Boštjan Rode, dipl. inž. matem.
Železarna Ravne

ASM/SLA: S 12j; U 4 k
DK: 519.28 : 681.3

Statistična analiza regresije z uporabo elektronskih računalnikov

V članku sta opisana dva programa iz področja analize regresije. Prvi program je narejen po metodi »Korak za korakom« in določi le pomembne spremenljivke, ki ena za drugo vstopajo v regresijsko enačbo.

Program je razložen na primerih rezultatov iz elektronskega računalnika ZUSE Z-23.

Prikazani so tudi nomogrami teh primerov, ki nam grafično predočijo rezultate. Taki nomogrami so se pri raziskovalnem delu v Železarni Ravne zelo udomačili.

Drugi program izračuna običajne in parcialne koeficiente koleracije.

Tudi ta program je razložen na praktičnem primeru iz Železarne Ravne.

UVOD

Pri modernem raziskovalnem delu z uporabo metod matematične statistike ima analiza regresije zelo pomembno vlogo[1]. Z njo ugotavljamo medsebojne odvisnosti in učinkovitost vplivnih faktorjev. Uporabimo jo lahko kot metodo za statistično obdelavo podatkov iz urejene dokumentacije, še bolj učinkovita pa je, če jo vključimo v strategijo planiranih raziskav[2], za katere po vnaprej pripravljenem sistemu v planu raziskav zbiramo podatke in to metodo primerno kombiniramo z uporabo drugih metod matematične statistike[3, 4, 5].

Najpreprostejše primere analize regresije lahko za razlago in boljše razumevanje osnovnih principov računsko izvedemo na preprost in razumljiv način. Prav uporabna pa je postala statistična analiza regresije šele s pomočjo elektronskih računalnikov. Ti po posebno pripravljenih programih opravijo delo, ki brez njih skoraj ni izvedljivo. Pri tem gre za veliko število spremenljivk in razne oblike nelinearnih medsebojnih odvisnosti.

Program je treba prilagoditi specialnim potrebam in ga primerno razviti. Ob primerno pripravljenem programu, zadostni kapaciteti računalnika in posebni dodatni iznajdljivosti programerja število spremenljivk in število podatkov skoraj ni omejeno.

Programi iz standardnih bibliotek računalniških sistemov največkrat zaradi svoje splošne in široko uporabne oblike ne zadovoljujejo zahtev nekaterih specialnih potreb raziskovalnega dela na določenih področjih.

V Železarni Ravne smo že v letu 1961 začeli za potrebe metalurških raziskav uvajati metodo analize regresije. Ker nismo imeli možnosti uporabe računalnika in ker matematika s programiranjem takrat še ni našla mesta v metalurških raziskavah, smo te analize opravljali po preprostem in precej poenostavljenem računskem postopku z uporabo namiznih pisarniških računskih strojev za osnovne operacije. To so bile le linearne regresije dveh spremenljivk, če pa je šlo za množično regresijo, smo nato več linearnih regresij s pomočjo regresijskih koeficientov kombinirali v enačbo množične regresije. Zaradi praktične uporabnosti smo v postopek računanja uvedli več poenostavitev, razen tega pa je bilo tako računanje izpostavljeno številnim napakam. Obseg dela je prerasel možnosti takega izračunavanja, zato smo že v prvih letih delovanja računalnika ZUSE Z-23 na Računskem centru v Ljubljani začeli iskati možnosti uporabe računalnika. Prvi korak smo ob težkem začetnem sporazumevanju metalurgov in matematikov napravili tako, da smo za postopek, ki smo ga že uporabljali, izdelali program, ki je le zamuden računski postopek zamenjal z izračunom na hitrem računalniku. Že to je pomenilo napredek, čeprav zelo skromen. Kmalu smo postali nezadovoljni z reševanjem linearnih regresij, ker je mnogo primerov, za katero je že tehnološko popolnoma jasno, da jih ni mogoče reševati s predpostavljeno linearno odvisnostjo.

Začeli smo uporabljati standardni program analize regresije na računalniku Elliot 803 v Metalurškem inštitutu Zenica.

Z letom 1967 smo v Železarni Ravne začeli z intenzivnim delom pri programiranju in širjenju uporabe računalnikov v raziskovalnem delu. Za

---

* Opomba:

Program za analizo regresije na računalnikih je bil izdelan v okviru raziskovalnega projekta »Uvajanja metod matematične statistike v kontrolo kvalitete in metalurške raziskave«, ki ga v Železarni Ravne vodi Jože Rodič, dipl. inž. met., vodja metalurškega oddelka. Pri izdelavi programa za analizo regresije je neposredno sodeloval z usmerjanjem razvoja in izboljšav programa za zadovoljevanje potreb metalurških raziskav.

Avtor članka, Boštjan Rode — matematik v raziskovalnem oddelku se ob tej publikaciji zahvaljuje tov. Zvonimirju Bohtetu iz matematičnega oddelka FNT za koristne napotke pri izdelavi programa za analizo regresije »Korak za korakom« in tov. Janezu Mencingerju za posredovanje metode izračuna koeficientov parcialne korelacije, primerne za elektronski računalnik.

tako delo je potreben matematik — programer, ki je v stalnem stiku z raziskovalci in njihovimi tehničnimi problemi. S standardnim programom iz biblioteke Elliot smo kmalu postali nezadovoljni in iskali smo novih, boljših možnosti. Pri nadaljnjem delu smo uporabljali računalnike ZUSE Z 23, CDC, IBM 360/30 in IBM 1130.

V članku želimo v kratkem opisati značilnosti, teoretične osnove in praktično uporabo dveh programov za analizo regresije, ki našim potrebam metalurškega raziskovalnega dela zaenkrat najbolje ustrezata. Ta dva programa redno uporabljamo in smo z njima rešili že mnogo praktičnih problemov s področja raziskav, tehnologije proizvodnje in lastnosti jekel.

Prvi program je narejen po metodi »korak za korakom« (step by step), ki ima velike prednosti pred ostalimi metodami analize regresije v tem, da določi le pomembne regresijske koeficiente, ostalih pa sploh ne računa. Drugi program pa izračuna običajne in parcialne koeficiente korelacije, ki nam omogočijo zelo pregledno sliko medsebojnih vplivov po dveh in dveh faktorjev ne glede na stanje drugih.

V industriji je mnogo proizvodnih procesov, pri katerih so lastnosti ali količine določenih proizvodov odvisne od števílnih vplivnih faktorjev, med katerimi jih nekaj kontroliramo in nekaj ne. Določen proces bomo bolje uravnavali, če bomo podrobneje spoznali velikost vpliva različnih faktorjev na kvaliteto proizvodov iz procesa. Pogosto je edini ali pa najboljši laboratorij, ki je na razpolago za študij vplivnih faktorjev, sam proizvodni proces, kjer je iz tehničnih in ekonomskih vzrokov največkrat nemogoče urediti vse potrebno za sistematični študij vsakega vplivnega faktorja. Največ, kar v takih primerih lahko naredimo, je, da med tekočo proizvodnjo zasledujemo velikost vplivnih faktorjev in kvaliteto vmesnih in končnih proizvodov. To skušamo narediti na tak način, da s tem ne preprečujemo normalnega poteka proizvodnje, ki naj bo tudi osnova raziskave.

Iz tako zbranih podatkov lahko ugotovimo zvezo med spreminjanjem velikosti vplivnih faktorjev in lastnostmi proizvodov. Metoda, ki jo pri tem uporabimo je analiza regresije.

Vrednosti določene karakteristike proizvoda, ki smo jih izmerili, vzamemo kot vrednosti odvisne spremenljivke Y. Ustrezne vrednosti vplivnih faktorjev pa so vrednosti neodvisnih spremenljivk $X_1, X_2 \ldots X_m$. Dobiti hočemo linearno zvezo

$$Y = b_o + b_1X_1 + b_2X_2 + \ldots + b_mX_m \qquad (1)$$

Pri tem so $b_1, b_2 \ldots b_m$ regresijski koeficienti, ki naj nam povedo, kako sprememba vsakega vplivnega faktorja X vpliva na spremembo vrednosti opazovane karakteristike proizvoda — spremenljivke Y.

V enačbi je b_o konstantna vrednost spremenljivke Y, ki ni pojasnjena s spreminjanjem X-ov v enačbi (1). Enačbo (1) imenujemo regresijska zveza ali enačba regresije.

Iz zbranih podatkov za vrednosti odvisne in neodvisnih spremenljivk z metodo analize regresije vedno lahko določimo regresijske koeficiente b. Vprašanje je le, ali zveza, ki smo jo tako dobili med vplivnimi faktorji in kvaliteto proizvoda, res obstaja, ali pa smo dobili take vrednosti za koeficiente b, čeprav zveze, ki jo le-ti določajo, v resnici ni. Kakor temu pravimo, ugotoviti hočemo statistično pomembnost regresijske zveze. V praksi navadno smatramo da je rezultat uporaben, če je verjetnost, da nastopi le zaradi slučaja, manjša od 5 %. Tak rezultat imenujemo tudi 95 % pomemben, ali dobljen na 5 % nivoju pomembnosti.

Podobno lahko dobimo tudi, da je rezultat 99 % pomemben ali 99,9 % pomemben, če je verjetnost, da nastopi le zaradi slučaja samo 1 % ali samo 0,1 %.

Pri izbiri vplivnih faktorjev za neodvisne spremenljivke moramo paziti, da nobeden med njimi ne vpliva na kateri drug izbran vplivni faktor, ker sicer ne moremo prav določiti pomembnosti regresijske zveze. S posebno metodo parcialnih koeficientov korelacije moremo določiti tiste pare vplivnih faktorjev, kjer vrednost enega faktorja močno vpliva na vrednost drugega. Potem enega od obeh faktorjev v paru ne vzamemo med neodvisne spremenljivke.

V tej publikaciji si bomo ogledali uporabo dveh programov za analizo regresije na elektronskem računalniku. Namen prvega programa je samo določitev pomembnih regresijskih koeficientov, medtem ko nepomembne vzamemo enake 0. Z drugim programom pa lahko izračunamo le vse regresijske koeficiente skupaj in parcialne koeficiente korelacije med posameznimi spremenljivkami.

POMEN REGRESIJSKE ZVEZE

Vzemimo, da imamo podatke za analizo regresije zbrane v tabeli,

X_1	X_2	. . . . . .	X_s
x_{11}	x_{21}		x_{s1}
x_{12}	x_{22}		x_{s2}
.	.		.
.	.		.
.	.		.
x_{1n}	x_{2n}		x_{sn}

Tabela 1

kjer so $X_1, X_2, \ldots X_s$ spremenljivke, x_{ij} pa njihove vrednosti. Vsaka spremenljivka ima v praksi tudi svoje ime, kot npr.: trdota, žilavost, ogljik, temperatura popuščanja ... in enoto: HRC, kpm/cm², %, °C ...

Vrednosti x_{ij} so v praksi števila, ki smo jih izmerili za vsako opazovano karakteristiko — spremenljivko.

V isti vrsti leže števila, izmerjena v isti enoti proizvodnega procesa. Primer proizvodnega procesa je npr.: proizvodnja določene vrste jekla. Števila, izmerjena za opazovane karakteristike (dodatki legur, temperatura izpusta, trdota ...) v isti šarži, bi ležala v isti vrsti tabele 1. Število vrst n pa je potem enako številu šarž.

Med spremenljivkami $X_1, X_2, \ldots X_s$, ki jih imenujemo tudi prvotne, lahko predpostavimo **najrazličnejše** zveze. Tisto izmed spremenljivk, ki jo vzamemo za odvisno, bomo označili z Y. Neodvisne spremenljivke pa bomo zaradi enostavnosti zapisa formul označili kar po vrsti z $X_1, X_2, \ldots X_m$, čeprav so to lahko katerekoli od prvotnih spremenljivk ali celo njihovi logaritmi, recipročne vrednosti ali produkti. Vzemimo za primer, da smo predpostavili odvisnost

$$Y = b_o + b_1X_1 + b_2X_1^2 + b_3 \ln X_2 + b_4 \frac{1}{X_3} + b_5 . \frac{1}{X_3} . \ln X_2$$

Neodvisne spremenljivke, ki jih zaradi enostavnosti označimo z $X_1, X_2, \ldots X_5$, so tu po vrsti

$$X_1, X_1^2, \ln X_2, \frac{1}{X_3}, \frac{1}{X_3} \ln X_2.$$

Zgornjo enačbo bi s tem dogovorom enostavno zapisali:

$$Y = b_o + \sum_{i=1}^{5} b_i X_i$$

Vrednosti odvisne spremenljivke označimo z $y_1, y_2, \ldots y_n$ in poljubno vrednost med njimi z y_j.

Kako bomo v enačbi (1) določili regresijske koeficiente $b_1, b_2 \ldots b_m$ in konstanto b_o?

Vzemimo, da jih že poznamo. Z $\hat{y}_j$, $j = 1, 2 \ldots n$ označimo vrednosti spremenljivke Y, ki jih izračunamo iz regresijske zveze (1) tako, da za neodvisne spremenljivke $X_1, X_2, \ldots X_m$ vstavimo njihove vrednosti iz tabele 1.

$$\hat{Y}_j = b_o + b_1x_{1j} + b_2x_{2j} \ldots + b_mx_{mj} \quad j = 1, 2 \ldots n \quad (2)$$

Želimo, da bi bile razlike med y_j in $\hat{y}_j$ čim manjše. Ta pogoj napišemo takole:

$$\sum_{n}^{j=1} (y_j - \hat{y}_j)^2 = \text{minimum} \quad (3)$$

Regresijske koeficiente $b_1, b_2 \ldots b_m$ in konstanto b_o določimo tako, da je pogoj (3) izpolnjen. Pravimo, da smo jih določili po metodi najmanjših kvadratov. Seveda ni nujno, da regresijske koeficiente določimo prav na ta način. Lahko nekatere kar predpostavimo enake nič, druge pa določimo po metodi najmanjših kvadratov. Posebno metodo določitve bomo spoznali kasneje pri opisu programa. Ko poznamo regresijske koeficiente in konstanto, lahko izračunamo vrednosti $\hat{y}_j$ za vse $j = 1, 2 \ldots n$ in nato še levo stran enačbe (3), ki jo bomo označili z V in jo imenujemo **nepojasnjena varianca** ali vsota kvadratov odstopanj od regresijskih vrednosti.

$$V = \sum_{j=1}^{n} (y_j - \hat{y}_j)^2 \quad (4)$$

Nepojasnjena jo imenujemo zato, ker odstopanja nismo pojasnili z regresijsko zvezo.

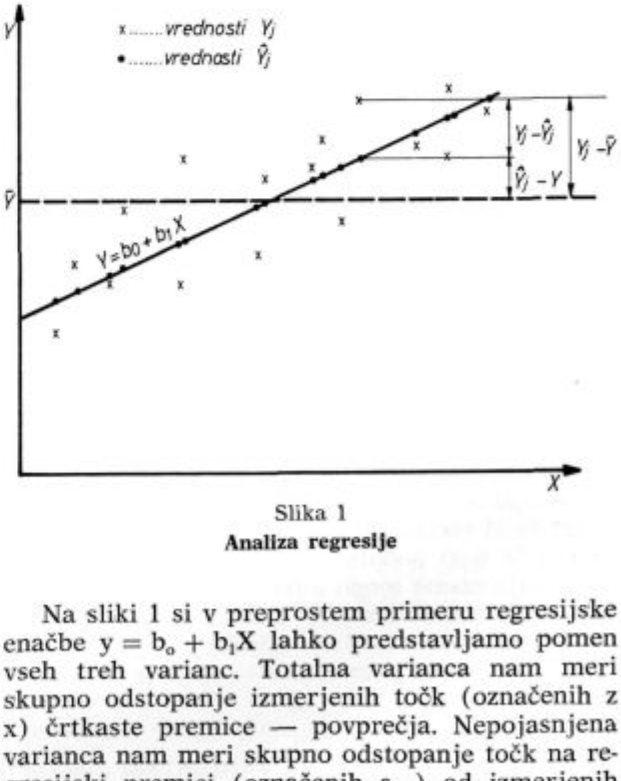
Srednjo vrednost vseh vrednosti y_j označimo z $\bar{y}$. Izračunamo lahko vsoto kvadratov odstopanj od srednje vrednosti.

$$K = \sum_{j=1}^{n} (y_j - \bar{y})^2 \quad (5)$$

Vrednost K imenujemo tudi **totalna ali celotna varianca.** Razliko vrednosti $K - V$ imenujemo **pojasnjena varianca.** Zakaj to ime? Dokazali bi lahko, da je

$$K - V = \sum_{j=1}^{n} (\hat{y}_j - \bar{y})^2 \quad (6)$$

kar je vsota kvadratov odstopanj regresijskih vrednosti od srednje vrednosti.



Slika 1
Analiza regresije

Na sliki 1 si v preprostem primeru regresijske enačbe $y = b_o + b_1X$ lahko predstavljamo pomen vseh treh varianc. Totalna varianca nam meri skupno odstopanje izmerjenih točk (označenih z x) črtkaste premice — povprečja. Nepojasnjena varianca nam meri skupno odstopanje točk na regresijski premici (označenih s .) od izmerjenih točk.

Pojasnjena varianca pa nam meri odstopanje točk na regresijski premici (prej imenovane regresijske vrednosti) od črtkaste premice.

Razmerje pojasnjene in totalne variance imenujemo **determinacijski koeficient** in ga označimo z R^2.

$$R^2 = \frac{K - V}{K} = 1 - \frac{V}{K} \qquad (7)$$

Determinacijski koeficient nam pove, kakšen del celotnega razsipanja vrednosti smo pojasnili z enačbo regresije. Kvadratni koren iz determinacijskega koeficienta imenujemo **koeficient množične korelacije R.** Ta nam meri jakost regresijske zveze. Njegova vrednost leži med 0 in 1. Če je $R = 0$, pravimo, da med odvisno spremenljivko in neodvisnimi ni zveze, ki smo jo predpostavili. V praksi ne bo R nikoli 0, pač pa je lahko zelo majhen. Seveda pa je $R = 0$, če vzamemo vse regresijske koeficiente enake 0. Pri $R = 1$, pravimo, da je odvisnost popolna. Tudi tega primera v praksi nikoli ne dosežemo, vendar se veselimo primerov z velikimi R, ker nam pojasnijo medsebojne odvisnosti in jakosti vplivov.

Zanima nas tudi pomembnost regresijske enačbe, o čemer smo že govorili. Pomembnost vseh regresijskih koeficientov skupaj ugotovimo s tako imenovanim **F — testom.** Izračunati moramo vrednost

$$R = \frac{R^2\,(n - m - 1)}{m\,(1 - R^2)} \qquad (8)$$

m = število neodvisnih spremenljivk

n = število podatkov za vsako spremenljivko

Nato pogledamo v tabele[4] za funkcijo $F\alpha$; ν_1, ν_2 za željen nivo pomembnosti (npr.: $\alpha = 5$) in prostostni stopnji $\nu_1 = m$ in $\nu_2 = n - m - 1$. Če je vrednost iz tabel manjša od vrednosti, izračunane po obrazcu (8), potem je enačba regresije $(100 - \alpha)$ % pomembna (npr.: 95 % pomembna). V primeru 95 % pomembnosti primerjamo vrednost F, izračunano po obrazcu (8) še z vrednostjo iz tabel za $\alpha = 1$. Tako ugotovimo, ali je enačba regresije tudi 99 % pomembna. V primeru, ko pa ni 95 % pomembnosti, primerjamo izračunani F še z vrednostjo iz tabel za $\alpha = 10$. Nivo pomembnosti, ki ga izberemo za primerjavo zavisi od primera do primera in je odvisen tudi od naših izkušenj. Če nismo ugotovili pomembnosti regresijske zveze, pa s tem še ni rečeno, da te zveze v resnici ni. Lahko so bili le naši podatki tako slabi, da je z večjo gotovostjo nismo mogli odkriti. Morda bi s ponovnim pazljivim zbiranjem zadostnega števila podatkov lahko odkrili, da je že prej predpostavljena regresijska zveza pomembna — to je gotova.

PROGRAM »KORAK ZA KORAKOM« ZA ANALIZO REGRESIJE[8]

To je program, za katerega smo v uvodu rekli, da določi le pomembne regresijske koeficiente. Uporabnost programa si bomo ogledali na praktičnem primeru v splošni obliki. Zbranih imamo po 30 vrednosti v obliki tabele 1 za 19 spremenljivk ($n = 30$, $s = 19$) $X_1, X_2, \ldots X_{19}$. Poleg tega imamo tudi seznam funkcijskih odvisnosti, ki jih predpostavimo glede na praktične izkušnje in jih želimo kvantitativno spoznati.

Plan analize regresije — praktičen primer

$$X_2 = f\,(X_1, X_3, X_4, X_6, X_8, X_9, X_{11})$$
$$X_2' = f\,(X_{12}, \ldots X_{19})$$
$$X_3 = f\,(X_1)$$
$$X_4 = f\,(X_1)$$
$$X_5 = f\,(X_1)$$
$$X_6 = f\,(X_1, X_3, X_4)$$
$$X_7 = f\,(X_2, X_3, X_4, X_6, X_8, X_9, X_{11})$$
$$X_7' = f\,(X_{12}, \ldots X_{19})$$
$$X_8 = f\,(X_2, X_3, X_4, X_6, X_7, X_9, X_{11})$$
$$X_8' = f\,(X_{12}, \ldots X_{19})$$
$$X_9 = f\,(X_2, X_3, X_4, X_6, X_7, X_8, X_{11})$$
$$X_9' = f\,(X_{12}, \ldots X_{19})$$
$$X_{11} = f\,(X_2, \ldots X_4, X_6, \ldots X_9)$$
$$X_{11}' = f\,(X_{12}, \ldots X_{19})$$
$$X_{16} = f\,(X_7, X_8)$$
$$X_{16}' = f(X_{12}, \ldots X_{15}, X_{17}, \ldots X_{19})$$
$$X_{17} = f\,(X_7, X_8)$$
$$X_{17}' = f\,(X_{12}, \ldots X_{16}, X_{18}, X_{19})$$
$$X_{18} = f\,(X_3, X_4, X_6 \ldots X_9, X_{11})$$
$$X_{18}' = f\,(X_{12}, \ldots X_{17} \,.\, X_{19})$$
$$X_{19} = f\,(X_1, X_2)$$
$$X_{12} = (X_1, \ldots X_4, X_6, \ldots X_9, X_{11}, X_{16}, X_{17}, X_{19})$$
$$X_{13} = f\,(X_3, X_5, X_6, X_8, X_{19})$$
$$X_{15} = f\,(X_{19})$$

V seznamu opazimo, da so nekatere spremenljivke dvakrat odvisne, vedno od druge skupine neodvisnih spremenljivk. Zato za take odvisne spremenljivke nekrat uporabimo ' poleg njihovega simbola X.

Določiti moramo še oblike funkcijskih zvez. V našem primeru naj bodo to polinomi druge stopnje brez mešanih produktov spremenljivk. Tako iščemo v tem primeru odvisnost npr.:

$X_6 = f\,(X_1, X_3, X_4)$ v obliki

$$X_6 = b_0 + b_1X_1 + b_2X_3 + b_3X_4 + b_4X_1^2 + b_5X_3^2 + b_6X_4^2 \qquad (9)$$

V računalnik bomo podali podatke tako, da se bo ta po programu pripravil na izračun vseh 24 regresijskih enačb iz seznama. Zato bo moral najprej izračunati kvadrate vrednosti spremenljivk

$$X_1, X_2 \ldots X_9, X_{11}, X_{12} \ldots X_{19}$$

(X_{10} ne nastopa nikjer v seznamu funkcijskih odvisnosti, zato kvadratov njenih vrednosti računalnik ne potrebuje, čeprav imamo X_{10} v podatkih.) Teh 18×30 kvadratov bo računalnik shranil v spomin naprej od 19×30 vrednosti prvotnih spremenljivk. Tako bo vrstni red spremenljivk za računalnik naslednji:

Tabela 2

1'	X_1
2'	X_2
3'	X_3
.	.
.	.
.	.
19'	X_{19}
20'	X_1^2
21'	X_2^2
22'	X_3^2
23'	X_4^2
24'	X_5^2
25'	X_6^2
26'	X_7^2
27'	X_8^2
28'	X_9^2
29'	X_{11}^2
30'	X_{12}^2
31'	X_{13}^2
32'	X_{14}^2
33'	X_{15}^2
34'	X_{16}^2
35'	X_{17}^2
36'	X_{18}^2
37'	X_{19}^2

To moramo imeti stalno pred očmi, tako za določitev odvisne in neodvisnih spremenljivk vsake funkcijske zveze, kot tudi za kasnejše tolmačenje rezultatov. Npr.: za funkcijsko zvezo $X_6 = f(X_1, X_3, X_4)$ računalniku določimo odvisno spremenljivko s številko 6', neodvisne spremenljivke pa določimo s števili 1' 3' 4' 20' 22' 23', ki nastopajo tudi v rezultatih poleg regresijskih koeficientov b_1, $b_2 \ldots b_6$ iz enačbe (9). Tako je npr. koeficient b_6 v rezultatih določen s številko 23'.

Podatke za računalnik pripravimo v posebni obliki tako, da računalniku »povemo«, kaj naj z njimi naredi, preden začne računati regresijske koeficiente za posamezne funkcijske zveze.

Računalnik računa regresijske koeficiente enega za drugim. V začetku so vsi enaki nič, kar pomeni, da nobene neodvisne spremenljivke ni v regresijski enačbi. Nato vstopajo spremenljivke ena za drugo v regresijsko enačbo in ustrezni koeficienti dobijo vrednosti različne od nič. Vsak vstop spremenljivke v regresijsko enačbo in izračun vseh regresijskih koeficientov za spremenljivke, ki so že v regresijski enačbi, imenujemo korak v računanju. V vsakem koraku vstopi spremenljivka, ki najbolj zmanjša nepojasnjeno varianco. Vstopajo spremenljivke samo toliko časa, dokler je njihov doprinos k zmanjšanju nepojasnjene variance pomemben.

Po nekaj korakih določena spremenljivka, ki je prej vstopila, lahko tudi izstopi iz regresijske enačbe, če njen doprinos k pojasnjeni varianci ni več pomemben. To pomembnost pri vstopu in izstopu spremenljivk urejata dve števili F_1 in F_2, ki ju navadno vzamemo obe 4, kar ustreza 95 % pomembnosti v primerih, ko je število podatkov za eno spremenljivko vsaj 30. Pri manjšem številu podatkov moramo F_1 povečati. Na koncu tega članka je tabela 3 za $F\alpha$; 1, ν iz katerih dobimo našo vrednost F_1 za določen nivo pomembnosti

Slika 2

Protokol rezultatov analize regresije za odvisnost
$x_9 = f(x_2, x_3, x_4, x_6, x_7, x_8, x_{11})$

```
KONTROLNE  VSOTE
.448260000_10+06     .152400000_10+04     .500700000_10+04
.136210000_10+05     .112969999_10+03     .288000000_10+02
.133199999_10+01     .445100000_10+02


F1= .400000_10+01
F2= .400000_10+01

STD. DEV. Y .358853_10-00


 1'
VSTOPA  X  8'
IZRACUNANI  F  .143925_10+03
STD. DEV. Y  .147382_10-00


KONSTANTA .880516_10-00

SPREM. KOEFICIENT      STD. NAP. KOEF.
 8'    .869928_10-00   .725128_10-01

KOEF. DET.     KOEF. KOREL.
.837138_10-00  .914952_10-00


 2'
VSTOPA  X  11'
IZRACUNANI  F  .685124_10+01
STD. DEV. Y  .134040_10-00


KONSTANTA .111846_10+01

SPREM. KOEFICIENT      STD. NAP. KOEF.
 8'    .818978_10-00   .687614_10-01
11'   -.456370_10+01   .174354_10+01

KOEF. DET.     KOEF. KOREL.
.870100_10-00  .932791_10-00
```

```
3'
VSTOPA  X  29'
IZRACUNANI  F  .863822_10+01
STD. DEV.  Y  .118342_10-00

KONSTANTA  .168195_10+01

SPREM.  KOEFICIENT      STD. NAP. KOEF.
 8'     .792119_10-00   .613923_10-01
11'    -.276481_10+02   .800370_10+01
29'     .219666_10+03   .747398_10+02

KOEF. DET.      KOEF. KOREL.
.902495_10-00   .949997_10-00


4'
VSTOPA  X  7'
IZRACUNANI  F  .656314_10+01
STD. DEV.  Y  .107408_10-00

KONSTANTA  .189306_10-00

SPREM.  KOEFICIENT      STD. NAP. KOEF.
 7'     .394532_10-00   .154002_10-00
 8'     .758644_10-00   .572315_10-01
11'    -.271463_10+02   .726683_10+01
29'     .223281_10+03   .678489_10+02

KOEF. DET.      KOEF. KOREL.
.922770_10-00   .960609_10-00


5'
VSTOPA  X  26'
IZRACUNANI  F  .453595_10+01
STD. DEV.  Y  .100533_10-00

KONSTANTA  -.304702_10+02

SPREM.  KOEFICIENT      STD. NAP. KOEF.
 7'     .169999_10+02   .779812_10+01
 8'     .762020_10-00   .535919_10-01
11'    -.316193_10+02   .711858_10+01
26'    -.223768_10+01   .105066_10+01
29'     .268238_10+03   .669226_10+02

KOEF. DET.      KOEF. KOREL.
.935046_10-00   .966977_10-00
```

$\alpha = 10; 5; 1; 0{,}1$, če za ν vzamemo število podatkov n za eno spremenljivko minus 2 ($\nu = n - 2$). F_2 izberemo tako, da je $F_1 \geq F_2$.

Poglejmo si rezultate za funkcijsko zvezo $X_9 = f(X_2, X_3, X_4, X_6, X_7, X_8, X_{11})$ na sliki 2. Pod naslovom »Kontrolne vsote« so najprej vsote vseh vrednosti neodvisnih spremenljivk X_2, X_3, X_4, X_6, X_7 X_8, X_{11} in nazadnje vsota odvisne spremenljivke X_9, torej vsega skupaj 8 vsot. Za vsotami sledita izpisani vrednosti F_1 in F_2, ki smo jih določili s podatki.

Sledi standardna deviacija odvisne spremenljivke, ki jo računalnik označi z naslovom »STD. DEV. Y«. Nato imamo izpise v vsakem koraku. Najprej zaporedno številko koraka, nato spremenljivka, ki vstopa v regresijsko enačbo oziroma izstopa iz nje, opremljena z napisom »VSTOPA« oziroma »IZSTOPA«. Sledi izračunana vrednost F, ki jo je računalnik primerjal z F_1 oziroma F_2. »IZRAČUNANI F«, kot piše pred to vrednostjo, je pri vstopu vedno večji od F_1 in pri izstopu vedno manjši od F_2. Potem računalnik izpiše še standardno deviacijo odvisne spremenljivke, vendar sedaj ne več glede na srednjo vrednost $\overline{Y}$, temveč glede na vrednosti $\hat{y}_j$, ki jih že poznamo (npr.: iz enačbe (2), kjer so b_0, $b_1 \ldots b_m$ koeficienti spremenljivk, ki so vključno s tem korakom že vstopile v regresijsko enačbo). Tako od koraka do koraka zasledujemo manjšanje te vrednosti, ko spremenljivke vstopajo. Vrednost pa se le nepomembno poveča, ko katera od spremenljivk izstopi.

Po manjšem presledku sledi še izpis v istem koraku. Zraven napisa »KONSTANTA« izpiše računalnik vrednost b_o. Pod napisom »SPREM.« dobimo števila, ki v smislu tabele 2 določajo spremenljivke, za katere slede pod napisom »KOEFICIENT« regresijski koeficienti. Pod napisom »STD. NAP. KOEF.« dobimo standardne napake teh regresijskih koeficientov.

V vsakem koraku računalnik izpiše še trenutni koeficient determinacije R^2 pod napisom »KOEF. DET.« in koeficient korelacije R pod napisom »KOEF. KOREL.« Z vstopanjem spremenljivk se tudi ta dva koeficienta večata. Pri izstopu spremenljivke pa se navadno le neznatno zmanjšata.

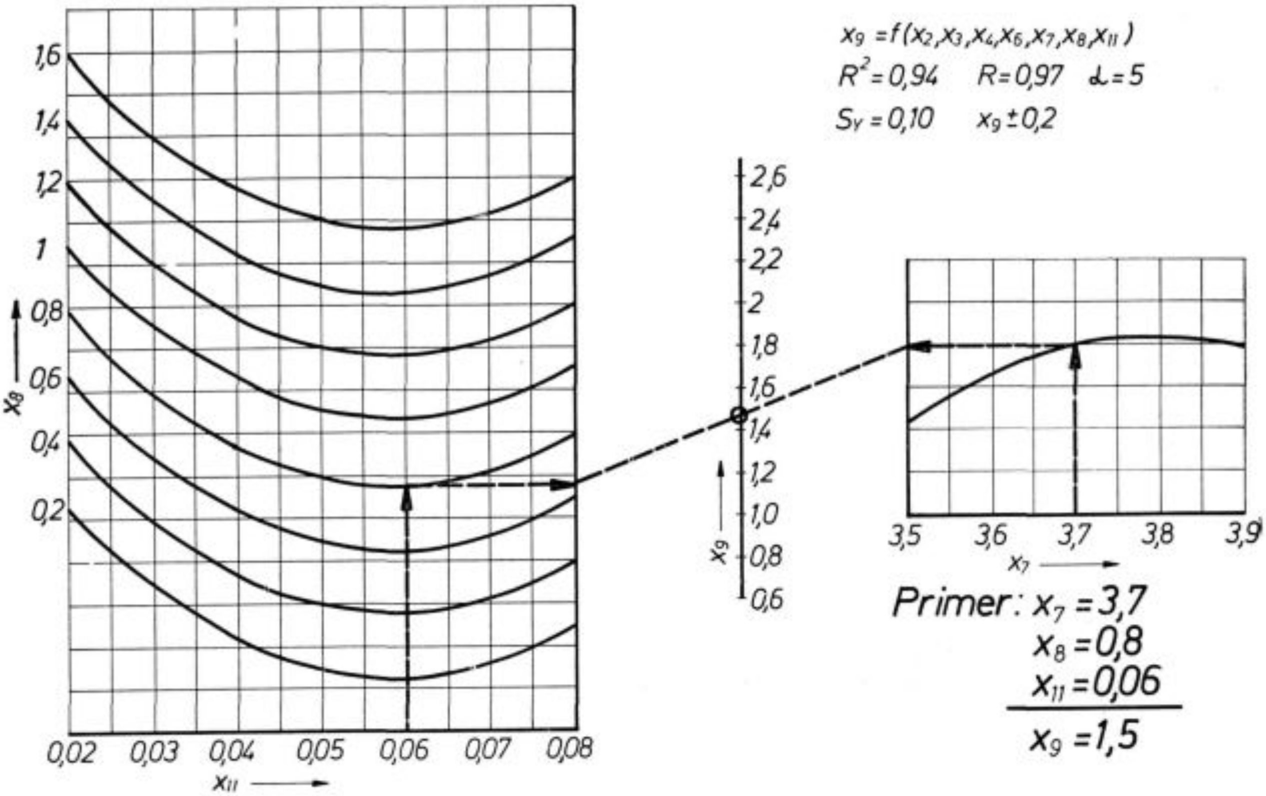
V našem primeru $X_9 = f(X_2, X_3, X_4, X_6, X_7 X_8, X_{11})$ je najprej vstopila spremenljivka X_8 z zelo velikim izračunanim $F = 144$ (zaokroženo na 3 mesta). Standardna deviacija odvisne spremenljivke se je zmanjšala na polovico iz 0,36 na 0,15. Koeficient determinacije je že v prvem koraku velik 0,84. Koeficient korelacije je z vrednostjo 0,915 tako že blizu 1.

V 2., 3., 4. in 5. koraku vstopajo po vrsti spremenljivke X_{11}, X_{11}^2 (z oznako 29' po tabeli 2!) X_7 in X_7^2 (z oznako 26'). Standardna deviacija X_9 se je zmanjšala do konca na 0,10 in koeficient determinacije je narastel na 0,94. Končna enačba regresije bi bila:

$$X_9 = -30{,}47 + 17\,X_7 - 2{,}238\,X_7^2 + 0{,}762\,X_8 - 31{,}62\,X_{11} + 268{,}2\,X_{11}^2. \quad (10)$$

Ker je v 5. koraku izračunani F = 4,5, moremo za to enačbo trditi le, da je 95 % pomembna. Po 1. koraku imamo regresijsko odvisnost $X_9 = 0,8805 + 0,8699\,X_8$, ki je 99,9 % pomembna, ker je izračunani F enak 144.

Pomembnost regresijske zveze v določenem koraku dobimo tako, da do vključno tega koraka poiščemo najmanjši izračunani F pri vstopu spremenljivke in ga primerjamo z vrednostjo $F\alpha$, 1, ν iz tabele 3 za določen nivo α in $\nu = n - 2$. Če je najmanjši izračunani F večji od vrednosti iz tabele, potem je regresijska enačba v tem koraku $(100 - \alpha)$ % pomembna. V našem primeru ni nobena regresijska enačba, razen prve več kot 95 % pomembna, ker je vrednost v tabeli 3 za F_1; 1,28 = 7,64 (glej tabele na koncu članka).



Slika 3
Primer nomograma za tri pomembne spremenljivke

Regresijsko enačbo (10) narišemo v nomogramu na sliki 3.

V nomogramu je narisan primer, kako iz vrednosti neodvisnih spremenljivk X_7, X_8 in X_{11} dobimo vrednost odvisne spremenljivke X_9. Na sliki so vpisane tudi vrednosti koeficientov R^2 in R, nivo pomembnosti ($\alpha = 5$), standardna deviacija odvisne spremenljivke S_y in območje 95 % gotovosti $X_9 \pm 0,2$. (To je približno $X_9 \pm 1,96 \,.\, S_y$.) Pas 95 % gotovosti pomeni, da le s 5 % verjetnostjo lahko pričakujemo dejansko vrednost odvisne spremenljivke zunaj mej $X_9 + 0,2$ in $X_9 - 0,2$, če za X_9 vzamemo vrednost, dobljeno iz nomograma. Tako je v našem primeru le 5 % verjetnosti, da dejanska vrednost X_9, ki bi jo izmerili v proizvodnem procesu pri vrednostih $X_7 = 3,7$, $X_8 = 0,8$ in $X_{11} = 0,06$ leži zunaj mej 1,3 in 1,7. Polovično širino pasu 95 % gotovosti izračunamo tako, da S_y pomnožimo z 1,96. V našem primeru 0,10 . 1,96 = 0,20, če zaokrožimo na 2 decimalki. Taka določitev tolerančnega območja je le približna in prav dobro upo-

Slika 4
Protokol rezultatov analize regresije za odvisnost $x_{14} = f(x_7, x_8)$

```
KONTROLNE  VSOTE
.112969999 10 +03     .208000000 10 +02     .274999999 10 +02


F1=  .400000 10 +01
F2=  .400000 10 +01

STD.  DEV.  Y  .295347 10 -00



KONSTANTA  .916666 10 -00

SPREM.  KOEFICIENT     STD. NAP. KOEF.

KOEF. DET.     KOEF. KOREL.
0              0
```

rabna, čeprav ne ustreza popolnoma teoriji, ki je dosti bolj zapletena. Iz nomograma se vidi tudi jakost vpliva posamezne neodvisne spremenljivke. V našem primeru očitno najbolj vpliva spremenljivka X_8, najmanj pa X_7. To se ujema tudi z vrstnim redom vstopanja spremenljivk v regresijsko enačbo.

Oglejmo si še primer rezultatov za funkcijsko zvezo $X_{16} = f(X_7, X_8)$ na sliki 4. Nobena od spremenljivk X_7, X_7^2, X_8, X_8^2 ni vstopila v regresijsko enačbo, zato so vsi regresijski koeficienti ostali enaki 0 in iz računalnika ne dobimo nobenega zapisa. Konstanta je srednja vrednost odvisne spremenljivke X_{16}, o čemer se prepričamo tudi, če zadnjo kontrolno vsoto delimo s številom podatkov 30.

$27\lambda : 30 = 0{,}916666\ldots$

Koeficienta determinacije in korelacije sta enaka 0.

Zanimivi so tudi rezultati za funkcijsko zvezo $X_{17}' = f(X_{12}, \ldots X_{16}, X_{18}, X_{19})$ na sliki 5.

Slika 5

Protokol rezultatov analize regresije za odvisnost
$x_{17} = f\ (x_{12} \ldots . x_{16},\ x_{18},\ x_{19})$

KONTROLNE VSOTE

$.113760999_{10}+04$	$.528549999_{10}+03$	$.958020000_{10}+03$
$.201820002_{10}+03$	$.274999999_{10}+02$	$.934000000_{10}+01$
$.223500000_{10}+06$	$.122050000_{10}+03$	

F1= $.400000_{10}+01$
F2= $.400000_{10}+01$

STD. DEV. Y $.890397_{10}-00$

1'
VSTOPA X 31'
IZRACUNANI F $.163049_{10}+02$
STD. DEV. Y $.720371_{10}-00$

KONSTANTA $.817038_{10}+01$

SPREM.	KOEFICIENT	STD. NAP. KOEF.
31'	$-.131605_{10}-01$	$.325921_{10}-02$

KOEF. DET.	KOEF. KOREL.
$.368016_{10}-00$	$.606643_{10}-00$

2'
VSTOPA X 14'
IZRACUNANI F $.158183_{10}+02$
STD. DEV. Y $.582533_{10}-00$

KONSTANTA $.181209_{10}+02$

SPREM.	KOEFICIENT	STD. NAP. KOEF.
14'	$-.295160_{10}-00$	$.742127_{10}-01$
31'	$-.148445_{10}-01$	$.266938_{10}-02$

KOEF. DET.	KOEF. KOREL.
$.601489_{10}-00$	$.775557_{10}-00$

3'
VSTOPA X 15'
IZRACUNANI F $.363296_{10}+02$
STD. DEV. Y $.383402_{10}-00$

KONSTANTA $.280381_{10}+02$

SPREM.	KOEFICIENT	STD. NAP. KOEF.
14'	$-.473878_{10}-00$	$.571395_{10}-01$
15'	$-.626735_{10}-00$	$.103980_{10}-00$
31'	$-.177121_{10}-01$	$.182016_{10}-02$

KOEF. DET.	KOEF. KOREL.
$.833766_{10}-00$	$.913108_{10}-00$

4'
VSTOPA X 12'
IZRACUNANI F $.265459_{10}+02$
STD. DEV. Y $.272298_{10}-00$

KONSTANTA $.596403_{10}+02$

SPREM.	KOEFICIENT	STD. NAP. KOEF.
12'	$-.567498_{10}-00$	$.110145_{10}-00$
14'	$-.709683_{10}-00$	$.611675_{10}-01$
15'	$-.771441_{10}-00$	$.790091_{10}-01$
31'	$-.198897_{10}-01$	$.136005_{10}-02$

KOEF. DET.	KOEF. KOREL.
$.919376_{10}-00$	$.958841_{10}-00$

```
 5'
VSTOPA  X  33'
IZRACUNANI  F  .899399_10+Ø1
STD.  DEV.  Y  .237Ø26_10-ØØ


KONSTANTA  .7Ø6764_10+Ø2

SPREM.  KOEFICIENT      STD. NAP. KOEF.
12'    -.6Ø773Ø_10-ØØ    .968116_10-Ø1
14'    -.718817_10-ØØ    .533313_10-Ø1
15'    -.347247_10+Ø1    .9Ø3268_10-ØØ
31'    -.196366_10-Ø1    .118688_10-Ø2
33'     .193349_10-ØØ    .644714_10-Ø1

KOEF. DET.     KOEF. KOREL.
.941353_10-ØØ   .97Ø233_10-ØØ


 6'
VSTOPA  X  16'
IZRACUNANI  F  .73529Ø_10+Ø1
STD.  DEV.  Y  .21Ø766_10-ØØ


KONSTANTA  .8Ø1484_10+Ø2

SPREM.  KOEFICIENT      STD. NAP. KOEF
12'    -.729Ø46_10-ØØ    .97Ø177_10-Ø1
14'    -.816299_10-ØØ    .595Ø88_10-Ø1
15'    -.359982_10+Ø1    .8Ø4569_10-ØØ
16'    -.479317_10-ØØ    .176763_10-ØØ
31'    -.216541_10-Ø1    .129129_10-Ø2
33'     .196071_10-ØØ    .573443_10-Ø1

KOEF. DET.     KOEF. KOREL.
.95556Ø_10-ØØ   .977527_10-ØØ
```

Imamo kar 6 korakov.

Končna regresijska enačba zajame vpliv petih neodvisnih spremenljivk.

$$X_{17}' = 80{,}15 - 0{,}729\ X_{12} - 0{,}02165\ X_{13}{}^2 - 0{,}8163\ X_{14} + 0{,}197\ X_{15}{}^2 - 3{,}6\ X_{15} - 0{,}4793\ X_{16} \quad (11)$$

Nomogram za to zvezo vidimo na sliki 6. Vidimo močan vpliv sprejemljivke X_{14}. Koeficient determinacije je zelo velik 0,96, ker je res skoraj vsa variacija spremenljivke X_{17}' pojasnjena z regresijsko enačbo (11). Takih rezultatov si v praksi seveda zelo želimo, a žal niso zelo pogosti.

Oglejmo si še primer rezultatov, ko spremenljivka izstopa iz regresijske enačbe. To vidimo na sliki 7. Rezultati nimajo več zveze s prejšnjimi podatki, ker pripadajo čisto drugemu primeru. Sedaj nas zanima odvisnost spremenljivke X_8 od 32 neodvisnih spremenljivk $X_1, X_2, \ldots X_{32}$. Iščemo torej $X_8 = f(X_1, X_2, \ldots X_{32})$.

Na sliki 7 vidimo, da vstopajo po vrsti spremenljivke X_{11}, X_{29}, X_4, X_{31}, nato izstopi spremenljivka X_{11}, ker so njen vpliv na odvisno spremenljivko pojasnile ostale spremenljivke, ki so za njo vstopile v regresijsko enačbo. Na koncu vstopi še spremenljivka X_{27}. V 5. koraku vidimo ob izstopu spremenljivke X_{11}, da je izračunani $F = 3{,}7$ — to je manj od vrednosti 4, ki smo jo vzeli v podatkih za F_2. Standardna deviacija odvisne spremenljivke se je v 5. koraku povečala za 0,02, koeficient determinacije pa se je zmanjšal za 0,016. Te spremembe so v našem primeru nepomembne, zato je spremenljivka X_{11} izstopila iz regresijske enačbe.

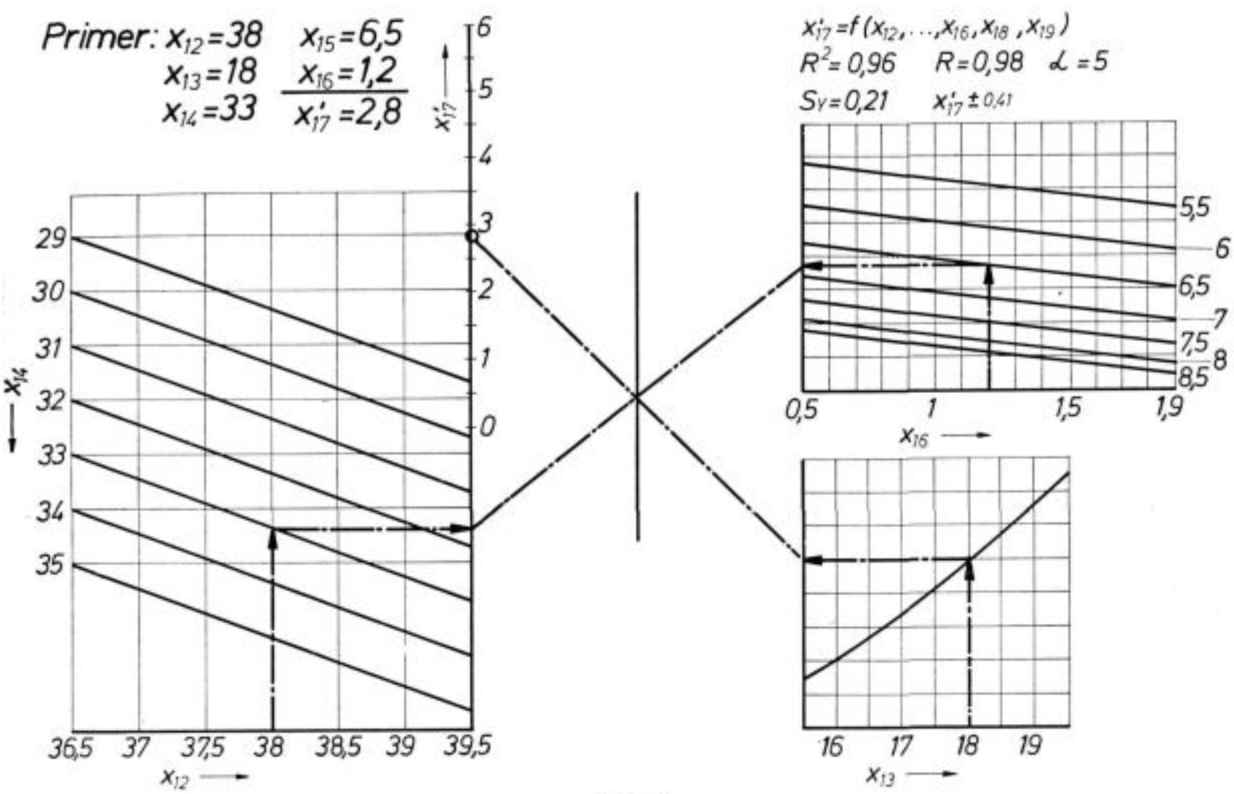
Vidimo tudi, da po izstopu spremenljivke spet lahko vstopi kakšna spremenljivka. Standardna deviacija odvisne spremenljivke se spet zmanjša in koeficient korelacije se spet poveča.

Tabela. $F_{\alpha, 1, \nu}$

ν	$\alpha = 10$	$\alpha = 5$	$\alpha = 1$	$\alpha = 0{,}1$
1'	39.86	161.4	4Ø52	4.Ø53_10+Ø5
2'	8.53	18.51	98.5	998.5
3'	5.54	1Ø.13	34.12	167.Ø
4'	4.55	7.71	21.2Ø	74.14
5'	4.Ø6	6.61	16.26	47.18
6'	3.78	5.99	13.74	35.51
7'	3.59	5.59	12.25	29.25
8'	3.46	5.32	11.26	25.39
9'	3.36	5.12	1Ø.57	22.85
1Ø'	3.29	4.97	1Ø.Ø5	21.Ø4
11'	3.23	4.85	9.65	19.69
12'	3.18	4.75	9.33	18.64
13'	3.14	4.67	9.Ø8	17.82
14'	3.1Ø	4.6Ø	8.87	17.15
15'	3.Ø7	4.55	8.69	16.59
16'	3.Ø5	4.5Ø	8.53	16.12
17'	3.Ø3	4.45	8.4Ø	15.72
18'	3.Ø1	4.42	8.29	15.38
19'	2.99	4.38	8.19	15.Ø8
2Ø'	2.98	4.35	8.1Ø	14.82
21'	2.96	4.33	8.Ø2	14.59
22'	2.95	4.3Ø	7.95	14.38
23'	2.94	4.28	7.88	14.2Ø
24'	2.93	4.26	7.83	14.Ø3
25'	2.92	4.24	7.77	13.88
26'	2.91	4.23	7.72	13.74
27'	2.9Ø	4.21	7.68	13.62
28'	2.92	4.2Ø	7.64	13.5Ø
29'	2.89	4.18	7.6Ø	13.39
3Ø'	2.88	4.17	7.57	13.29
4Ø'	2.84	4.Ø9	7.32	12.61
6Ø'	2.79	4.ØØ	7.Ø8	11.97
12Ø'	2.75	3.92	6.85	11.38
∞	2.71	3.84	6.63	1Ø.83

Program »Koeficienti parcialne korelacije«

Dodatno k programu analize regresije smo pri raziskavah jekel razvili še poseben program z naslovom: »Koeficienti parcialne korelacije«. Razlog za razvoj tega programa je bil v tem, da se pri nekaterih raziskavah nismo več zadovoljili samo s končno ugotovitvijo regresijske enačbe, ampak



Slika 6

Primer monograma za pet pomembnih spremenljivk

smo želeli ugotoviti še medsebojno povezanost spremenljivk. Začetni del tega programa nam sedaj uspešno služi tudi za orientacijsko testiranje medsebojnih odvisnosti, ki nam s preglednim prikazom vseh linearnih medsebojnih zvez lahko precej pripomore k učinkovitejšemu in vsebinsko boljšemu planiranju podrobnejše analize regresije.

Oglejmo si uporabo tega programa na tehnično zelo zanimivem primeru s področja raziskav lastnosti brzoreznih jekel. Tako bo vrednost in praktična uporabnost tega programa predvsem metalurgom mnogo bolj razumljiva.

V obširnem programu raziskav lastnosti brzoreznih jekel smo posebej ugotavljali vpliv velikosti karbidov na mehanske in tehnološke lastnosti tega jekla. Za uvodne raziskave in ugotavljanje medsebojnih odvisnosti kvalitetnih karakteristik v žarjenem, kaljenem in popuščenem stanju smo imeli na voljo za poizkuse 153 palic iz ene in iste šarže z različnimi velikostmi karbidov.

Za vsakega od 153 vzorcev smo določili:

— trdoto (x_1) in velikost karbidov (x_2) v žarjenem stanju,

— oceno preloma (x_3), trdoto (x_4), velikost karbidov (x_5) in velikost avstenitnega zrna (x_6) v kaljenem stanju,

— trdoto (x_7) in popuščno obstojnost (x_8) v popuščenem stanju.

Ker so nas zanimale skoraj vse medsebojne odvisnosti omenjenih spremenljivk, smo morali izvršiti veliko število analiz regresije. Nekaj najpomembnejših ali najbolj značilnih smo prikazali v članku o brzoreznih jeklih[9]. Tak način je zamuden, zaključki s prikazom ugotovljenih odvisnosti pa so zelo nepregledni. Rešitev smo našli z razvojem omenjenega programa, ki nam v svojem prvem delu poda medsebojno povezanost vseh osmih spremenljivk in to vsake z vsako s tabelarično zbranimi koeficienti korelacije v obliki pregledne matrice. To lahko tudi grafično še bolj pregledno prikažemo.

Na sliki 8 je izpis računalnika po tem programu. Pod naslovom »OBIČAJNI KOREL. KOEF.« je matrica koeficientov, ki nam pove, kakšna je zveza med dvema spremenljivkama ob predpostavki, da so druge spremenljivke pri tem nespremenljive. V tej matrici so v prvi vrsti koeficienti vseh spremenljivk z x_1, v drugi vrsti koeficienti vseh spremenljivk z x_2, . . ., v zadnji vrsti pa koeficienti vseh spremenljivk z x_8. Matrica je seveda simetrična, saj je koeficient med x_2 in x_3 isti kakor med x_3 in x_2. Predznak minus pri teh koeficientih pomeni obratno sorazmernost, če pa pred koeficientom ni predznaka, pomeni to premosorazmernost spremenljivk.

Poljubno si lahko v poenostavljeni matrici z zaokroženimi vrednostmi koeficientov označimo pomembnost medsebojnih odvisnosti z jakostnimi rangi, kakor kaže slika 9. Še bolj praktično uporaben in pregleden je grafični prikaz s krogom, v katerem povezave posameznih spremenljivk prikazujejo pomembnost medsebojne povezanosti in premo ali obratno sorazmernost.

Slika 7

Protokol rezultatov analize regresije za primer, ko spremenljivka izstopa iz enačbe regresije

KONTROLNE VSOTE

$.425599999_{10}+02$	$.881999999_{10}+01$	$.475199999_{10}+02$
$.376500000_{10}+02$	$.504000001_{10}+01$	$.123099999_{10}+02$
$.311800000_{10}+05$	$.220219999_{10}+04$	

F1= $.400000_{10}+01$
F2= $.400000_{10}+01$

STD. DEV. Y $.126250_{10}+01$

1'
VSTOPA X 11'
IZRACUNANI F $.792023_{10}+02$
STD. DEV. Y $.715490_{10}-00$

KONSTANTA $.589776_{10}+02$

SPREM.	KOEFICIENT	STD. NAP. KOEF.
11'	$-.487547_{10}-00$	$.547832_{10}-01$

KOEF.DET.	KOEF.KOREL.
$.687506_{10}-00$	$.829160_{10}-00$

2'
VSTOPA X 29'
IZRACUNAI F $.624830_{10}+01$
STD. DEV. Y $.668423_{10}-00$

KONSTANTA $.593256_{10}+02$

SPREM.	KOEFICIENT	STD. NAP. KOEF.
11'	$-.499104_{10}-00$	$.513879_{10}-01$
29'	$-.390413_{10}+03$	$.156186_{10}+03$

KOEF. DET.	KOEF.KOREL.
$.734842_{10}-00$	$.857229_{10}-00$

3'
VSTOPA X 4'
IZRACUNANI F $.191112_{10}+02$
STD. DEV. Y $.542616_{10}-00$

KONSTANTA $.533855_{10}+02$

SPREM.	KOEFCIENT	STD. NAP. KOEF.
4'	$.401806_{10}+01$	$.919119_{10}-00$
11'	$.629746_{10}-00$	$.261569_{10}-00$
29'	$-.890041_{10}+03$	$.170697_{10}+03$

KOEF.DET.	KOEF.KOREL.
$.830255_{10}-00$	$.911183_{10}-00$

4'
VSTOPA X 31'
IZRACUNANI F $.602599_{10}+01$
STD. DEV. Y $.506471_{10}-00$

KONSTANTA $.531332_{10}+02$

SPREM.	KOEFICIENT	STD. NAP. KOEF.
4'	$.323013_{10}+01$	$.915975_{10}-00$
11'	$.495347_{10}-00$	$.251131_{10}-00$
29'	$-.882500_{10}+03$	$.159356_{10}+03$
31'	$.289914_{10}-11$	$.117694_{10}-11$

KOEF. DET.	KOEF. KOREL.
$.856465_{10}-00$	$.925454_{10}-00$

5'
IZSTOPA X 11'
IZRACUNANI F $.373509_{10}+01$
STD. DEV. Y $.526448_{10}-00$

KONSTANTA $.554663_{10}+02$

SPREM.	KOEFICENT	STD. NAP. KOEF.
4'	$.149358_{10}+01$	$.184894_{10}-00$
29'	$-.686436_{10}+03$	$.127739_{10}+03$
31'	$.342193_{10}-11$	$.118933_{10}-11$

KOEF.DET.	KOEF.KOREL.
$.840219_{10}-00$	$.916635_{10}-00$

```
6'
VSTOPA  X  27'
IZRACUNANI  F  .482964_10+01
STD. DEV.   Y  .499093_10-00


KONSTANTA  .543326_10+02

SPREM.KOEFICIENT     STD.NAP. KOEF.
 4'    .255194_10+01  .512496_10-00
27'    .518102_10-01  .235753_10-01
29'   -.905040_10+03  .156717_10+03
31'    .299773_10-11  .114392_10-11

KOEF. DET.    KOEF. KOREL.
.860618_10-00  .927695_10-00
```

Oglejmo si le del strokovno tehničnega komentarja k sliki 9:

Iz razpoložljivih podatkov smo lahko ugotovili in kvantitativno potrdili strokovno razumljivo domnevo, da je med velikostjo karbidov v žarjenem in kaljenem stanju, zrnatostjo preloma po Shepherdovi metodi in metalografsko oceno velikosti avstenitnega zrna po metodi Snyder-Graffa zelo ozka medsebojna zveza (koeficienti korelacije so vsi večji od 0,8 — sl. 9).

Regresijska odvisnost $x_5 = f(x_2)$ je podana z visokim koeficientom korelacije $R = 0{,}911$. Koeficient determinacije $R^2 = 0{,}83$ v nekoliko poenostavljeni obliki pomeni, da je 83 % variacij spremenljivke x_5 (velikost karbidov v kaljenem stanju) pojasnjenih z variacijami spremenljivke x_2 (velikost karbidov v žarjenem stanju), preostalih 17 % variacij pa ostaja ob tej regresiji še nepojasnjenih ali praktično odvisnih od drugih vplivov.

Ker je poznano, da so grobi karbidi v kaljenem brzoreznem jeklu lahko posledica različnih vplivov (npr. pregretje jekla pri kaljenju), je omenjena ugotovitev potrjena kvantitativno s koeficientom zelo pomembna in nam omogoča neposredno tehnološki zaključek z identifikacijo nastanka grobih karbidov. V našem primeru lahko z gotovostjo trdimo, da karbidi, ki smo jih opazili v kaljenem stanju niso nastali med kaljenjem, ampak so bili v mikrostrukturi jekla v skoraj enaki obliki prisotni že prej — v žarjenem stanju. Za nastanek takih karbidov med žarjenjem ni potrebnih pogojev, zato na osnovi take regresije lahko z gotovostjo trdimo, da so grobi karbidi v našem primeru nastali pred in med vročo predelavo jekla.

Na tak način lahko komentiramo še ostale medsebojne odvisnosti in lahko se odločimo, katere odvisnosti bomo smatrali v primerjavi z drugimi za nepomembne: Takih sploh ne prikazujemo! Tak prikaz pa nam lahko pomaga tudi kot pomoč pri začetni selekciji oziroma pri odločitvi, katere odvisnosti in katere kombinacije so res interesantne za podroben izračun enačbe analize regresije in za prikaz z monogramom. Lahko si prihranimo mnogo podrobnih izračunov nepomembnih in neinteresantnih regresij.

Izračun matrice običajnih koeficientov korelacije lahko torej smatramo za zelo priporočljiv uvodni korak pri analizah množične regresije.

Poglejmo na istem primeru še pomen koeficientov parcialne korelacije. Ti koeficienti, zbrani v podobni matrici, nam povedo, kakšna je zveza med dvema spremenljivkama ne glede na to, kakšne vrednosti imajo pri tem ostale spremenljivke. V računskem postopku se namreč spreminjanje ostalih spremenljivk izloči tako, da se vzame za parcialni koeficient korelacije med dvema spremenljivkama povprečje običajnih koeficientov korelacije pri različnih konstantnih vrednostih ostalih spremenljivk. Vrstni red v matrici teh koeficientov je zopet isti, le predznaki koeficientov imajo drugačen pomen. Pri parcialnih koeficientih pomeni predznak minus premo sorazmernost, medtem ko pri običajnih pomeni obratno sorazmernost.

Če označimo koeficient parcialne korelacije med x_i in x_j s p_{ij}, potem ima matrica vseh koeficientov obliko:

$$\begin{array}{ll}
p_{11} & p_{12} \cdots\cdots p_{18} \\
p_{21} & p_{22} \cdots\cdots p_{28} \\
\cdot & \cdot \qquad\quad \cdot \\
\cdot & \cdot \qquad\quad \cdot \\
\cdot & \cdot \qquad\quad \cdot \\
p_{81} & p_{82} \cdots\cdots p_{88}
\end{array}$$

kjer so

$$\begin{array}{ll}
p_{ii} = 1; & i = 1, 2 \ldots . 8 \\
p_{ij} = p_{ji}; & i, j = 1, 2 \ldots . 8 \text{ in} \\
|p_{ij}| \leq 1 &
\end{array}$$

Za primer vzemimo koeficient parcialne korelacije med spremenljivkama x_4 in x_7, to je $p_{47} = -0{,}9957$. To je vrednost zelo blizu 1. Zato sklepamo, da sta spremenljivki x_4 in x_7 tesno povezani med seboj. Tudi za parcialne koeficiente si za lepši pregled narišemo enako sliko 10 in na črte, ki povezujejo spremenljivke, lahko napišemo vrednosti koeficientov. Tiste z višjimi vrednostmi koeficientov narišemo debelejše.

Med rezultati na protokolu slike 8 so tudi koeficienti regresijske enačbe, če vzamemo za odvisno spremenljivko x_8 in za neodvisne $x_1, x_2 \ldots . x_7$. Pri programu »Koeficienti parcialne korelacije« računalnik vedno izračuna koeficiente za vse neodvisne spremenljivke, potem pa določi pomembnost regresijske enačbe z F testom. Zato izpiše prostostni stopnji N_1 in N_2 ter vrednost F (glej sliko 8).

Slika 8

Protokol rezultatov analize regresije po programu, ki izračuna običajne in parcialne koeficiente korelacije

KONTROLNE VSOTE

.40696000$_{10}$+05	.10510000$_{10}$+04	.13012500$_{10}$+04	.975964999$_{10}$+04	.10659000$_{10}$+04
.18260000$_{10}$+04	.987269993$_{10}$+04	.11451000$_{10}$+03		

OBICAJNI KOREL. KOEF.

.1000$_{10}$+01	-.7212$_{10}$-01	-.1016$_{10}$-00	-.2424$_{10}$-01	-.4310$_{10}$-01	-.8476$_{10}$-01	.1276$_{10}$-00	.1641$_{10}$-00
-.7212$_{10}$-01	.1000$_{10}$+01	.9151$_{10}$-00	.6569$_{10}$-00	.9111$_{10}$-00	.8415$_{10}$-00	.4062$_{10}$-00	-.1780$_{10}$-00
-.1016$_{10}$-00	.9151$_{10}$-00	.1000$_{10}$+01	.6638$_{10}$-00	.9320$_{10}$-00	.9201$_{10}$-00	.3788$_{10}$-00	-.2092$_{10}$-00
-.2424$_{10}$-01	.6569$_{10}$-00	.6638$_{10}$-00	.1000$_{10}$+01	.6127$_{10}$-00	.6582$_{10}$-00	.3947$_{10}$-00	-.4775$_{10}$-00
-.4310$_{10}$-01	.9111$_{10}$-00	.9320$_{10}$-00	.6127$_{10}$-00	.1000$_{10}$+01	.8723$_{10}$-00	.3707$_{10}$-00	-.1725$_{10}$-00
-.8476$_{10}$-01	.8415$_{10}$-00	.9201$_{10}$-00	.6582$_{10}$-00	.8723$_{10}$-00	.1000$_{10}$+01	.4078$_{10}$-00	-.1767$_{10}$-00
.1276$_{10}$-00	.4062$_{10}$-00	.3788$_{10}$-00	.3947$_{10}$-00	.3707$_{10}$-00	.4078$_{10}$-00	.1000$_{10}$+01	.6172$_{10}$-00
.1641$_{10}$-00	-.1780$_{10}$-00	-.2092$_{10}$-00	-.4775$_{10}$-00	-.1725$_{10}$-00	-.1767$_{10}$-00	.6172$_{10}$-00	.1000$_{10}$+01

KONSTANTA -.3516$_{10}$-00

SPREM.	KOEFICIENT	STD. NAP. KOEF.	T
1'	.5151$_{10}$-03	.9828$_{10}$-04	.5241$_{10}$+01
2'	-.4501$_{10}$-02	.4540$_{10}$-02	.9914$_{10}$-00
3'	.5217$_{10}$-02	.7918$_{10}$-02	.6588$_{10}$-00
4'	-.9922$_{10}$-00	.6622$_{10}$-02	.1498$_{10}$+03
5'	.2505$_{10}$-03	.4770$_{10}$-02	.5251$_{10}$-01
6'	-.7870$_{10}$-03	.1877$_{10}$-02	.4191$_{10}$-00
7'	.9957$_{10}$-00	.4828$_{10}$-02	.2062$_{10}$+03

N1	N2	F
7'	.1450$_{10}$+03	.8721$_{10}$+04

KOEFICIENTI PARCIALNE KORELACIJE

.1000$_{10}$+01	.4033$_{10}$-02	.1300$_{10}$-00	-.4004$_{10}$-00	-.1292$_{10}$-00	.1217$_{10}$-01	.3892$_{10}$-00	-.3991$_{10}$-00
.4033$_{10}$-02	.1000$_{10}$+01	-.3649$_{10}$-00	.6775$_{10}$-01	-.4116$_{10}$-00	.1023$_{10}$-00	-.8967$_{10}$-01	.8205$_{10}$-01
.1300$_{10}$-00	-.3649$_{10}$-00	.1000$_{10}$+01	-.6265$_{10}$-01	-.4263$_{10}$-00	-.5475$_{10}$-00	.5871$_{10}$-01	-.5463$_{10}$-01
-.4004$_{10}$-00	.6775$_{10}$-01	-.6265$_{10}$-01	.1000$_{10}$+01	.5114$_{10}$-02	.2250$_{10}$-01	-.9957$_{10}$-00	.9967$_{10}$-00
-.1292$_{10}$-00	-.4116$_{10}$-00	-.4263$_{10}$-00	.5114$_{10}$-02	.1000$_{10}$+01	-.1356$_{10}$-00	.6493$_{10}$-02	-.4361$_{10}$-02
.1217$_{10}$-01	.1023$_{10}$-00	-.5475$_{10}$-00	.2250$_{10}$-01	-.1356$_{10}$-00	.1000$_{10}$+01	-.4307$_{10}$-01	.3478$_{10}$-01
.3892$_{10}$-00	-.8967$_{10}$-01	.5871$_{10}$-01	-.9957$_{10}$-00	.6493$_{10}$-02	-.4307$_{10}$-01	.1000$_{10}$ 01	-.9982$_{10}$-00
-.3991$_{10}$-00	.8205$_{10}$-01	-.5463$_{10}$-01	.9967$_{10}$-00	-.4361$_{10}$-02	.3478$_{10}$-01	-.9982$_{10}$-00	.1000$_{10}$+01

TOTALNA VARIANCA	NEPOJASNJENA VARIANCA
4105$_{10}$+02	.9727$_{10}$-01

KOEF. MULT. KOREL.	DET. KOEF.
.9988$_{10}$-00	.9976$_{10}$-00

Žarjeno stanje

x_1 - trdota HB

x_2 - index velikosti karbidov I_k

Kaljeno stanje

x_3 - ocena preloma F po Shepherdu

x_4 - trdota HRC

x_5 - index velikosti karbidov I_k

x_6 - velikost austenitnega zrna SG po Snyder-Graffu

Popuščeno stanje

x_7 - trdota HRC

x_8 - popuščna obstojnost D_{HRC}

$D_{HRC} = HRC_{popuščeno} - HRC_{kaljeno}$

Koeficienti korelacije:

$r \geqq 0,80$ ———

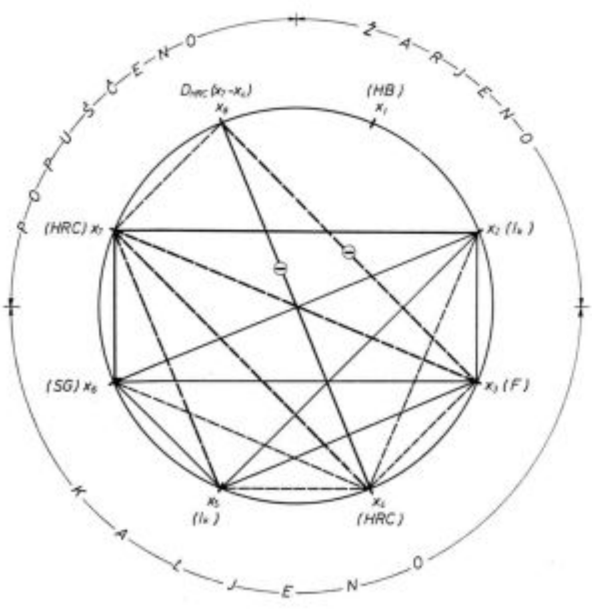
$r = 0,79 - 0,60$ – – – –

$r = 0,59 - 0,40$ ———

$r = 0,39 - 0,20$ – – – –

$r < 0,20$ (grafično ni prikazano)

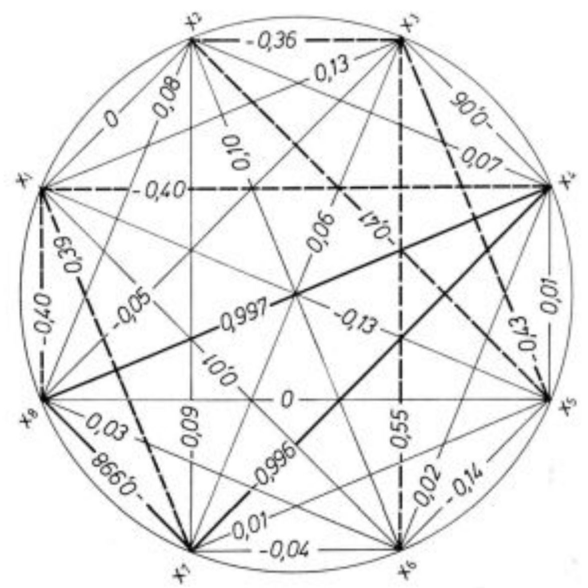
x_1	1							
x_2	-0,072	1						
x_3	-0,102	0,915	1					
x_4	-0,024	0,657	0,664	1				
x_5	-0,043	0,911	0,932	0,613	1			
x_6	-0,085	0,842	0,920	0,658	0,872	1		
x_7	0,128	0,406	0,379	0,395	0,371	0,408	1	
x_8	0,164	-0,178	-0,209	-0,477	-0,112	-0,177	0,617	1
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8



Slika 9

Običajni koeficienti korelacije

V našem primeru je F = 8721. To je velika vrednost in regresijska enačba je gotovo 99,99 % pomembna.



Slika 10

Koeficienti parcialne korelacije med spremenljivkami $x_1 \ldots\ldots\ldots x_8$

Za koeficienti parcialne korelacije dobimo tudi vrednosti totalne in nepojasnjene variance, nato pa koeficient množične korelacije in koeficient determinacije.

Literatura

1. Rodič J.: »Metode matematične statistike«, Železarski zbornik, 1967, I., št. 2., str. 137—154
2. Rodič J.: »Sistemi kontrole in metodika reševanja tehnoloških problemov«, Železarski zbornik, 1968., II., št. 3., str. 153—163
3. Rode B.: »Analiza statistične porazdelitve na elektronskem računalniku«, Železarski zbornik, 1967., I., št. 3, str. 189—203
4. Rode B., J. Rodič: »Statistično planiranje in vrednotenje metalurških raziskav — Analize variance s programi na računalniku ZUSE Z-23«, Železarski zbornik, 1968, II., št. 2., str. 99—111
5. Rode B.: »Statistično planiranje in vrednotenje metalurških raziskav — Latinski kvadrat s programom na elektronskem računalniku«, Železarski zbornik, 1969, III., št. 2, str.
6. Zanella A.: Programi di calcolo automatico nel controllo della qualità e nella programmazione degli esperimenti, AICQM, maj 1962, str. 11—26
7. Miller S. A., R. J. Taylor: Pegasus Program Multiple Regression Analysis, BISRA, List 143, Oktober November 1965
8. Ralston A., H. S. Wilf: Mathematical Methods for Digital Computers, John Wiley and Sons, str. 191—203.
9. Rodič J., A. Rodič: Brzorezna jekla III. del, Železarski zbornik, 1968, II. št. 3, str. 167—170

ZUSAMMENFASSUNG

Im Artikel sind zwei Programme aus dem Bereich der Regresionsanalyse beschrieben. Der erste Programm ist nach der »Schrittweise« Methode ausgefertigt und bestimmt nur die wichtigsten Variablen, welche eine hinter der anderen in die Regresionsgleichung eintretten.

Der Programm ist an den Beispielen aus der Elektronenrechenmaschine Z-23 gedeutet.

Es sind auch die Nomograme dieser Beispiele gezeigt, an welchen die Rezultate graphisch dargestellt sind. Solche Nomograme haben sich bei der Vorschungsarbeit im Hüttenwerk Ravne schon eingebürgert.

Der zweite Programm rechnet die Üblichen und die partialen Koeffiziente der Korrelation.

Auch dieser Programm ist an einem praktišchen Beispiel aus dem Hüttenwerk Ravne gedeutet.

SUMMARY

In the paper two computer programs are described from the field of regression analysis. The first program is made by the stepwise procedure and it determines only the basic variables which one after the other enter the regression equation.

The program is illustrated with examples of results obtained from computer ZUSE Z-23.

Also nomograms of these examples are given which graphically interprete the results. Such nomograms are very much used in the research work at the Iron and Steel Works Ravne.

The second program calculates simple and parcial correlation coefficients.

Also this program is illustrated with a practical example from Iron and Steel Works Ravne.

ЗАКЛЮЧЕНИЕ

В статье описаны два программа из отрасли анализа регрессии. Первая программа сделана методом »Шаг за шагом«; она определяет только важные переменные данныя которые один за другим вступают в регрессионное уравнение. Программа объяснена на целом ряде результатов полученых при помощи электроннога счётчика типа Zuse Z-23.

Приведены также номограммы тех примеров которые дают результаты в форме графика. Такие номограммы уже вошли в обшее употребление при работах исследования металлургическога завода Равне. Другая программа вычисляет обыкновенные и парциальные коэфициенты взаимнога соответствия.

Также и эта программа объяснена при помощи практическога примера металлургическога завода Равне.