Mr VLADIMIR HITREC
Šumarski fakultet Zagreb

UDC 519.2:634.0.83/86

O nekim koeficijentima koji određuju vezu između dvije veličine

— PRIMJENA U DRVNOJ INDUSTRIJI —

SAŽETAK

Rad je kritički osvrt na dosta često upotrebljavanu metodu prognoze pomoću koeficijenata. Ukazano je da se prognoze rade metodama regresione analize, dakle metodama matematičke statistike.

Pokušaj da se nađe koeficijent pretvorbe nekog obilježja X u obilježje Y, tj. da se na neki način izračuna konstanta C, kojom možemo množiti X i na taj način procijeniti Y, može biti često krivo interpretiran, a često i besmislen.

Konstantu C možemo upotrijebiti u slijedećim slučajevima:

1. Obilježja X i Y moraju biti proporcionalna, što znači da je njihova stohastička veza pravac koji prolazi kroz ishodište ($Y = CX$). U tom se slučaju konstanta C i greške procjene mogu izračunati poznatim metodama regresione analize.

2. Obilježje X jednako je nekoj konstanti ($X = K$). U tom slučaju vrijednost CK je jedna točkovna procjena obilježja Y.

3. Na elementima slučajnog uzorka izmjerena je vrijednost obilježja X i odgovarajuća vrijednost obilježja Y, te je izračunata konstanta $C = \Sigma Y / \Sigma X$. Sada će, uz određenu pouzdanost, vrijediti $\Sigma Y = C \Sigma X$ na svim slučajnim uzorcima uzetim iz istog osnovnog skupa.

Na jednom uzorku različitih trupaca izračunati koeficijent volumnog iskorišćenja možemo primijeniti na drugi skup trupaca samo ako je ovaj potonji iste ili slične strukture kao i prvi, tj. ako se sastoji od jednakog broja trupaca istog volumena.

KLJUČNE RIJEČI: prognoza — regresiona analiza — koeficijent iskorišćenja trupaca — koeficijent protoka materijala

ABOUT SOME COEFFICIENTS DETERMINING THE RELATION BETWEEN TWO PROPERTIES — APPLICATION IN WOODWORKING INDUSTRY

SUMMARY

The paper is a critical review of the rather frequently used method of prognosis by means of coefficent. It is pointed out that prognoses are performed by the methods of regression analysis, hence by the methods of mathematical statistics.

The attempt to find the coefficient of conversion of some random variable Y, viz. to compute in a way constant C, with which latter we can multyply X and thus estimate Y, may frequently be wrongly interpreted or be without meaning, even.

We can use constant C in the following cases:

1. Random variables X and Y ought to be proportional, which means that their stochastic connection is a straight line passing through the origin of coordinates ($y = CX$). In this case constant C and errors of estimate can be computed by the well-known methods of regression analysis.

2. Random variable X is equal to a certain constant ($X = K$). In this case the value of CK is a point estimate of the random variable Y.

3. On the elements of a random sample were mesaured the value of the random variable X and the corresponding value of the random variable Y, and thus the constant $C = \Sigma Y / \Sigma X$ was computed. Under a determinate confidence, valid will be $\Sigma Y = C \Sigma X$ on all the random samples taken from the same basic set.

The coefficient of logs volume yield computed on a sample of different logs may be applied to another set of logs only if the latter is of the same or similar structure as the first, i. e. if it consists of an identical number of logs of the same volume.

KEY WORDS: prognosis — regression analysis — log volume yield — flow of material

1. PROBLEM

Kada se u praksi govori o koeficijentu, najčešće se misli na stanovitu konstantu k koja povezuje dvije veličine. Označimo li te veličine sa X odnosno Y, tada je koeficijent koji im pripada

$$k = \frac{Y}{X} \quad (1)$$

Govori se npr. o koeficijentu iskorišćenja kapaciteta, koeficijentu iskorišćenja trupaca, koeficijentu protoka materijala itd.

Jedan od zadataka ovog rada jest da pokaže gdje smijemo koeficijente takve vrste računati, odnosno, gdje oni imaju smisla, a gdje računanje koeficijenta (1) treba zamijeniti drugim metodama.

Budući da se kod računanja koeficijenata (1), a i kod drugih metoda o kojima će kasnije biti riječ, radi o procjenama dobivenim uzorcima, nužno će biti odrediti greške takvih procjena. To je drugi zadatak ovog rada.

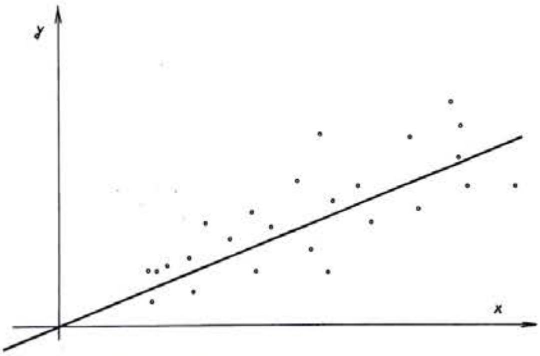
2. MATEMATIČKE OSNOVE

Na elementima osnovnog statističkog skupa definirana su dva obilježja: obilježje X i obilježje Y, za koje smatramo da su u stanovitoj vezi. (Ako su npr. elementi osnovnog skupa trupci, onda Y može biti volumen građe, a X volumen trupaca).

Stavimo li da je kao u (1), k = Y/X onda smo time pretpostavili da je veza između X i Y LINEARNA, i čak specijalnije da je Y u p r a v o p r o p o r c i o n a l a n s X.

Ta je pretpostavka ekvivalentna s odnosom veličina X i Y kako je to prikazano na slici 1, odnosno s pretpostavkom da se stohastička veza veličina Y i X može prikazati pravcem regresije k o j i p r o l a z i k r o z i s h o d i š t e, te ima jednadžbu

$$Y = k \cdot X.$$



Slika 1. — Stohastički odnos veličina Y i X koji mora postojati, želimo li računati prema jednadžbi (1).

* Materija razmatrana u ovom poglavlju može se naći u literaturi navedenoj pod brojem (1)

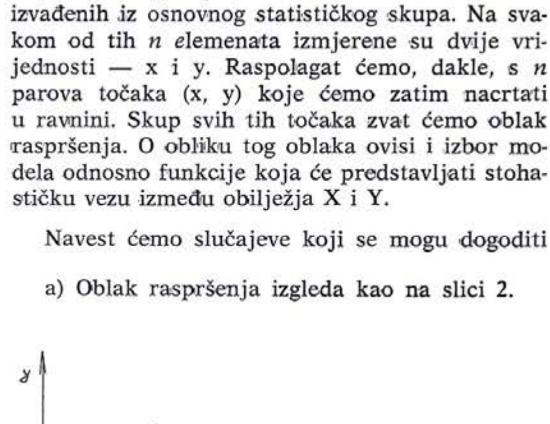
Evidentno je da ta prepostavka ne mora u svim slučajevima u praksi biti zadovoljena. Ako pretpostavka proporcionalnosti nije zadovoljena, to još naravno ne znači da ne postoji veza između obilježja Y i obilježja X. Ta veza može postojati, a njezino određivanje i analiza spadaju u dio statistike koji se zove regresiona analiza. U regresionu analizu spada i određivanje koeficijenta k iz jednadžbe (1), dakle i slučaj proporcionalnosti. Ako postoji, k je, dakle, koeficijent regresije (ne proporcija!) te se njegova greška računa također poznatim metodama regresione analize.

U metode regresione analize nećemo ovdje ulaziti, no smatramo da je potrebno izložiti način odabiranja adekvatnog modela koji će reprezentirati promatranu stohastičku vezu, te omogućiti da se za poznatu vrijednost obilježja X nađe pripadna vrijednost obilježja Y. Spomenut ćemo i metode koje omogućuju da se za tu procjenu odredi greška.

Do generalnih zakonitosti najčešće dolazimo uzorkom čije elemente mjerimo, analiziramo, te dobivene rezultate generaliziramo. U problemu u kojem je ovdje riječ, uzorak će biti n elemenata izvađenih iz osnovnog statističkog skupa. Na svakom od tih n elemenata izmjerene su dvije vrijednosti — x i y. Raspolagat ćemo, dakle, s n parova točaka (x, y) koje ćemo zatim nacrtati u ravnini. Skup svih tih točaka zvat ćemo oblak raspršenja. O obliku tog oblaka ovisi i izbor modela odnosno funkcije koja će predstavljati stohastičku vezu između obilježja X i Y.

Navest ćemo slučajeve koji se mogu dogoditi

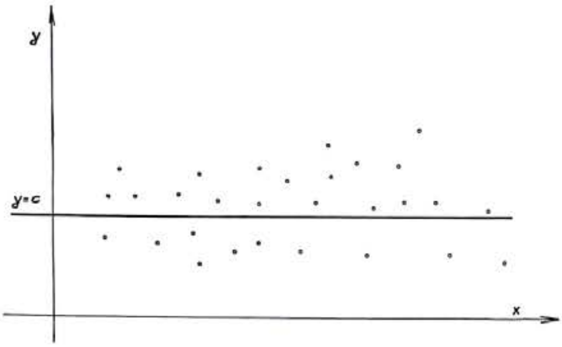
a) Oblak raspršenja izgleda kao na slici 2.



Slika 2. — Ne postoji nikakva veza između obilježja X i obilježja Y.

To je slučaj kada ne postoji nikakva veza između obilježja X i obilježja Y, te nismo u mogućnosti dati nikakvu informaciju o vrijednosti obilježja Y ako poznamo vrijednost obilježja X.

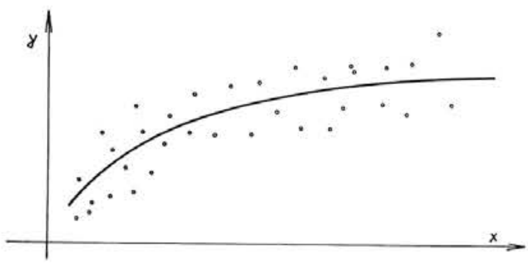
b) Oblak raspršenja izgleda kao na slici 3.



Slika 3. — Promjena vrijednosti obilježja X ne utječe na promjenu vrijednosti obilježja Y.

Vrijednosti obilježja Y su distribuirane oko konstantne vrijednosti C, te ne »reagiraju« na promjenu vrijednosti X. Za bilo koju vrijednost obilježja X, Y je uvijek stohastički konstantan, te ga možemo analizirati poznatim metodama potpuno neovisno o obilježju X.

c) Oblak raspršenja izgleda kao na slici 4.

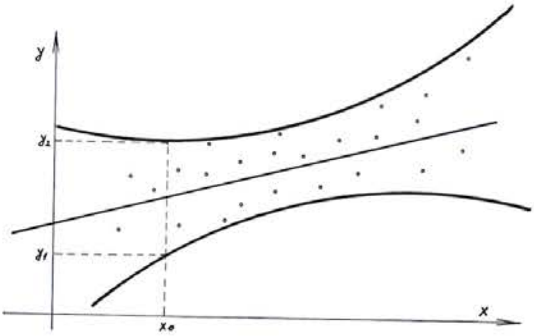


Slika 4. — Postoji nelinearna veza između obilježja X i obilježja Y.

U tom slučaju ima smisla govoriti o stohastičkoj vezi između obilježja X i Y. No, ne samo da je iz slike vidljivo da Y nije proporcionalan sa X, već je očito da ta veza nije linearna. Prema obliku oblaka raspršenja, moguće je odrediti adekvatnu funkciju koja interpretira vezu obilježja X i Y, a čije parametre možemo odrediti metodom najmanjih kvadrata. Primjer prikazan na slici 4. upućuje na traženje veze oblika k = Y/X.

d) Oblak raspršenja izgleda kao na slici 5.

Slika upućuje na postojanje linearne veze između obilježja X i Y. Oblik te veze je pravac Y = a + bX, čiji parametri se mogu odrediti metodom najmanjih kvadrata. Zakrivljene linije na slici 5. su granice konfidencije koje se također mogu izračunati poznatim metodama. Jednadžba pravca regresije zajedno s granicama konfidencije omogućuje da za zadani x_0 pročitamo



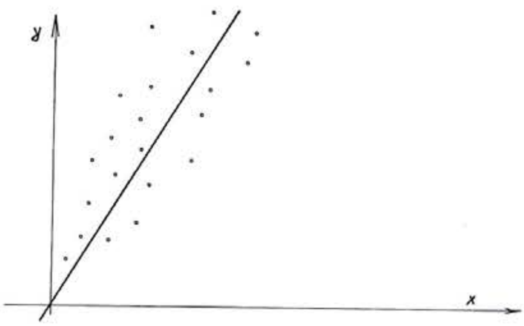
Slika 5. — Linearna regresija između obilježja X i Y, s granicama konfidencije.

(ili izračunamo) pripadnu vrijednost Y, s time da smo npr. 95% sigurni (ako su granice konfidencije računate za 95%-tnu sigurnost) da će se pripadni Y kretati u intervalu (y_1, y_2).

Ovdje ima smisla govoriti o koeficijentu Y/X, no on nije konstantan, već ovisi o vrijednosti X. Podijelimo li, naime, jednadžbu pravca regresije s X, dobit ćemo: k(x) = Y/X = a/X + b.

O vrijednostima parametara a i b ovisi da li će se koeficijent k(x) povećavati ili smanjivati s porastom vrijednosti X.

e) Oblak raspršenja izgleda kao na slici 6.



Slika 6. — Obilježje Y je proporcionalno s X.

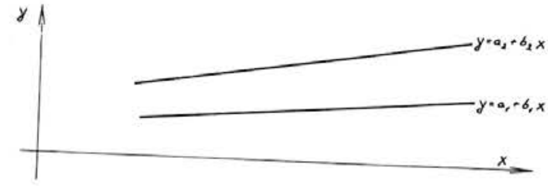
Veza prikazana na slici 6. je specijalan slučaj linearne veze. Ovdje se radi o proporcionalnosti, te možemo naći takav k da je pravac: y = k · X, jednadžba regresije za promatranu pojavu.

Kao procjenu parametra *k* možemo uzeti jedan od izraza

$$\Sigma XY/\Sigma X^2, \qquad \Sigma Y/\Sigma X, \qquad (\Sigma Y/X)/n.$$

Sva su tri gornja izraza nepristrane procjene parametra *k*. Koja će od tih procjena biti preciznija, ovisi o varijanci podataka. U većini slučajeva najbolja će biti druga od njih, dakle: $k = \Sigma Y / \Sigma X$.

Na sličan način kao i u slučaju pod d) možemo izračunati granice povjerenja procijenjenih vrijednosti Y.



Slika 7. — Pravci čiju različitost (nagiba i nivoa) možemo testirati.

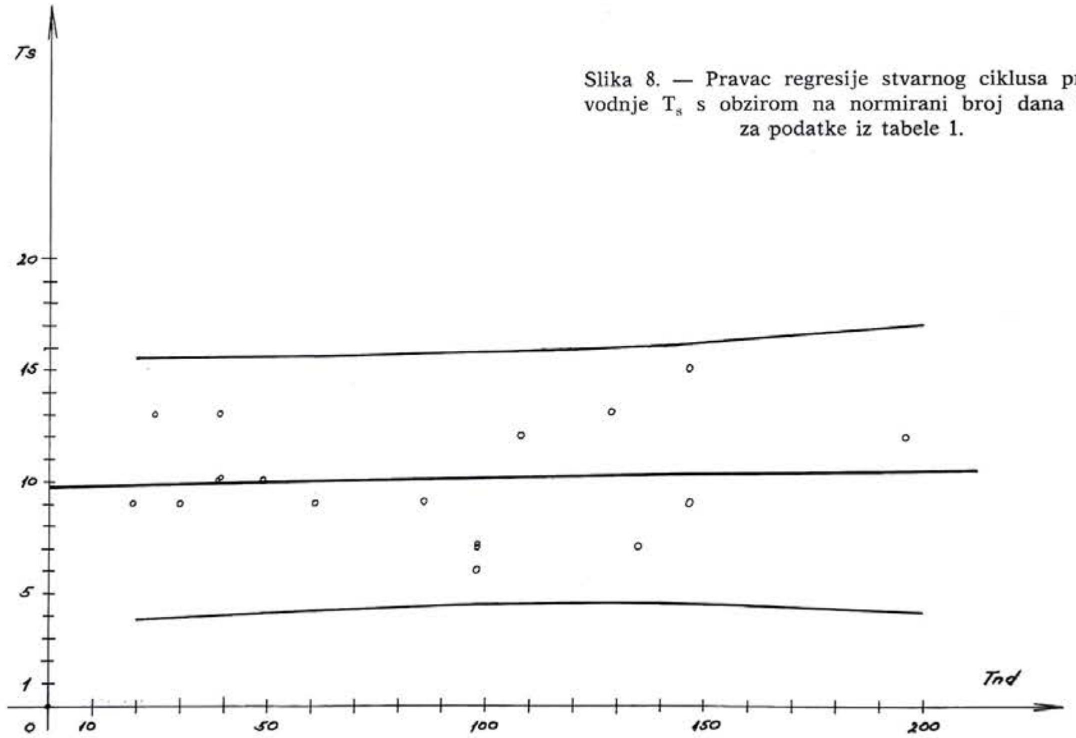
Budući da se radi o slučajnom uzorku i stohastičkoj vezi, postavlja se pitanje kada možemo kazati da je oblak raspršenja takav da pravac regresije prolazi kroz ishodište. Pomoću slike možemo procijeniti da pravac »prolazi negdje blizu ishodišta, ili možda kroz samo ishodište«, ali da li bi i drugi uzorak iz istog osnovnog skupa dao isti rezultat? Koliko uzorkom dobiveni pravac smije odstupati od ishodišta pa da to odstupanje bude još uvijek slučajno, dakle takvo da možemo smatrati da je veza koju promatramo proporcionalnost. Često je za tu odluku dovoljna slika i poznavanje prirode pojave. Ukoliko slika izaziva nedoumicu, potrebno je izvršiti testiranje hipoteze da pravac prolazi kroz ishodište.

Spomenimo još da su u statistici razrađene metode kojima uspoređujemo dva ili više pravaca regresije. Pomoću tih metoda možemo zaključiti da li se pravci dobiveni iz dva uzorka razlikuju slučajno ili bitno, i to bilo u nivou bilo u nagibu (slika 7.).

3. PRIMJERI

1. Navest ćemo prvo primjer iz organizacije rada u vezi s računanjem prosječnog koeficijenta protoka.

U jednom poduzeću drvne industrije izmjerene su za 18 radnih naloga veličine T_s (stvarni ciklus proizvodnje računat u danima), a poznat je bio i normirani broj dana T_{nd} za pripadni radni nalog (Broj potrebnih norma-sati za stanoviti nalog podijeljen brojem radnih sati u jednom danu). Dobiveni podaci su unijeti u tabelu 1, a pripadni oblak raspršenja, pravac regresije i 95%-tne granice konfidencije prikazani su na slici 8.



Slika 8. — Pravac regresije stvarnog ciklusa proizvodnje T_s s obzirom na normirani broj dana T_{nd}, za podatke iz tabele 1.

Tabela 1. — Broj normiranih dana s pripadnim stvarnim ciklusom proizvodnje i pripadnim koeficijentom protoka.

T_{nd} (1)	T_s (2)	f (3)	T_{nd} (1)	T_s (2)	f (3)
147	15	0,10	98	6	0,06
39	13	0,33	39	10	0,26
196	12	0,06	98	7	0,07
19	9	0,47	39	10	0,26
129	13	0,10	61	9	0,15
135	7	0,05	98	7	0,07
147	9	0,06	49	10	0,20
30	9	0,30	86	9	0,10
108	12	0,11	24	13	0,54

Pravac regresije dobiven je metodom najmanjih kvadrata, a jednadžba mu je

$$T_s = 9,6 + 0,005\ T_{nd}.$$

Iz slike 8, iz jednadžbe pravca (koeficijent uz T_{nd} jest vrlo malen) i iz izračunatog koeficijenta korelacije (r = 0,08) vidljivo je da nalozi koji su promatrani ne pokazuju nikakvu uzročnu vezu između normiranih dana i stvarnog ciklusa proizvodnje, što znači da nismo u mogućnosti — u ovom slučaju — poznavajući broj normiranih dana, dati bilo kakvu prognozu o stvarnom ciklusu proizvodnje. Zašto je to tako, stvar je analize koju će izvršiti tehnolog.

Primjer će nam poslužiti da pokažemo do kakvih bismo zaključaka došli da smo se umjesto regresijom služili koeficijentima. U tabeli 1, kolona (3), dani su koeficijenti protoka za svaki radni nalog. Srednja vrijednost tih koeficijenata jest

$$\bar{f} = \Sigma f_i/18 = 0,18.$$

Stavimo li

$$T_s = 0,18\ T_{nd},$$

dobili smo potpuno neadekvatan model. Uvrstimo li u taj model vrijednosti $T_{nd} = 100$, dobit ćemo $T_s = 18$, što je vrlo loša procjena. Slično bi bilo i s drugim vrijedostima T_{nd}.

Primjer 2. Odabran je slučajni uzorak od 40 jelovih trupaca III klase. Svakom trupcu određen je volumen (M) i volumen dobijene građe (G). Podaci su unijeti u tabelu 2.

Na slici 9 nanijet je pripadni oblak raspršenja koji svojim prostiranjem ukazuje na mogućnost linearne regresije. Podaci su izjednačeni pravcem, te je dobivena jednadžba: G= —0,0612+ + 0874 M. Izračunat je i koeficijent korelacije koji iznosi r = 0,96, što pokazuje veliki stupanj linearne zavisnosti volumena trupaca i dobivenog volumena građe.

Tabela 2. — Volumeni trupaca i volumeni piljene građe

M [m³]	G [m³]	M [m³]	G [m³]	M [m³]	G [m³]	M [m³]	G [m³]
0,67	0,51	0,72	0,55	0,69	0,55	0,38	0,28
0,75	0,60	0,64	0,69	0,64	0,55	0,39	0,29
0,36	0,25	0,45	0,32	0,61	0,52	0,72	0,58
0,69	0,55	0,43	0,40	0,69	0,56	0,67	0,50
0,49	0,35	0,72	0,61	0,72	0,50	0,72	0,59
0,69	0,56	0,45	0,34	0,67	0,52	0,61	0,49
0,41	0,29	0,48	0,35	0,45	0,33	0,36	0,26
0,39	0,26	0,50	0,37	0,79	0,59	0,58	0,45
0,53	0,40	0,50	0,36	0,50	0,33	0,55	0,39
0,55	0,34	0,55	0,47	0,50	0,34	0,58	0,48

Izračunate su i na slici 9 nacrtane i granice konfidencije za 95%-tnu sigurnost. Granice konfidencije su nacrtane za pojedinačne vrijednosti, što znači da za određeni volumen trupca možemo očitati interval u kome će se s 95% sigurnosti nalaziti volumen dobivene građe. Analogne granice mogu se konstruirati i za srednje vrijednosti. One će biti znatno uže i ovisit će o broju trupaca koji je uzet za uzorak.

Mala vrijednost koeficijenta a = —0,0612 navodi na pomisao da se radi o regresiji čiji pravac prolazi kroz ishodište, tj. o proporciji. Izvršili smo testiranje hipoteze

$$H_0\ ;\ a = O$$

prema alternativnoj

$$H_1\ ;\ a < 0.$$

Izračunata je vrijednost varijable *u*, te je dobiveno

$$u = 2,4,$$

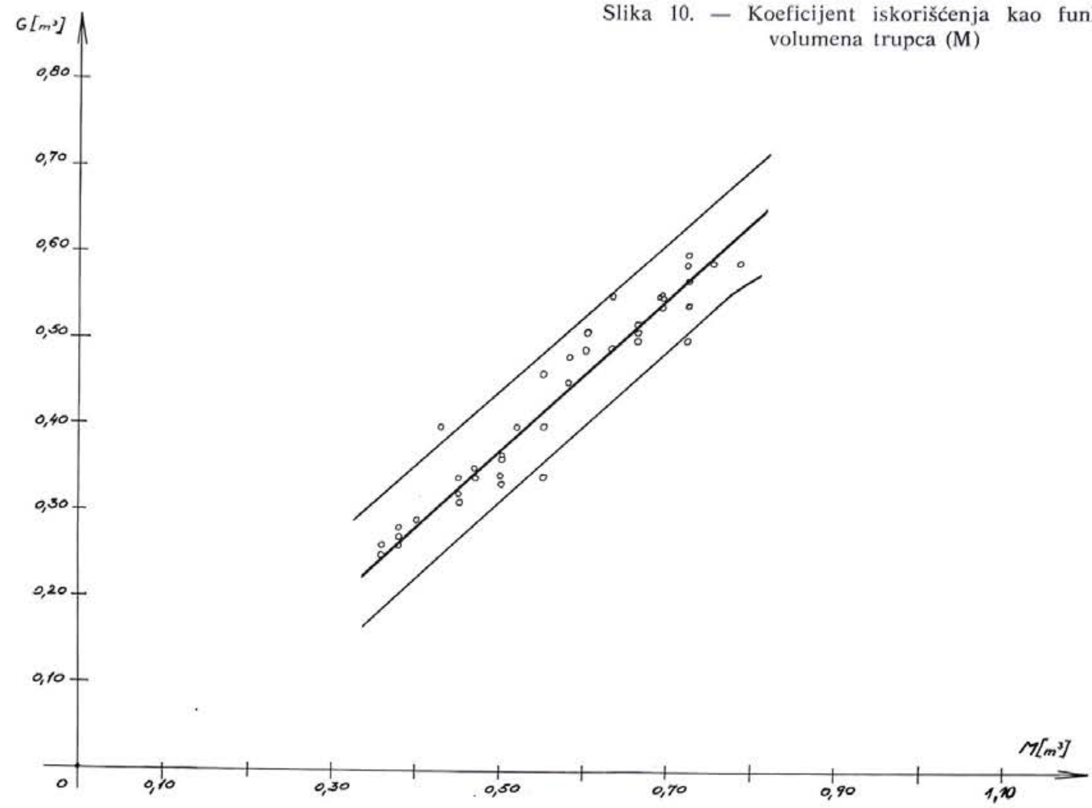
što omogućuje odbacivanje H_0 hipoteze uz kritični nivo testa manji od 1%.

Zaključujemo, dakle, da građa nije proporcionalna s volumenom trupca, te jednostavnim dijeljenjem jednadžbe regresije s M dobivamo koeficijent iskorišćenja kao funkciju od M:

$$k\ (M) = —0,0612/M + 0,874.$$

Ta je funkcija nacrtana na slici 10.

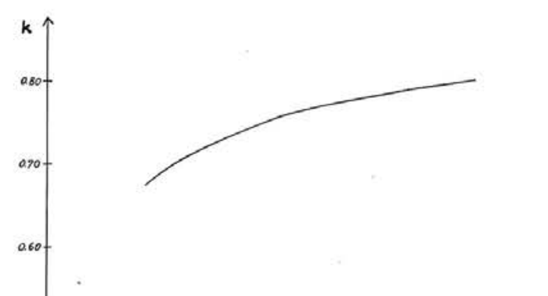
Dobili smo poznati oblik veze između koeficijenta iskorišćenja i volumena trupca (vidi literaturu pod (3) i (4)). Statističke metode linearne regresije omogućuju dalju analizu tih odnosa. O kojim se odnosima radi, bilo je spomenuto u točki 2. ovog rada.



Slika 10. — Koeficijent iskorišćenja kao funkcija volumena trupca (M)

4. DODATAK

Ako je interval u kojem se kreće vrijednost obilježja X dovoljno uzak, odnosno ako se radi o slučajevima prikazanim na slici 2 ili 3, tada se analiza svodi na analizu varijable Y, s time da se ima na umu da dobivene zakonitosti vrijede samo za taj uzak interval, odnosno da vrijede svugdje potpuno neovisno o vrijednosti X.



Slika 9. — Oblak raspršenja, pravac regresije i granice pouzdanosti za trupce iz primjera 2.

LITERATURA:

1. Snedecor, G. W. i Cochran, W. G.: Statistical Methods, Ames, Iowa, USA, 1976.
2. Vila, A. i Leicher, Z.: Planiranje proizvodnje i kontrola rokova, Informator, Zagreb, 1972.
3. Knežević, M.: Racionalna prerada drveta na gateru, Beograd, 1976.
4. Knežević, M.: Najbolje kvalitativno iskorišćenje trupaca pri rezanju jednakih debljina dasaka, Glasnik Šum. fak., br. 8/1955.