok stepen korelacije, a dobijena vrednost $r_{xy} = 0,94$ predstavlja stvarnu meru korelacije pa se zaključak može da uopšti na celu populaciju.

$t = 8,712 > t_{(10 \text{ i } 0,01)} = 3,17$ i $p < 0,01$

Kako je dobijena t vrednost od 8,712 veća od granične tablične vrednosti,

t=3,17, za broj stepeni slobode 10 i prag značajnosti od p=0,01, odbacujemo nultu i prihvatamo alternativnu hipotezu sa greškom p<0,05 i sigurnošću P većim i od 99%.

Pirsonov koeficijent korelacije daje informacije da li je povezanost varijabli slaba, umerena, jaka ili veoma jaka. Medutim, on nam ne daje i informaciju koliko je zavisna promenljiva uslovljena vrednostima nezavisno promenljive, a koliko drugim faktorima.

Ovaj problem rešava koeficijent determinacije, koji se najlakše izračunava kao drugi stepen koeficijenta proste linearne korelacije I on je mera za *objašnjeni* varijabilitet:

koeficijent determinacije = $\mathbf{r_{xy}^2} = 0,94^2 = 0,8836$

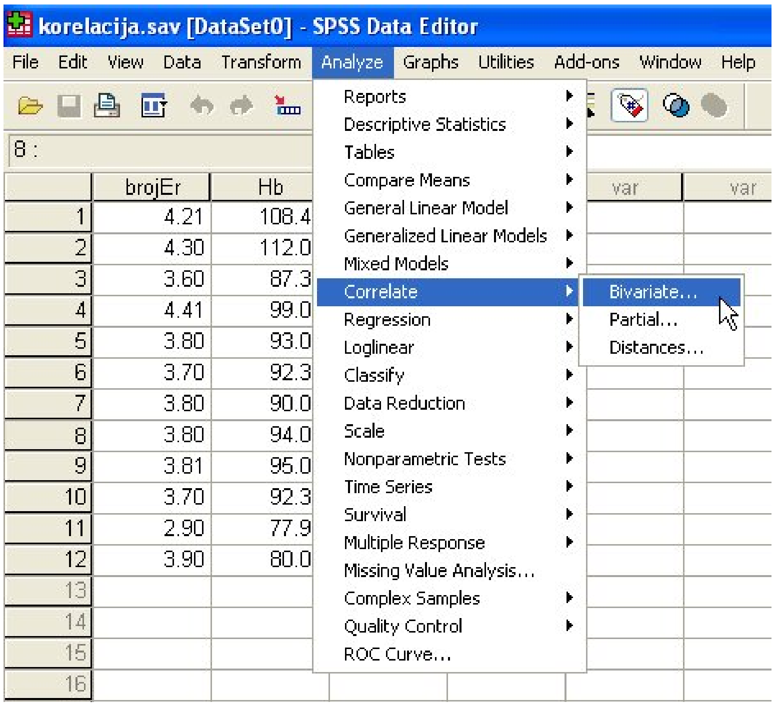
Vrednost koeficijenta determinacije od 0,8836 nam pokazuje da su vrednosti hemoglobina sa 88,36% određene (determinisane) brojem eritrocita.

Ostatak od 1 je **koeficijent alijenacije**: $\mathbf{1 - r_{xy}^2} = 1 - 0,8836 = 0,1164$ tj. od 11,64% uslovljen je drugim faktorima i on predstavlja, koji je mera za *neobjašnjeni* varijabilitet.

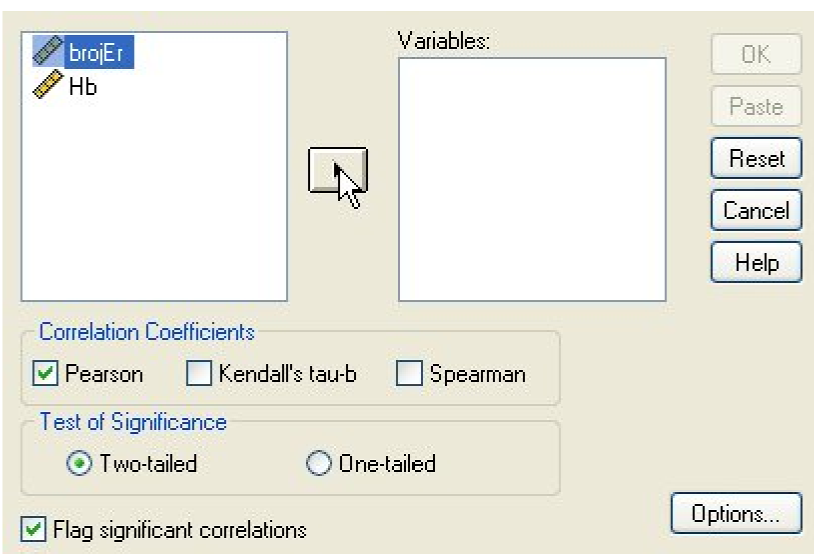
Zbir objašnjenog (determinisanog) varijabiliteta i neobjašnjenog varijabiliteta je uvek jednak jedinici, odnosno 100%.

U SPSS-u se koeficijent proste linearne korelacije se određuje na sledeći način:

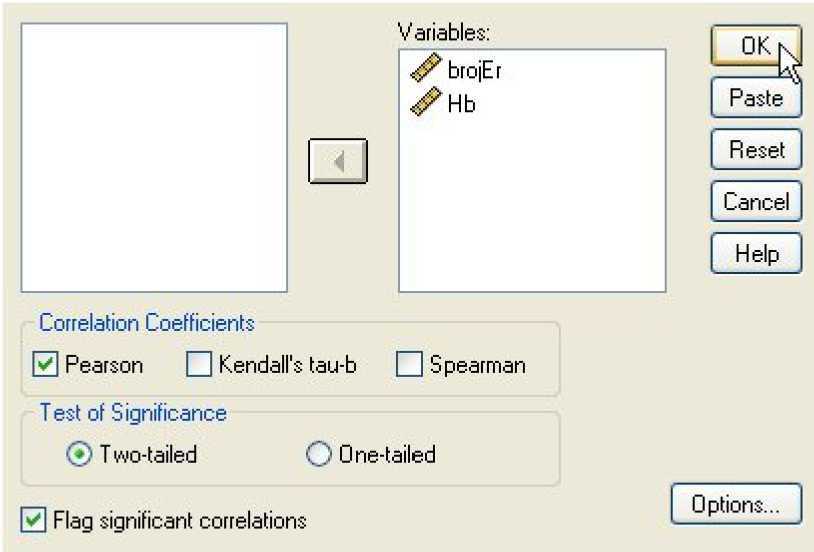
Obeleži se *Analyse / Correlate* i u desnom grananju *Bivariate*:



Na ekranu se dobije:



Opcije koje je ponudio računar *Pearson*, *Two-tailed* i *Flag significant correlations* se zadrže. Željene varijable se prebace iz levog u desni prozor, u našem slučaju varijabla sa brojem eritrocita (“brojEr”*)* i visinom hemoglobina (“Hb”).



Klikne se na *OK* I dobiju rezultati:

Correlations

		brojEr	Hb
brojEr	Pearson Correlation	1	.778**
	Sig. (2-tailed)		.003
	N	12	12
Hb	Pearson Correlation	.778**	1
	Sig. (2-tailed)	.003	
	N	12	12

**. Correlation is significant at the 0.01 level

U tabeli su date vrednost Pirsonovog koeficijenta korelacije koji je za dati primer 0,778, a vrednosti p se čita u *Sig. (2-tailed)* i iznosi 0,003 u datom zadatku.

SPEARMAN-OV KOEFICIJENT RANG KORELACIJE

Spearmanov koeficijent rang korelacije je neparametrijski ekvivalent Pearsonovom koeficijentu linearne korelacije.

Razlika je u tome što se računske operacije ne izvode iz numeričkih vrednosti zavisne i nezavisno promenljive pojave, već iz njihovih relativnih odnosa tj. rangova.

Računamo ga ako je ispunjen jedan ili više sledećih uslova:

- Barem jedna od varijabli, x ili y, merena je ordinalnom skalom
- Ni x ni y nemaju normalnu distribuciju
- Uzorak je mali
- Treba nam mera povezanosti između dve varijable kada ta povezanost nije linearna

Postupak se odvija u dve etape:

1. Stvarne numeričke vrednosti i zavisne i nezavisne pojave sređuju se po veličini, od najmanje do najveće (ili obrnuto) i određuje se njihov rang tj. obeleže se kao: prvi rang (prvo mesto = 1), drugi rang (drugo mesto = 2), treći rang (treće mesto po velicini = 3) i tako do n-tog ranga.

Kod rangova ne znamo stvarne razlike između numeričkih vrednosti, nego jedino razlike između rangova.

2. Kada smo podatke pravih vrednosti transformisali u rangove pristupa se izračunavanju tzv. rang korelacije, tj. izračunava se korelacija među rangovima.

Ovaj postupak je kraći i praktičniji od Pearson-ove linearne korelacije, pogotovu ako broj parova nije veliki.

Spearman-ov koeficijent rang korelacije se izračunava po formuli:

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n \cdot \left(n^2 - 1\right)},$$

gde je:

ρ - (ro) Spearmanov koeficijent,

d – razlika (diferencija) između rangova x i y,

n - broj parova rangova promenljivih x i y.

Stepen slobode se izračunava: S.S. = n.

I koeficijent rang korelacije, može da ima vrednosti od -1 do + 1.

Što je razlika između rangova obeležja x i y manja, to se njegova vrednost više približava vrednostima + 1 i -1, a to znači i da je stepen korelacije veći između posmatranih pojava.

Razlika između Pirsonovog koeficijenta proste linearne korelacije i Spirmanovog koeficijenta rang korelacije, je u tome što se ovaj poslednji može da izračunava iz podataka, kada je merenje vršeno na ordinalnoj skali.

Spirmanov koeficijent može da zameni Pirsonov, ako se intervalni podaci prevedu u ordinalne tj. ako se rangiraju po velicini. Obrnuto, ako su podaci dati u ordinarnoj skali, može da se primeni samo Spirmanov koeficijent.

Bitna razlika je i u sledećem:

Statistička snaga "power" Pirsonovog koeficijenta je znatno veća nego Spirmanovog, pa zato ako su podaci dati intervalno, prednost treba dati Pirsonovom koeficijentu, a Spirmanov zbog lakoće izračunavanja primeniti kao pilot probu.

Primer 3. Asistent je rangirao 7 studenata iz svoje grupe u odnosu prema nastavi (x) i stepenu obučenosti za praktičan rad (y).
Kao rang 1 je koristio najpovoljniju, a kao rang 7 najnepovoljniju ocenu za oba modaliteta.

Student	A	B	C	D	E	F	G
Rang za x	4	2	6	1	3	7	5
Rang za y	3	1	6	2	4	7	5

Da li je odnos prema nastavi u međuzavisnosti sa obučenošću za praktičan rad ?

Ho: Između odnosa prema nastavi i obučenosti za praktičan rad ne postoji međuzavisnost.

Ha: Između odnosa prema nastavi i obučenosti za praktičan rad postoji međuzavisnost.

Konstruišemo radnu tabelu i sređujemo rangove po veličini:

Student	Rang za x	Rang za y	d	D^2
A	4	3	1	1
B	2	1	1	1
C	6	6	0	0
D	1	2	-1	1
E	3	4	-1	1
F	7	7	0	0
G	5	5	0	0
Σ	-	-	-	4

Spirmanov koeficijent rang korelacije je:

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 4}{7 \cdot (7^2 - 1)} = 0{,}93$$

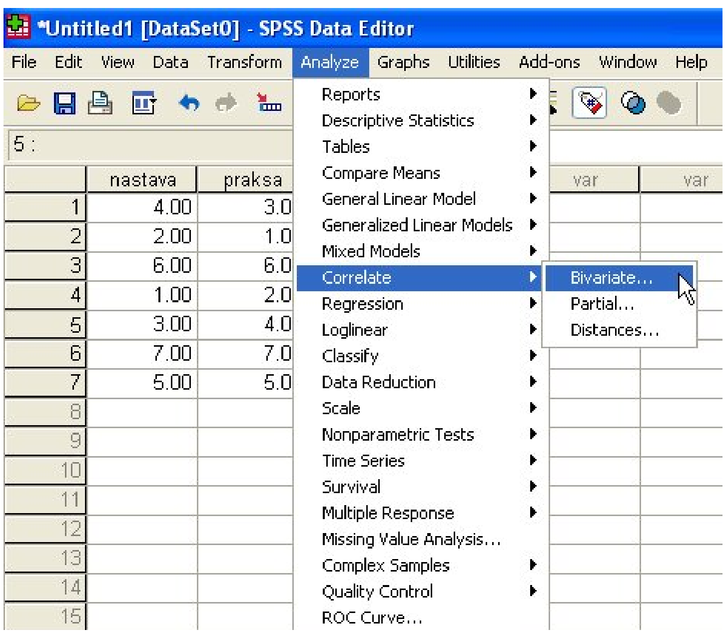
S.S. = 7

$\rho = 0{,}93 > \rho_{(7 \text{ i } 0,05)} = 0{,}786$ i $p < 0{,}05$

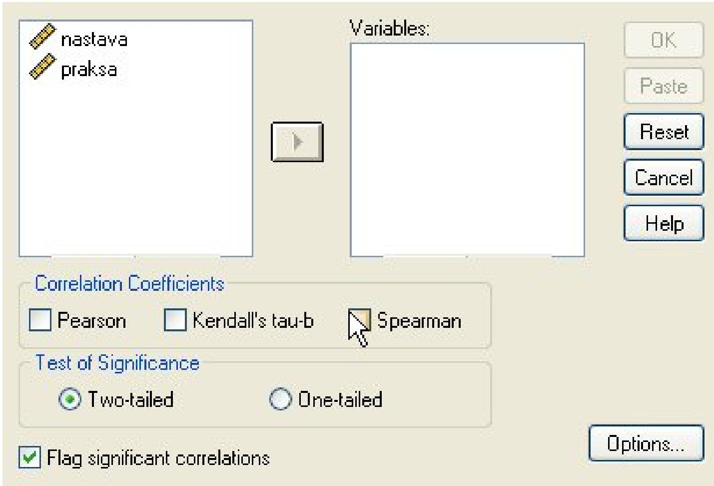
Dakle, između odnosa prema nastavi i obučenosti za praktičan rad postoji jaka međuzavisnost, što tvrdimo sa $p < 0{,}05$.

U SPSS-u se koeficijent proste linearne korelacije se određuje na sledeći način:

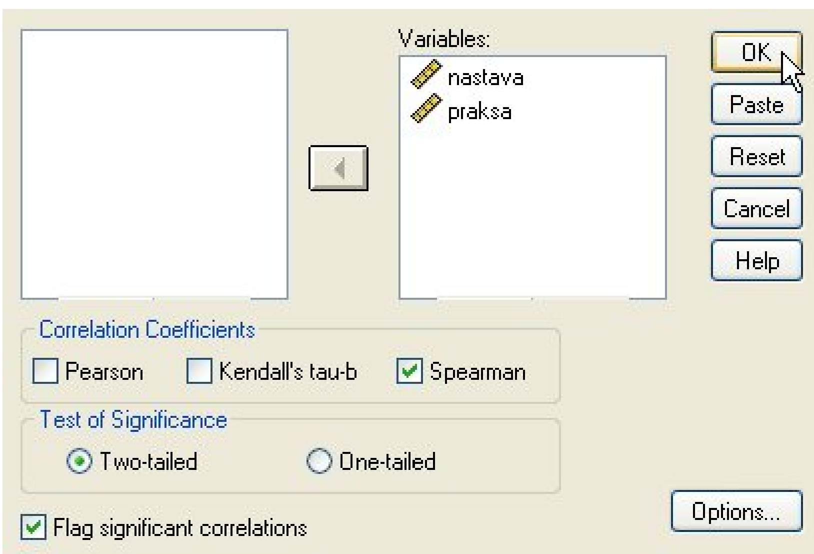
Obeleži se *Analyse / Correlate* I u desnom grananju *Bivariate*:



Na ekranu se dobije:



Opcije koje ponudi računar *Two-tailed* i *Flag significant correlations* se zadrže. Obeleži se opcija za željeni test, tj. *Spearman*. Ispitivane varijable se prebace iz levog u desni prozor, u našem slučaju varijabla sa odnosom prema nastavi ("nastava") i stepenom obučenosti za praktični rad ("praksa").



Klikne se na *OK* i dobiju rezultati:

Correlations

			nastava	praksa
Spearman's rho	nastava	Correlation Coefficient	1.000	.929**
		Sig. (2-tailed)	.	.003
		N	7	7
	praksa	Correlation Coefficient	.929**	1.000
		Sig. (2-tailed)	.003	.
		N	7	7

**. Correlation is significant at the 0.01 level (2-tailed).

U tabeli su date vrednosti Spirmanovog koeficijenta ρ (0,929) i vrednost p (0,003) - *Sig. (2-tailed).*

Zadaci za vežbanje

1. Izmerena je koncentracija kiseonika kod 10 zdravih odraslih muškaraca i dobijene su sledeće vrednosti:

N	1	2	3	4	5	6	7	8	9	10
arterijska	23	17	19	20	22	21	21	19,5	20,5	24
venska	16	11,5	13	14	12,5	14,5	14,2	11,5	17	18

Da li postoji povezanost između koncentracije kiseonika u arterijskoj i venskoj krvi.

Dobijene podatke rangiraj i izračunaj Spirmanov koeficijent rang korelacije.

2. U grupi 11 obolelih od skleritisa udruženog sa sistemskim bolestima vezivnog tkiva registrovan je broj recidiva i nivo cirkulišućih imunih kompleksa (CIC). Može li se na osnovu nivoa CIC prognozirati učestralčst recidiva?

Br.recidiva	6	10	15	3	4	2	6	2	4	15	3
CIC (mg%)	171	149	265	168	55	7	37	55	146	144	82

3. Ugrupi od 10 pacijenata sa glaukomom izmerena je aktivnost enzima superoksiddismutaze (SOD) u očnoj vodici, prilikom operacije, a takođe i koeficijent lakoće isticanja očne vodice (C). Da li je aktivnost enzima povezana sa promenom koeficijenta lakoće isticanja očne vodice (C)?

SOD	1,5	1,6	1,7	1,9	1,1	1,3	1,4	1,3	1,4	1,7
C	0,08	0,06	0,09	0,06	0,14	0,12	0,10	0,16	0,12	0,06

4. Kod trinaestoro dece izvršeno je tuberkulinsko testiranje, a rezultate su čitala dva lekara sa sledećim nalazima:

r.br. deteta	1	2	3	4	5	6	7	8	9	10	11	12	13
lekar A	6	4	8	12	14	9	15	5	5	3	7	11	13
lekar B	5	4	7	10	15	10	16	7	5	3	7	9	14

Da li postoji slaganje između lekara A i B u očitavanju veličine tuberkulinske reakcije?

5. Kod 40 osoba meren je sistolni pritisak i dobijeni su sledeći rezultati u odnosu na godine starosti:

- starost: $\overline{x} = 49{,}2$, $SD_x = 8{,}2$, $\sum x^2 = 1968$, $\sum xy = 1968$
- sistolni pritisak: $\overline{y} = 140$, $SD_y = 14$, $\sum y^2 = 5600$

Utvrdi da li postoji veza između godina starosti i sistolnog pritiska?

Prilog. Dvostrane 5% i 1% tačke raspodele koeficijenta korelacije

n	5%	1%	*n*	5%	1%	*n*	5%	1%
5	0.88	0.96	18	0.47	0.59	40	0.31	0.40
6	0.81	0.92	19	0.46	0.58	50	0.28	0.36
7	0.75	0.87	20	0.44	0.56	60	0.25	0.33
8	0.71	0.83	21	0.43	0.55	70	0.24	0.31
9	0.67	0.80	22	0.42	0.54	80	0.22	0.29
10	0.63	0.77	23	0.41	0.53	90	0.21	0.27
11	0.60	0.74	24	0.40	0.52	100	0.20	0.25
12	0.58	0.71	25	0.40	0.51	200	0.14	0.18
13	0.55	0.68	26	0.39	0.50	500	0.09	0.12
14	0.53	0.66	27	0.38	0.49	1000	0.06	0.08
15	0.51	0.64	28	0.37	0.48			

n = broj posmatranja