RASPRAVE, 18, 1983, str. 41—52

UDK 551.509.314
551.509.323
Originalni znanstveni rad

Statistička metoda prognoze minimalne temperature

A Statistical Method for Forecasting Minimum Temperature

GORAN BELAMARIĆ i MIHOVIL KISEGI
Republički hidrometeorološki zavod SR Hrvatske, Zagreb

Sažetak: Za dobivanje objektivne prognoze minimalne temperature u mjesecu siječnju za Zagreb korištena je jednadžba mnogostruke linearne regresije. Odabrano je deset potencijalnih prediktora, mjerenih na opservatoriju Zagreb-Maksimir. Određeno je ukupno osam jednadžbi regresije, koristeći kombinacije od tri do pet, po mogućnosti međusobno nezavisnih, prediktora. Sve jednadžbe određene su iz petogodišnjeg niza 1975-79. Ocjena jednadžbi učinjena je na podacima siječnja 1982. Za kombinaciju od pet međusobno nezavisnih prediktora koji su dali najveći doprinos redukciji varijance prediktanda, određena je jednadžba iz desetogodišnjeg niza podataka 1972-81, čime su dobiveni stabilniji koeficijenti regresije.

Ključne riječi: Minimalna temperatura, statistička prognoza temperature, mnogostruka linearna regresija.

Abstract: This study employs the equation of multiple linear regression with ten potential predictors of minimum temperature in January at the observatory Zagreb-Maksimir. A total of eight equations of regression are used as a combination of three to five possibly mutually independent predictors. All equations are determined from five years data series 1975-79. The test of the method is conducted on data taken in January 1982. For the combination of five mutually independent predictors, giving the highest contribution to the variance of the predictand the equation from a 10-year (1972-81) data series is constructed, which resulted in more stable coefficients of regression.

Key words: Minimum temperature, statistical temperature forecasting, multiple linear regression.

1. UVOD

Cilj rada bio je dobiti objektivnu prognozu minimalne temperature u mjesecu siječnju za zagrebačko područje. Prognoza minimalnih temperatura je od velikog značenja, posebno zimi, kada je i naročito teška. Teškoće prognoze za šire područje su i u tome, što minimalna temperatura jako ovisi o lokalnim uvjetima — konfiguraciji terena, karakteristikama tla te mogućem snježnom pokrivaču.

Do pojave numeričke prognoze vremena korištena je »klasična metoda«, kod koje se na osnovi opaženih vrijednosti meteoroloških elemenata do trenutka izdavanja prognoze formulira prognoza određene meteorološke veličine.

Numerička prognoza vremena nam ne daje izravno prognozu minimalne temperature. Već i sam termin »numerička prognoza vremena« nije ispravan, jer se ne prognozira vrijeme, već polje tlaka, geopotencijala itd. Tek savršeniji modeli numeričke prognoze daju i polja temperature, horizontalnih i vertikalnih brzina, vlažnosti, oborina itd. Stoga prognoza većine meteoroloških elemenata, koji su značajni za čovjeka i njegove aktivnosti, ostaje »na dušu« meteorologa-sinoptičara, koji interpretira prognostičke karte služeći se svojim znanjem, iskustvom, praćenjem vremena i poznavanjem klimatskih karakteristika krajeva za koje daje prognozu. Ta interpretacija je međutim subjektivna, pa se stoga prišlo razvijanju objektivnih metoda prognoze vremena.

Pregled tih metoda dali su Glahn (1965), Barry i Perry (1973) i Belov (1975).

U Republičkom hidrometeorološkom zavodu SR Hrvatske koristi se isključivo »PERFECT PROGNOSTIC METHOD«. Do sada određene su operativne jednadžbe za prognozu kondicionalne vjerojatnosti snježnih oborina u zimskom dijelu godine (razdoblje od IX do IV mjeseca) za osam mjesta u zapadnom području unutrašnjosti Hrvatske (Kisegi, 1971). Tom prilikom korišten je i logit model (Brelsford i Jones, 1967; Jones, 1968).

U ovom radu za prognozu minimalne temperature koristili smo metodu mnogostruke linearne regresije. Regresija naravno ne mora biti linearna, ali se pokazalo da uključivanje nelinearnih članova u jednadžbu regresije ne povećava bitno točnost prognoze. Po prvi put primijenili su ju za prognozu minimalne i maksimalne temperature, za 131 stanicu na području SAD uz korištenje četiri prediktora, Klein i Lewis (1970). Daljnja poboljšanja, uvođenjem novih prediktora učinili su Klein i Marshall (1973), Hammons, Dallavalle i Klein 1976) te Carter i ost. (1978).

2. OBJEKTIVNA PROGNOZA VREMENA

2.1. **Statističke veze**

Zadatak statističke prognoze je, pomoću modela vjerojatnosti, iz prošlog i trenutnog stanja atmosfere prognozirati buduće stanje. Karakteristike prošlog, odnosno trenutnog stanja zovu se prediktori, a karakteristike budućeg prediktandi.

Model vjerojatnosti odražava činjenicu da prediktori i prediktand (ili više njih) nisu funkcijski (dakle jednoznačno), već stohastički povezani. Kod stohastičke veze promjena vrijednosti jedne varijable (prediktora) utječe na promjenu razdiobe druge varijable (prediktanda), odnosno jednoj vrijednosti prediktora pripada više vrijednosti prediktanda, ali su te vrijednosti raspodijeljene po nekom zakonu vjerojatnosti.

Stohastičke veze odgovaraju cjelokupnim kolektivima vrijednosti meteoroloških elemenata (populaciji). Čitava populacija obično nije dostupna, pa se zadovoljavamo jednim njezinim dijelom, uzorkom. Tada govorimo o statističkim vezama. Jasno je da povećanjem broja članova uzorka statističke veze teže stohastičkim.

Uzorak bi morao biti slučajan, tj. njegove članove trebali bi nasumce odabrati iz populacije. Češći je slučaj da raspolažemo vremenskim nizom vrijednosti meteoroloških elemenata, dakle uzorkom koji nije, strogo uzevši, slučajan. Važno je ipak da su karakteristike tog vremenskog niza stacionirane, tj. da se ne mijenjaju u vremenu (dakle da ne postoji trend, periodičnost i slično). Periodičnost, koja dakako kod većine meteoroloških elemenata postoji, može se eliminirati ako razmatramo otklone elemenata od normi, tj. od njihovih srednjih vrijednosti za pojedini dan u godini (ili sat u danu), koje pokazuju godišnji, odnosno dnevni hod.

2.2. **Dinamičko-statističke metode**

Većina numeričkih modela prognoze vremena ne daje neposredno i prognozu pojedinih meteoroloških elemenata, ali oni sadrže varijable koje su u prilično dobroj povezanosti s pojedinim meteorološkim elementom.

Interpretacija prognostičkih karata do 1970. godine oslanjala se uglavnom na iskustvo, koje se stjecalo proučavanjem klimatoloških podataka, uključujući praćene vremenskih prilika, sačuvanih podataka iz prethodnih prognostičkih karata i kompjuterskih analiza. Sve te podatke možemo nazvati »historijski materijal« Kada se taj materijal analizira u interpretacijske svrhe i uskladišti kao »banka iskustva«, dobijemo ono što nazivamo »interpretacijska klimatologija«. Veličina niza historijskog materijala mijenja se u ovisnosti o slučaju. Važno je da su svi tipovi vremenskih situacija predstavljeni u materijalu, uključujući i ekstremne slučajeve. Mora se provesti i neka vrsta verifikacije, možda na štetu dijela historijskog materijala. Pored tog problema, prije odabiranja same metode interpretiranja i prognoze, treba riješiti još niz pitanja, o čemu je detaljnu analizu dao Lönnqvist (1978).

Sve do nedavno metode objektivne prognoze vremena mogli smo svrstati u jednu od dvije kategorije — dinamičke i statističke. U novije vrijeme istražuje se relativno novo područje stohastičko-dinamičke prognoze, a započeli su ga svojim radovima Epstein (1969) i Fleming (1971). Premda će se dinamičko-stohastička prognoza i dalje razvijati korištenjem sve jačih elektroničkih računala, mi ćemo u praksi morati koristiti neku kombinaciju dinamičkih i statističkih metoda. Poslije 1970. godine razvile su se dvije takve metode.

Prva statistička metoda, koju su upotrebili Klein i ost. (1959), općenito je nazvana »PERFECT PROGNOSTIC METHOD« (»metoda savršene prognoze«). Kod ove metode traži se jedan istovremeni statistički odnos između varijable koju želimo procijeniti-prediktanda i niza izabranih varijabli-prediktori, koje mogu biti prognozirane dinamičkim modelom ili modelima. Obje varijable, predktand i prediktor, su opažene veličine uzete iz jednog određenog uzorka, koji nikako nije slučajan. Koristi se dakle »historijski materijal«. Dobivene statističke veze primjenjuju se zatim na izlazne vrijednosti prediktora iz numeričkog modela, i dobiva se prognoza prediktanda za isti prognostički termin, što ne mora uvijek biti, jer među njima može postojati korak u vremenu od 24 sata pa i više.

Druga metoda, koju zovemo »MODEL OUTPUT STATISTICS«, ili MOS, sastoji se u određivanju veze između prediktanda i prediktora dobivenih iz numeričkog modela, ali u istom terminu za koji su prognozirani prediktori. Primjena se zatim provodi na točno isti način kao kod metode savršene prognoze. Danas se smatra da je MOS najpraktičniji način prilaženja objektivnoj prognozi većine elemenata vremena. Tako se danas već rade prognoze vjerojatnosti oborina, kondicionalne vjerojatnosti snježnih oborina, količine oborina, vjetra pri tlu, maksimalne i minimalne temperature, vertikalne i horizontalne vidljivosti, količine naoblake te grmljavinske oluje i snažne lokalne oluje (Klein, 1978).

Budući da mi još ne raspolažemo s arhiviranim dužim nizom vrijednosti varijabli dobivenih numeričkim modelom, prisiljeni smo za sada koristiti metodu savršene prognoze. Ukoliko su prognostičke karte točne, ona ima i svojih prednosti. Dobra strana te metode je da je interpretacija to točnija što je bolja prognostička karta. Druga je prednost da se statistički parametri koji se interpretiraju iz modela mogu izvesti jednom zauvijek, i da se dobivaju boji rezultati svaki put kada se numerički model poboljša. Glavni nedostatak je što treba biti oprezan u izboru prediktora — ne mogu se koristiti oni koji se ne mogu dobro prognozirati numeričkim modelom.

2.3. Korelacija i regresija

2.3.1. Linearna korelacija

Da bi danu varijablu-prediktand doveli u vezu s bilo kojom od izbranih varijabli-prediktori, uobičajeno je računanje jednostavnog linearnog koeficijenta korelacije, koji je za uzorak dan izrazom

$$r(x,y) = \frac{(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (2.1)$$

gdje su x_i i y_i i-te vrijednosti dvije varijable koje koreliramo, $\bar{x}$ i $\bar{y}$ su njihovi srednjaci dobiveni iz uzorka, sumacija se provodi preko n slučajeva. Ukoliko se prediktand s jedne točke korelira s vrijednostima prediktora uzetim s mreže točaka, dobivaju se korelaciona polja.

2.3.2. Djelomična korelacija

Može se izvesti da je za slučaj dvije nezavisne varijable koeficijent djelomične korelacije između y i x_1 (x_2 isključen).

$$r(y,x_1;x_2) = \frac{r(y,x_1) - r(y,x_2)\,r(x_1\,x_2)}{\sqrt{[1 - r^2(y,x_2)]\;[1 - r^2(x_1,\,x_2)]}} \qquad (2.2)$$

Za tri nezavisne varijable:

$$r(y,x_i;x_j,x_k) = \frac{r(y,x_i;x_j) - r(y,x_k;x_j)\,r(x_i,x_k;x_j)}{\sqrt{[1 - r^2(y,x_k;x_j)]\;[1 - r^2(x_i,x_k;x_j)]}} \qquad (2.3)$$

$i, j, k = 1,2,3;\ i \neq j \neq k.$

Tu je dakle isključen utjecaj varijabli x_j, x_k na korelaciju među y i x_i.

Ako su nezavisne varijable nekorelirane $r(x_i,x_k;x_j) = 0$, pa je $r(y,x_i;x_j,x_k) = r(y,x_i;x_j) = r(y,x_i)$.

Za više nezavisnih varijabli koeficijente djelomične korelacije dobivamo iterativnim postupkom, isključujući utjecaj jedne po jedne varijable računajući (2.2), uvrštavajući u (2.3) itd.

2.3.3. Linearna regresija

Koeficijent maksimalne korelacije (bez obzira na predznak), u korelacionom polju, u točki mreže, ili na meteorološkoj stanici, dobiven između prediktanda i izabranog prediktora, pokazuje koliko dobro dana varijabla može biti prognozirana jednostavnom jednadžbom linearne regresije, oblika

$$\hat{y} = b_0 + b_1 x \qquad (2.4)$$

gdje su b_0 i b_1 koeficijenti regresije, $\hat{y}$ prognostička vrijednost zavisne varijable ili prediktanda, a x je prediktor ili nezavisna varijabla. Vrijednosti konstanti b_0 i b_1 određuju se metodom najmanjih kvadrata, koja minimalizira sumu kvadrata razlika između prognostičkih vrijednosti $\hat{y}$ i opaženih vrijednosti y, tj.

$$\sum_{i=1}^{n} (y_i - \hat{y})^2 = \min$$

Koeficijent b_1 je nagib pravca regresije i dan je izrazom

$$b_1 = r(x,y)\,\frac{s(y)}{s(x)}, \qquad (2.5)$$

dok se b_0 dobije izrazom.

$$b_0 = \bar{y} - b_1\bar{x} = \bar{y} - r(x,y)\,\frac{s(y)}{s(x)}\,.\,\bar{x} \qquad (2.6)$$

s(x) i s(y) su standardne devijacije, čija neiskrivljena procjena je dana izrazom

$$= (x)s\ \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}} \qquad (2.7)$$

Kvadrat standardne devijacije zove se varijanca (V).

2.3.4. Mnogostruka linearna regresija

Poboljšani rezultati mogu se dobiti korištenjem dodatnih prediktora u jednadžbi mnogostruke linearne regresije ovog oblika

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k \qquad (2.8)$$

gdje su b_0 i b_i koeficijenti regresije određeni metodom najmanjih kvadrata, a $x_1, x_2 \ldots x_k$ su k različitih prediktora.

Koeficijente jednadžbe možemo procjenjivati kao kod jednostavne linearne regresije samo ako nezavisne varijable x_i, $i = 1,2 \ldots k$ nisu i same međusobno korelirane. Ako jesu, koristimo se koeficijentima djelomične (parcijalne) korelacije. Ti koeficijenti pokazuju korelaciju zavisne i jedne od nezavisnih varijabli, ako ostale nezavisne varijable imaju neku stalnu vrijednost, čime se isključuje njihov utjecaj. Procjenjivanje koeficijenata može se napraviti preko matrica; vektor koeficijenata (B) može se izračunati kao produkt matrica:

$$B = (X'X)^{-1} \cdot X'Y = C \cdot X'Y \qquad (2.9)$$

gdje su

Y — vektor vrijednosti zavisne varijable
X — matrica vrijednosti nezavisnih varijabli
' — je oznaka za transponiranu matricu.

Zgodno je uvesti matricu korelacije, koja je kvadratna i simetrična, te raditi s njenom determinantom. Determinanta kompletne matrice koeficijenata jednostavne korelacije (D) uključuje koeficijente između prediktanda i svakog pojedinog prediktora (u prvom stupcu i prvom retku) i koeficijente korelacije između raznih prediktora (u ostatku matrice). Svi članovi ove determinante su u intervalu [0,1], za razliku od matrice X'X, kojoj članovi mogu jako varirati. Može se pokazati da se koeficijenti mogu dobiti kao:

$$b_j = \frac{s(y)}{s(x_j)} \; \frac{D(y,x_j)}{D(x,x)} \qquad (2.10)$$

$D(x,x)$ i $D(y,x_j)$ su minori determinante D; $D(x,x)$ dobijemo tako da u D precrtamo prvi redak i prvi stupac, a $((y,x_j)$ ako precrtamo prvi redak i j-ti stupac (ili prvi stupac i j-ti redak, što je svejedno, jer je D simetrična).

Koeficijenti se mogu dobiti i pomoću izraza:

$$\begin{vmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_k \end{vmatrix} = s(y) \cdot D^{-1}(x,x) \begin{vmatrix} \frac{r(y,x_1)}{s(x_1)} \\ \frac{r(y,x_2)}{s(x_2)} \\ \cdot \\ \cdot \\ \frac{r(y,x_k)}{s(x_k)} \end{vmatrix} \qquad (2.11)$$

gdje je $D^{-1}(x,x)$ determinanta matrice inverzne matrici $D(x,x)$. Primijetimo da, radimo li izrazima (2.10) ili (2.11), koeficijent b_0 se ne može odrediti. Djelujemo li s operatorom osrednjavanja E( ) na jednadžbu mnogostruke linearne regresije (2.8) dobivamo:

$$E(y) = b_0 + b_1 E(x_1) + b_2 E(x_2) + \ldots + b_k E(x_k) \qquad (2.12)$$

no $E(y) = \bar{y}$, $E(x_i) = \bar{x}_i$, pa je

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \ldots - b_k \bar{x}_k \qquad (2.13)$$

Uvrstimo li taj izraz u (2.8) dobiva se:

$$(y - \bar{y}) = b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2) + \ldots + b_k(x_k - \bar{x}_k) \qquad (2.14)$$

Umjesto s početnim varijablama $x_1 x_2, \ldots x_k$, radimo s njihovim odstupanjima od srednjaka, što je čak i praktičnije. Koeficijenti dobiveni pomoću (2.10) ili (2.11) odnose se i na jednadžbu (2.8), u kojoj bi još trebali procijeniti parametar b_0 (2.13).

2.3.5. Koeficijent mnogostruke korelacije

Ovaj koeficijent pokazuje vezu između zavisne varijable i nezavisnih varijabli, a može se izračunati pomoću jednadžbe:

$$R = \sqrt{1 - \frac{D}{D(x,x)}} = \sqrt{1 - \frac{(SGP)^2}{[s(y)]^2}} \qquad (2.15)$$

SGP je standardna greška prognoza dobivenih jednadžbom regresije, a $s(y)$ je standardna devijacija prediktanda.

Iz jednadžbe (2.15) slijedi da je postotak totalne varijance (EV) prediktanda oko srednjaka y, objašnjen jednadžbom mnogostruke linearne regresije, jednak kvadratu koeficijenta mnogostruke korelacije, ili:

$$RV = \frac{EV}{100} = R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (2.16)$$

gdje je y_i opažena vrijednost prediktanda, $\hat{y}_i$ prognozirana vrijednost jednadžbom regresije (2.8), a $\bar{y}$ je srednjak zavisne varijable. Izraz je poznat kao redukcija varijance (RV) i on je mjera valjanosti jednadžbe za procjenu y.

Iz gornje jednadžbe je jasno da smanjenje sume kvadrata grešaka procjene je istog značenja kao i povećanje redukcije varijance, i smanjenje SGP greške (ili standardne greške procjene), gdje je

$$SGP = \left[\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \right]^{1/2} \qquad (2.17)$$

Ako su koeficijenti korelacije $r(y,x_i) = 0$ ($i = 1, 2, \ldots k$) i $r(x_i,x_j) = 0$ ($i, j, = 1,2, \ldots k$, za $i \neq j$), onda je

$D = 1$ i $D(x,x) = 1$, pa je $R = 0$.

Ako su varijable funkcijski povezane, $r = \pm 1$, pa je i $R = 1$. Za $k > 1$, R je veći od bilo kojeg koeficijenta korelacije, osim onih na dijagonalnoj liniji u D[gdje je $r(x,x) = r(x_1, x_1) = \ldots = r(x_k, x_k) = 1$].

Standardna greška procjene može se računati i pomoću koeficijenta mnogostruke korelacije. Iz izraza (2.16) dobivamo

$$SGP = s(y) \cdot \sqrt{1 - R^2} \qquad (2.18)$$

Za $R = 1$, $SGP = 0$, što znači da su opažene i prognozirane (izračunate) vrijednosti jednake, za $R = 0$, $SGP = s(y)$, tj. korištenje jednadžbe regresije nema smisla.

2.4. Metoda odabiranja prediktora

Većinom nije poznato koje prediktore, i koliko njih, uključiti u jednadžbu regresije. Čak da prediktand i može biti koreliran s mnogo varijabli, jednadžba regresije treba sadržavati samo nekoliko njih koji objašnjavaju varijancu prediktanda isto tako dobro kao i jedna jednadžba koja sadržava mnogo varijabli. Treba dakle birati samo one prediktore koji značajno i neovisno doprinose prognozi prediktanda. Korištenjem manjeg niza prediktora sprečava se i nestabilnost jednadžbe i osiguravaju se dobri rezultati kad se ona primjenjuje na nove podatke.

Metoda odabiranja prediktora zove se screening (ili stepwise) regresija, a uveo ju je u meteorologiju Miller (1958), koji je slijedio jedan raniji nepublicirani rad Bryana iz 1944. g. Od tada ova metoda ima mnoge primjene u meteorologiji, Klein i ost. (1959) te niz primjena diskutiran u radu Glahna (1965).

U praksi je primijenjeno nekoliko varijanti ove općenite metode. Jedna od njih, možda i najjednostavnija, je tzv. forward stepwise method. Tehnika ove metode je u postupnom uključivanju nezavisnih varijabli, u jednadžbu regresije. Prvo se uključuje ona koja ima najveću vrijednost koeficijenta korelacije sa zavisnom varijablom. Izabrani prediktor i najviše reducira varijancu prediktanda. Tada se kao drugi prediktor izabire onaj koji zajedno s prvim izabranim reducira varijancu više nego bilo koji drugi prediktor od preostalih, i tako se nastavlja izbor jednog po jednog prediktora. Ovaj postupak je ekvivalentan računanju koeficijenta djelomične korelacije između prediktanda i svakog pojedinog od preostalih prediktora, držeći prethodno odabrane prediktore konstantnim, i odabire se kao slijedeći onaj koji daje najveću djelomičnu korelaciju. Završna točka odabiranja određuje se po volji u ovisnosti o F testu, kako su to predložili Lubim i Summerfield (1951). Međutim, budući da se odabiranje prediktora ne čini slučajno, F test nije strogo primjenljiv. Miller (1958) je predložio da se umjesto korištenja kritične vrijednosti F ($1-\alpha$) u svakom koraku odabiranja prediktora upotrijebi vrijednost F ($1-\alpha$) / ($P-S+1$); gdje je P ukupan broj mogućih varijabli, S je redni broj odabrane varijable, dok se α obično uzima 0.05. Ovom modificiranom F testu, koji daje putokaz, izbora broja prediktora, ne može se pripisati egzaktan signifikantni nivo. Obično se uzima da se odabiranje prediktora zaustavi kada slijedeći najbolji prediktor reducira ukupnu varijancu prediktanda za manje od recimo 1%. Diskusija ove metode neizbježno vodi na operacije s matricama, i detaljno ju je opisao Efroymson (1960).

Međutim, ni ova metoda ne osigurava da ćemo naći najbolji niz prediktora. Traženje najboljeg niza prediktora ograničeno je i kompjuterskim vremenom. Dakle ovdje, kao i u bilo kojoj tehnici predviđanja, meteorolog mora iskustvom odabrati prediktore, pomoću kojih bi mogao očekivati dobre rezultate.

2.5. **Kolmogorov — Smirnov test**

Metoda mnogostruke linearne regresije primjenljiva je samo u slučaju ako su i nezavisne varijable (prediktori) i zavisna varijabla (prediktand) razdijeljeni po normalnoj razdiobi. Za testiranje odstupanja empiričkih razdioba (razdioba podataka) odgovarajućih normalnih razdioba (s istim srednjacima i standardnim devijacijama) korišten je Kolmogorov-Smirnov test. Taj test kao kriterij uzima u obzir najveću razliku teoretske i empiričke kumulativne razdiobe. Kao nivo značajnosti, uzeto je $\alpha=0{,}01$; to je tzv. greška prve vrste — greška odbacivanja istinite hipoteze — ovdje da su razdiobe zaista normalne. Kolmogorov-Smirnov test za ovaj nivo daje kritičnu vrijednost 0,1352. Treba međutim istaći da je taj test primjenljiv, strogo gledajući, samo ako su parametri razdioba (dakle srednjak i varijanca) unaprijed poznati. U protivnom test je dosta »blag« (Benjamin i Cornell, 1970).

2.6. **Greške koeficijenata**

2.6.1. **Greške koeficijenata jednadžbe regresije**

Koeficijenti jednadžbe regresije b_i određeni su s izvjesnom greškom, jer su procijenjeni iz uzorka. Greške se mogu najzgodnije prikazati u matričnom obliku:

$$s^2(B)=B \,.\, B' = (x'x)^{-1} \,.\, x'y \,.\, y'x(x'x)^{-1} \qquad (2.19)$$

Pri izvodu se koristimo jednadžbom (2.9), te nekim poznatim svojstvima matrica, napose:

$$(x')' = x \quad \text{te} \quad [(x'x)^{-1}]' = (x'x)^{-1}$$

pa je $B'=y'x(x'x)^{-1}$.

Izraz (2.19) možemo prikazati u sažetijem obliku, uzevši da je:

$yy' = s^2(y)$ — varijanca zavisne varijable

$(x'x)\,(x'x)^{-1} = 1$ jedinična matrica

pa je, uz $(x,x)^{-1} = C$

$$s^2(B) = s^2(y) \cdot C \qquad (2.20)$$

Dijagonalni elementi ove matrice (matrice kovarijanci) daju varijance, a vandijagonalni elementi kovarijance procjene koeficijenata jednadžbe regresije.

Varijance se mogu računati i preko determinante transformirane matrice kovarijance (Δ).

$$s^2(b_j) = \frac{s(y)\,.\,\Delta_{jj}}{(n-k)\,.\,\Delta} \qquad (2.21)$$

koja je simetrična. Δ_{jj} je minor determinante, dobiven precrtavanjem j-tog retka i j-tog stupca; n je broj mjerenja, a k broj nezavisnih varijabli. Iz (2.21) slijedi da greške koeficijenata regresije rastu s povećanjem broja varijabli. Raspolažemo li s malim brojem opažanja (mjerenja), a s velikim brojem varijabli, polučujemo nepouzdane koeficijente regresije i manje statistički stabilna rješenja. Razvijanjem determinanti Δ i Δ_{jj} dobivamo za greške koeficijenata jednadžbe s 3 nezavisne varijable:

$$s^2(b_i) = \frac{s(y)}{s(x_i)} \quad \frac{1-r^2(x_j,x_k)}{(n-3)\,D(x,x)} \qquad (2.22)$$

$$i,j,k, = 1,2,3; \quad i\neq j\neq k.$$

2.6.2. **Greške koeficijenata korelacije**

Potječu li koeficijenti korelacije iz populacije koja se podvrgava zakonu bivarijatne normalne razdiobe, za dovoljno velik broj opažanja i za koeficijente koji nisu suviše blizu ± 1, razdioba koeficijenata korelacije iz uzoraka teži k normalnoj, s pravim (a nepoznate veličine) koeficijentom korelacije ρ kao srednjakom, i s varijancom

$$s^2(r) = \frac{(1-r^2)^2}{n-1} \qquad (2.23)$$

[U literaturi se nalazi i izraz $s^2(r) = \frac{(1-r^2)^2}{n}$]

Ta procjena varijance ima manju grešku od procjene dane s (2.23), ali za razliku od nje nije nepristrana. Za velike n te su procjene, uostalom, gotovo jednake].

Budući da srednjak te razdiobe nije poznat, interval povjerenja za koeficijente određujemo pomoću t-razdiobe:

$$r - t_{n-1}\, s(r) < \rho < r + t_{n-1} s(r) \qquad (2.24)$$

t je vrijednost varijable t, koji odgovara nekom nivou signifikantnosti (obično se odabire takva vrijednost da je u intervalu danom s (2.24) 95% svih vrijednosti koeficijenata korelacije dobivenih iz uzoraka).

Slično računamo i varijancu koeficijenata mnogostruke korelacije

$$s^2(R) = \frac{(1-R^2)^2}{n-k-1} \qquad (2.25)$$

k je broj nezavisnih varijabli.

Varijanca (2.23), koja je određena s točnošću $n^3/_2$, vrijedi i kao varijanca razdiobe koeficijenata korelacije bilo koje veličine, ali ta razdioba to više odstupa od normalne što su koeficijenti po apsolutnoj vrijednosti veći.

R. Fisher je dao transformaciju

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r} \quad \text{i} \quad \xi = \frac{1}{2} \frac{1+\rho}{1-\rho} \qquad (2.26)$$

i pokazao da je, za čak ne suviše velik broj opažanja (dovoljno je već $n \geqq 50$), veličina Z razdijeljena po zakonu normalne razdiobe sa srednjakom

$$E(z) = \xi + \frac{\rho}{2(n-1)}$$

(u računu se, umjesto nepoznatog ρ, uzima koeficijent korelacije iz uzorka r) i varijancom

$$s^2(z) = \frac{1}{n-3} \qquad (2.27)$$

Za male n postoji tablica sa Z-razdiobu. Fisherova Z-razdioba omogućava rješavanja raznih problema s kojima se češće susrećemo:

a) koliko je bitna razlika među koeficijentom korelacije iz uzorka (r) i koeficijentom korelacije cijele populacije ρ (poznatim ili pretpostavljenim);

b) koliko je bitna razlika među koeficijentima korelacije izračunatim iz dva ili više uzoraka;

c) koja je najbolja procjena koeficijenta korelacije, ako raspolažemo s 2 ili više uzoraka i njima pripadnim koeficijentima korelacije.

U praksi najčešće testiramo hipotezu da ne postoji linearna povezanost između dviju varijabli, tj. da je koeficijent korelacije $\rho=0$. Tada je veličina razdijeljena po t-razdiobi s (n—2) stupnja slobode.

$$t = \frac{r\sqrt{n-2}}{1-r^2} \qquad (2.28)$$

2.6.3. Greške zbog autokorelacije

Pri razmatranju jednadžbe linearne regresije i procjenjivanju njenih koeficijenata, pretpostavilo se da su vrijednosti zavisne varijable međusobno nekorelirane. U praksi, pri analizi meteoroloških podataka, obično statističke parametre procjenjujemo iz vremenskih nizova meteoroloških elemenata. Članovi tih nizova su obično povezani, tj. međusobno korelirani. Pritom to ne mora značiti da su i kauzalno povezani, već da stanja atmosfere pokazuju veliku perzistenciju.

Međusobna korelacija (autokorelacija) smanjuje obim informacije iz niza podataka, snižavajući točnost procjena statističkih parametara (povećavajući njihovu varijancu).

U teoriji linearne regresije pokazuje se da su procjene koeficijenata jednadžbe linearne regresije, uvaživši autokorelaciju:

$$B = (X'MX)^{-1}\, X'M^{-1}Y \qquad (2.29)$$

$$s^2(B) = (X'X)^{-1}\, X'MX\, (X'X)^{-1} \qquad (2.30)$$

gdje je M kovarijacijska matrica zavisne varijable y. Za međusobno nezavisne podatke y_i (i=1, ... n), mjerene s jednakom točnošću

$$M = s^2(y)$$

što izraz (2.29) svodi na izraz (2.19).

Koeficijenti autokorelacije su nam, naravno, nepoznati, pa ih procjenjujemo iz podataka. U praksi ih je mukotrpno procjenjivati, a ako ne raspolažemo s velikim brojem podataka i nemoguće. Zato se pri procjenjivanju koeficijenata regresije služimo formulom (2.9). Koeficijenti određeni prema toj formuli imaju veću srednju kvadratnu grešku od onih određenih prema (2.29), ali su također nepristrani.

2.6.4 Greške zaokruživanja

Računamo li koeficijente jednadžbe regresije s više nezavisnih varijabli, prilikom zaokruživanja brojeva s kojima računamo na određeni broj decimala, potreban je poseban oprez. To zaokruživanje može biti izvor velikim greškama, čemu se ne posvećuje dovoljna pažnja. Radimo li s matricom, inverznom matrici kovarijance, čiju determinantu označavamo s $D^{-1}(x,x)$, vidi izraz (2.11), tada se ona može računati na dva načina:

a) da podijelimo svaki član determinante s D(x,x)

b) da zajednički faktor izlučimo ispred determinante — tada množimo s $[\frac{1}{D(x,x)}]$

Aritmetički je to naravno svejedno, ali u praktičnom radu rezultati se mogu prilično razlikovati. Naročito ako je D(x,x) mali broj, a to je slučaj kad su koeficijenti korelacije među nezavisnim varijablama $r(x_1,x_2)$, $r(x_2,x_3)$ i $r(x_1,x_3)$ veliki (blizu 1), moramo biti oprezni. U graničnom slučaju za $r(x_i,x_j)=1$, $D^{-1}(x,x)$ je singularna matrica.

Drugi glavni uzrok grešaka je ako su brojevi koji ulaze u račun koeficijenata regresije znatno različitog reda veličine.

Posebno nije dobro ako su brojčane vrijednosti nezavisnih varijabli znatno veće od vrijednosti zavisne varijable. Tada je vrijednost izračunata jednadžbom regresije vrlo osjetljiva.

Najbolji je savjet da se radi s korelacijskom matricom, i da se pri računanju svih koeficijenata zadrži što više decimalnih mjesta, a sva zaokruživanja da se ostave za konačni račun — nakon svih množenja i dijeljenja.

3. PROGNOZA MINIMALNE TEMPERATURE U MJESECU SIJEČNJU ZA ZAGREB

3.1. Izbor prediktora i njihove statističke karakteristike

Da bismo dobivenu jednadžbu regresije mogli koristi u prognostičke svrhe, moraju se birati prediktori koji se prognoziraju numeričkim modelom ili modelima. Izbor je bio sužen činjenicom da prognostičke karte dobivene HIBU modelom, koje emitira preko faksimila služba SHMZ, daju samo vrijednosti prizemnog tlaka, geopotencijala na 850, 700 i 500 mbara, temperaturu na 850 mbara te položaj prizemnih fronti. Tako su kao potencijalni prediktori odabrani: prizemni tlak, temperatura na 850 mbara te relativne topografije 850/1000, 700/1000 i 500/1000 mbara, ukupno pet prediktora.

Kao daljnji prediktori uvedeni su vjetar na AT 850 mbara i vjetar pri tlu, i to srednji dnevni vjetar u m/sek. Za oba prediktora potrebna je daljnja interpretacijska obrada, pomoću koje bi mogli procijeniti njihovu vrijednost iz prognostičkih karata. Tako je broj potencijalnih prediktora povećan na sedam.

Budući da naoblaka ima velik utjecaj na minimalnu temperaturu, kao osmi prediktor uzeta je vlaga na 850 mbara, i to relativna vlaga. Smatrali smo da vlaga na 700 i 500 mbara nije toliko značajna, jer srednja i visoka naoblaka ne utječu bitnije na minimalnu temperaturu. Za sada HIBU model još ne daje vlagu, pa bi se za prognostičke svrhe i taj prediktor morao procijeniti prema razvoju sinoptičke situacije, ili advekcijom na karti AT 850 mbara. Često je međutim, naročito zimi, oblačni sloj na nižoj visini od 850 mbarske plohe, pa bi stoga bilo pogodnije uzimati srednju vlagu između tla i 850 mbara, što mi sada nismo u mogućnosti. Međutim neki modeli, kao npr. model evropskog centra za srednjoročnu prognozu vremena, u mogućnosti su da prognoziraju sve ove prediktore.

Kao deveti prediktor, za prognozu minimalne temperature, uzeta je prizemna relativna vlaga u 13 sati, dakle podatak iz prethodnog dana. Prizemnu vlagu od 13 sati koristili su i drugi autori kod određivanja regresionih jednadžbi za prognozu mraza kao Penzar (1957) te formula Mihajlevskog.

Kao posljednji, deseti, prediktor navedena je minimalna temperatura izmjerena prethodnog dana, i to klimatološkši minimum. Time je uzeta u račun perzistencija.

Prvih pet i osmi prediktor izvađeni su iz radiosondažnih podataka stanice Zagreb—Maksimir od 01 SEV, a ostali iz klimatoloških mjesečnih izvještaja. U račun je uzet petogodišnji niz, razdoblje 1975/79 godina. U tabeli 1. prikazane su osnovne statističke karakteristike odabranih prediktora, dobivene iz petogodišnjeg niza.

Tabela 1. Osnovne statističke karakteristike varijabli: minimalna i maksimalna vrijednost, srednjak, standardna devijacija (δ), odstupanje srednjaka (d) i Kolmogorov-Smirnov test (KS).

Table 1. Basic statistical characteristics of variables: minimum and maximum values, mean, standard deviation (δ), deviation from the mean (d) and Kolmogorov-Smirnov test (KS).

Varijabla	Min.	Max.	Srednjak Mean	δ	d	KS
RT 850/100	1236	1361	1306,9	23,1	3.7	0,0452
temp. na 850 mb	—17,5	10,2	—1,3	5,1	0,8	0,0582
RT 500/1000	5006	5536	5352,8	99,0	15,7	0,0658
tlak-pressure	975,5	1023	1002,0	9,8	1,6	0,0288
rel. vl. 850 mb humidity	18	100	73,7	22,6	3,6	0,1234
RT 700/1000	2663	2930	2832,2	52,5	8,3	0,0409
vjetar na 850 mb wind at 850 mb	0,8	22,1	9,2	5,3	0,8	0,0892
prizemni vjetar surface wind	0,4	4,3	1,6	0,8	0,1	0,0864
rel. vl. 13 SEV rel. humidity	28	100	73,1	16,5	2,6	0,0702
perzistencija	—17,3	7,1	—2,4	4,8	0,7	0,0896

Odstupanje srednjaka (d) znači da je vjerojatnost 0,95 da će pravi srednjak (μ) pasti u granice $\pm$d oko dobivenog srednjaka tj.

$$P(\bar{x}-d<\mu<\bar{x}+d) = 0,95.$$

U posljednjem stupcu, Kolmogorov - Smirnov test, dana su najveća odstupanja teoretske i empiričke kumulativne razdiobe (prethodno su razdiobe standardizirane, tj. suma svih podataka=1). Razdioba relativne vlage na 850 mbra najviše odstupa od normalne razdiobe, ali ipak ne toliko da bi prema ovom testu mogli odbaciti hipotezu o normalnoj razdiobi, jer je odstupanje još uvijek manje od kritičnog, koje je 0,1352. Sva ostala odstupanja su znatno manja, a najmanje odstupanje od normalne razdiobe daje tlak.

3.2. Matrica korelacije prediktora

U tabeli 2. prikazana je matrica korelacije odabranih prediktora. Vandijagonalni elementi te matrice su koeficijenti linearne korelacije između prediktora. Vidi se da su prediktori (koje ćemo sada skraćeno pisati) RT 850/1000, RT 700/1000 i RT 500/1000 mbara (RT85, RT70 i RT50), i temperatura na 850 mbara (T85) jako korelirani, što je i logično budući da svaki odražava toplinsku strukturu nižeg sloja atmosfere. Od preostalih prediktora najveće koeficijente korelacije, s već navedenim prediktorima, ima minimalna temperatura, dan ranije (TNDR), i to naročito s RT85. Preostali prediktori, prizemni tlak (p), relativna vlaga na 850 mbara (RV85), vjetar na 850 mbara (VJ85), srednji prizemni vjetar (VJPR) i relativna vlaga u 13 SEV, određena dan ranije (RVDR) razmjerno su nezavisni i njihova međusobna korelacija, kao i korelacija s prije navedenim prediktorima, je mala. To je, svakako, vrlo poželjno svojstvo kada se radi o izboru prediktora za jednadžbu mnogostruke regresije. Većina njihovih međusobnih koeficijenata korelacije je negativna, osobito za tlak koji jedino s prediktorom RT50 ima pozitivnu korelaciju.

3.3. Statističke karakteristike regresije

U daljnjem radu na prognozi minimalne temperature izvedena je jednadžba mnogostruke linearne regresije uz korištenje svih deset prediktora. Za računanje regresionih koeficijenata koristili smo programe s kojima raspolaže Sveučilišni računski centar u Zagrebu, a sva računanja su provedena na elektroničkom računalu UNIVAC 1110. U tabeli 3. prikazane su statističke karakteristike regresije. U prvom stupcu su jednostavni linearni koeficijenti korelacije između minimalne temperature i pojedinih prediktora. Najveću vezu minimalna temperatura pokazuje s RT85 i perzistencijom (TNDR), a zatim slijede RT70, T85 te RT50. Ostali koeficijenti korelacije su znatno manji. Korelacija tlaka i minimalne temperature je negativna: to je dobro poznata činjenica da je u zimskim mjesecima visoki tlak obično povezan s niskim temperaturama i obratno. U 2. stupcu testirana je hipoteza da ne postoji linearna povezanost između minimalne temperature i pojedinog prediktora, tj. da je koeficijent korelacije ρ=0. Vjerojatnost je dana na pet decimala, i ona je za prediktore koji

Tabela 2. Matrica korelacije prediktora.
Table 2. Matrix of correlation predictors.

	RT85	T85	RT50	p	RV85	RT70	VJ85	VJPR	RVDR	TNDR
RT85	1	0,898	0,817	—0,215	—0,042	0,921	0,246	0,149	0,059	0,474
T85		1	0,895	—0,028	—0,171	0,954	0,165	0,082	0,155	0,295
RT50			1	0,042	—0,025	0,944	0,227	0,006	0,169	0,237
p				1	—0,262	—0,077	—0,148	—0,254	—0,033	—0,268
RV85					1	—,060	—0,016	—0,109	0,286	0,136
RT70						1	0,205	0,089	0,147	0,328
VJ85							1	0,226	—0,190	—0,023
VJPR								1	—0,418	0,140
RVDR									1	0,066
TNDR										1

Tabela 3. Statističke karakteristike regresije.
Table 3. Statistical characteristics of regression.

	r	$\rho=0$	r(y,x$_i$)	b$_i$(S)	s[b$_i$(S)]	s[b$_i$(S)] / b$_i$(S)	$\beta=0$	b$_i$	RV
RT85	0,691	0,00000	0,156	0,29083	0,15464	0,53	0,06207	0,05679	20,11
T85	0,550	0,00000	0,076	0,15408	0,16956	1,10	0,36504	0,13556	8,47
RT50	0,525	0,00000	0,013	0,02575	0,15994	6,21	0,87231	0,01176	1,35
p	—0,273	0,00065	0,047	0,03029	0,05351	1,77	0,57225	0,01395	—0,83
RV85	0,269	0,00077	0,401	0,27822	0,05328	0,19	0,00000	0,05567	7,49
RT70	0,600	0,00000	0,002	0,00637	0,25464	40,0	0,98009	0,00055	0,38
VJ85	0,206	0,01052	0,150	0,08951	0,04948	0,55	0,07257	0,07667	1,85
VJPR	0,232	0,00397	0,180	0,11424	0,05249	0,46	0,03117	0,62683	2,65
RVDR	0,056	0,48920	—0,052	—0,03388	0,05485	1,62	0,53778	0,09264	—0,19
TNDR	0,690	0,00000	0,570	0,45727	0,05527	0,12	0,00000	0,43116	31,55

su imali veliki koeficijent korelacije s minimalnom temperaturom mala. Najveća je za relativnu vlagu u 13 SEV, što je logično jer ona ima i najmanji koeficijent korelacije.

Koeficijenti djelomične korelacije između minimalne temperature i pojedinog prediktora, uz pretpostavku da je isključen utjecaj preostalih prediktora, 3. stupac, pokazuju da su gotovo jedini relevantni faktori u predikciji perzistencija i relativna vlaga na 850 mbara, a zatim slijede $\overline{VJPR}$, RT85 te VJ85. Ostali koeficijenti djelomične korelacije su znatno manji, dok je za relativnu vlagu u 13 SEV negativan.

Četvrti stupac odnosi se na regresijske koeficijente, koji su u stvari koeficijenti jednadžbe regresije za standardizirane varijable, $b_i(S)$, tj. za varijable podijeljene s odgovarajućim standardnim devijacijama. U 5. stupcu dana je standardna devijacija, a u 6. stupcu relativna greška $b_i(S)$ koeficijenta, dobivena dijeljenjem vrijednosti u 5. stupcu s vrijednosti u 4. stupcu. Najmanje relativne greške daju koeficijenti perzistencije, relativne vlage na 850 mbara, srednjeg prizemnog vjetra, relativne topografije 850/1000 mbara i vjetra na 850 mbara, što ukazuje da su ovi prediktori najprikladniji da se koriste u jednadžbi mnogostruke linearne regresije, za prognozu minimalne temperature. Na isti zaključak navodi nas i testiranje hipoteze $\beta=0$, 7. stupac.

U 8. stupcu dani su pravi koeficijenti regresijske jednadžbe, dobiveni tako da se odgovarajući regresijski $b_i(S)$ koeficijenti množe sa standardnom devijacijom minimalne temperature i dijele s pripadnom standardnom devijacijom nezavisne varijable. Poželjno je da koeficijenti budu što stabilniji, tj. da regresijski koeficijenti iz nekog drugog uzorka ne odstupaju od prvotnih (u našem slučaju regresijski koeficijenti izvedeni su iz niza od 153 podataka). Slobodni član b_o računa se po (2.13). On se može i izbjeći, ako kao varijable uzmemo njihova odstupanja od pripadnih srednjaka.

U 9. stupcu dani su doprinosi pojedinog prediktora objašnjenju varijance (RV). Ukupan doprinos svih prediktora je 73,85%, dakle velik, a time je i koeficijent mnogostruke korelacije razmjerno velik i iznosi 0,8594. Najveći doprinos redukciji varijance prediktanda daje perzistencija i relativna topografija 850/1000 mbara, a zatim slijede temperatura na 850 mbara i relativna vlaga na istom nivou. Srednji prizemni vjetar daje veći doprinos od vjetra na 850 mbara.

3.4. Jednadžbe regresije za prognozu minimalne temperature

Ako jednadžbu mnogostruke regresije uz korištenje svih deset prediktora, tabela 3., primijenimo u praksi, ona usprkos velikom doprinosu redukciji varijance prediktanda ne daje dobre rezultate. Razlozi su velik broj međusobno zavisnih prediktora i premalen niz podataka za njeno određivanje. Ipak tabele 2. i 3. su nam pomogle da biramo najbolju kombinaciju prediktora za prognozu minimalne temperature, uzimajući one koji daju najveću redukciju varijance, a međusobno su nezavisni. Određeno je više regresijskih jednadžbi uz korištenje većine prediktora, osim prizemnog tlaka i RT 700/1000 mbara, (p je ispitan ranije, Belamarić i Kisegi, 1981).

U prvoj kombinaciji odabrani su prediktori koje daje HIBU model, i to RT50, T85 i RT85. Dobivena je jednadžba regresije za prognozu minimalne temperature kojoj pridodajemo uz TN i broj 1.

$$TN1 = -0{,}0008(RT50) - 0{,}2311(T85)+0{,}1622(RT85) - 209{,}8 \qquad (3.1)$$

Međutim i ova jednadžba ne daje dobre rezultate. Jedan od razloga je što ne sadržava prediktor naoblake, a drugi što su korišteni međusobno zavisni prediktori.

Zatim je određeno nekoliko jednadžbi regresije tako da su birane kombinacije međusobno nezavisnih prediktora. Odabrani su RT85 i RV85, a uz njih još u pojedinim kombinacijama i prediktori VJ85, TNDR i $\overline{VJPR}$. Tako su dobivene slijedeće jednadžbe:

$$TN2 = 0{,}0912(RT85)+0{,}046(RV85)+0{,}09(VJ85)+0{,}41(TNDR) - 124{,}7 \qquad (3.2)$$

$$TN3 = 0{,}1325(RT85)+0{,}063(RV85)+0{,}007(VJ85)+0{,}89(\overline{VJPR}) - 181{,}7 \qquad (3.3)$$

$$TN4 = 0{,}0981(RT85)+0{,}047(RV85)+0{,}4(TNDR) - 133{,}0 \qquad (3.4)$$

$$TN5 = 0{,}0958(RT85)+0{,}05(RV85)+0{,}7(\overline{VJPR})+0{,}38(TNDR) - 131{,}2 \qquad (3.5)$$

Budući da najveći doprinos objašnjenju varijance prediktanda daje perzistencija, tabela 3, određena je jednadžba regresije uz korištenje samo tog prediktora

$$TN6 = 0{,}69(TNDR) - 0{,}7\,. \qquad (3.6)$$

U tabeli 4. navedeni su koeficijenti korelacije, dobiveni iz petogodišnjeg niza, kada bismo minimalnu temperaturu koristili za prognozu i u slijedeće dane, tj. za drugi, treći, ... do sedmog dana.

Tabela 4. Koeficijenti korelacije minimalnih temperatura.
Table 4. Correlation coefficients of minimum temperature.

	1 dan	2 dana	3 dana	4 dana	5 dana	6 dana	7 dana
koeficijent kor.	0,690	0,4475	0,2832	0,2024	0,1297	0,0302	—0,0387

Koeficijent korelacije se postupno smanjuje i sedmog dana postaje negativan, tako da perzistencija ima smisla za prognozu do 3 dana.

U daljnjem radu uveden je kao prediktor relativna vlaga u 13 SEV, pa je dobivena jednadžba.

$$TN7 = 0{,}0948(RT85)+0{,}043(RV85)+0{,}55(\overline{VJPR})+0{,}026(RVDR)+0{,}41(TNDR) - 130{,}9 \qquad (3.7)$$

Isti ovi prediktori korišteni su i za određivanje koeficijenata regresije iz desetgodišnjeg niza, razdoblje 1972/81, pa je dobivena jednadžba

$$TN8 = 0{,}09693(RT85)+0{,}04441(RV85)+0{,}83226 (\overline{VJPR}) + 0{,}02341(RVDR) + 0{,}37823(TNDR) - 134{,}41 \,. \quad (3.8)$$

Na ovaj način dobiveni su stabilniji koeficijenti. Oni su se za prediktore RT85 i RV85, a osobito za $\overline{VJPR}$, povećali, dok su se nešto smanjili za prediktore RVDR i TNDR.

3.5. Ocjena jednadžbi regresije na podacima siječnja 1982.

Mjesec siječanj 1982. odabran je zato što se nalazio u razdoblju programa ALPEX, a nije se nalazio u nizu iz kojeg su određivane jednadžbe.

Provjera uspješnosti regresijskog modela, tj. jednadžbi regresije, može se provesti pomoću analize reziduala (Draper i Smith, 1966). Reziduali (e_i) su diferencije vrijednosti zavisne varijable dobivene pomoću jednadžbe regresije ($\hat{y}_i$) i stvarnih (opaženih) vrijednosti (y_i):

$$e_i = \hat{y}_i - y_i \qquad i = 1, \ldots, n$$

i je redni broj podatka.

Vidi se da su reziduali iznos koji regresija nije uspjela objasniti. Nadalje, ako je regresijski model ispravan, reziduali predstavljaju slučajne greške.

U regresijskoj analizi uobičajene pretpostavke za greške su:

1). greške su nezavisne,
2) njihov srednjak jednak je nuli,
3) varijanca je konstantna,
4) slijede normalnu razdiobu.

Pretpostavka o nezavisnosti nije sasvim ispunjena. Što je više parametara (koeficijenata korelacije, odnosno koeficijenata jednadžbe) procijenjeno iz podataka, to su reziduali više korelirani (tj. nisu potpuno nezavisni). Ipak, ako je broj stupnjeva slobode ($S=n-p$) velik, tj. $n>p$ (n=broj podataka, p=broj parametara-nezavisnih varijabli) autokorelaciju reziduala možemo zanemariti.

Ako se jednadžbe regresije primjene na niz iz kojeg su određene, onda su pretpostavke 2) i 4) ispunjene (Belamarić i Kisegi, 1981).

Pretpostavka 3) da je varijanca konstantna, tj. neovisna o vrijednostima zavisne i nezavisnih varijabli, kao i o vremenu, nije provjerena u ovom radu.

U tabeli 5. prikazana je ocjena jednadžbi regresije (3.1) do (3.8) analizom reziduala. Računali smo srednjak, standardnu devijaciju i varijancu reziduala, standardnu grešku procjene, koeficijent mnogostruke korelacije i redukciju varijance. U posljednja četiri retka dana su maksimalna pozitivna i negativna odstupanja prognozirane od stvarne minimalne temperature s datumom za koji su dobivene.

Standardna devijacija minimalne temperature za siječanj 1982. je 3,855.

Ocjena je rađena samo za 30 dana. Nije uzet u račun 9. siječanj budući da nije bilo TEMP podataka od 01 SEV.

Iz tabele 5. se vidi, kad se jednadžbe regresije primijene na podatke koje nismo koristili za njihovo određivanje, da srednjak reziduala nije jednak nuli. Većina jednadžbi dala je u prosjeku nižu vrijednost minimalne temperature. Najveću srednju grešku daju (3.1) koja ne sadržava prediktor perzistencije, te (3.6) koja uzima u račun samo perzistenciju. Obje su jednadžbe u prosjeku dale za oko 1°C višu temperaturu.

Standardna devijacija reziduala kreće se od 2,58 kod (3.6) do 3,17 kod (3.3) koja također ne sadrži kao prediktor perzistenciju.

Razmatrajući srednju grešku procjene, koja se kreće od 2,56 kod (3.4) do 3,14 kod (3.3), vidimo da je ona manja od standardne devijacije prediktanda [$s(y)=3{,}855$], a to prema (2.18) znači da korištenje svih ocijenjenih jednadžbi regresije ima smisla, i da one u najvećem broju dana daju zadovoljavajuće rezultate prognoze.

Koeficijent mnogostruke korelacije je kod svih jednadžbi dosta visok, i kreće se od 0,58 (3.3) do 0,75 (3.4), dok je kod (3.7) neznatno niži (0,74).

Analogno je i s redukcijom varijance, koja je jednaka R^2.

Uspoređujući (3.4) i (3.7), vidimo da druga sadrži dva prediktora više, $\overline{VJPR}$ i RVDR, pa zato uključuje i veće mogućnosti za prognozu minimalne temperature. To su razlozi zbog kojih jednad-

Tabela 5. Ocjena jednadžbi regresije, analizom reziduala, na podacima siječnja 1982.

Table 5. Test of validaty of the regression equations by the residual analysis on January 1982 data.

	(3.1)	(3.2)	(3.3)	3.4)	,3.5)	(3.6)	(3.7)	(3.8)
e	1,11	—0,22	—0,33	—0,05	—0,08	1,08	0,26	—0,07
s_e	2,73	2,73	3,17	2,61	2,84	2,58	2,60	2,74
V_e	7,46	7,45	10,05	6,79	8,07	6,66	6,75	7,48
SGP	2,91	2,69	3,14	2,56	2,79	2,76	2,58	2,69
R	0,66	0,72	0,58	0,75	0,69	0,70	0,74	0,72
RV	0,43	0,51	0,34	0,56	0,48	0,49	0,55	0,51
$+e_{max}$	5,9	6.4	9,0	6,9	9,5	6,1	8,0	8,7
datum	16.	7.	7.	7.	7.	7.	7.	7.
$-e_{max}$	—3,1	—5,2	—7,3	—4,8	—5,0	—5,1	—4,3	—4,9
datum	26.	21.	26.	21.	21.	28.	21. i 26.	26.

žbu (3.7), usprkos neznatno manje vrijednosti koeficijenta mnogostruke korelacije i redukcije varijance, treba smatrati najboljom. Zato su korištenjem istih prediktora određeni stabilniji koeficijenti regresije iz desetogodišnjeg niza, jednadžba (3.8). Međutim, ona je kod primjene za prognozu minimalne temperature u siječnju 1982. dala nešto lošije rezultate od (3.7). Razlog treba tražiti što je siječanj 1982. bio izuzetno perzistentan. U većem dijelu mjeseca naši krajevi su se nalazili pod utjecajem hladnog arktičkog zraka, koji je prodro 7. ujutro, a u razdoblju od 14. do 21. siječnja nad područjem većeg dijela zemlje, pa tako i nad Zagrebom, postojala je jaka anticiklonalna inverzija. Zato je jednadžba (3.6), koja kao prediktor koristi samo perzistenciju, dala dobru prognozu u siječnju 1982.

Prilikom ocjene uspješnosti prognoze, odstupanja koja su veća od $2s_e$ smatrat ćemo grubim greškama.

U tabeli 5. vidimo da je kod svih jednadžbi maksimalno pozitivno odstupanje bilo veće od $2s_e$, dok su maksimalna negativna odstupanja bila većinom manja od $2s_e$.

Sve jednadžbe, osim (3.1), dale su najveće pozitivno odstupanje 7. siječnja. Tog dana u jutarnjim satima došlo je do prodora izrazito hladne arktičke zračne mase. Prognoza minimalne temperature izračunata je na osnovi prediktora izvađenih iz TEMP izvještaja od 01 SEV, a oni su se već do jutra znatno izmijenili. Samo relativna topografija 850/1000 mbara od 7. na 8. siječnja smanjila se za 61 geopotencijalni metar, a temperatura na 850 mbara pala je za 14,9°C. Da se u jednadžbe išlo sa vrijednostima prediktora od 8. siječnja 01 SEV rezultati prognoze bili bi daleko bolji. Samo smanjenje RT 850/1000 za 61 gpm daje po jednadžbi (3.8) smanjenje minimalne temperature za 6,9°C, tako bi odstupanje, koje je iznosilo 8,7°C, bilo svega 1,8°C.

Maksimalna negativna odstupanja su kod većina jednadžbi nastupila 21. ili 26. siječnja, a samo kod (3.6) 28 siječnja. U oba slučaja radilo se o pritjecanju toplije zračne mase u razdoblju poslije 01 SEV, a 21. siječnja došlo je tokom jutra i do naoblačenja, tako da bi i u tim slučajevima uzimanje u račun ovih značajnih promjena prediktora, do kojih je došlo neposredno poslije termina 01 SEV, dalo bolji rezultat.

Ovdje bismo spomenuli da je poznata i teoretski objašnjena činjenica da je regresijska prognoza bolja u području oko srednjaka prediktanda a mnogo lošija za ekstremne vrijednosti prediktanda (Draper i Smith, 1966; Benjamin i Cornell, 1970).

U tabeli 6. prikazana je razdioba reziduala dobivena jednadžbama (3.4), (3.7) i (3.8).

Granice klasa su: za odstupanje 0, od —0,5 do 0,5, za odstupanje ±1, od ±0,6 do ±1,5 itd.

Vidimo da (3.8), usprkos što u ekstremnim slučajevima ima najveću čestinu odstupanja, 2 slučaja u klasi —5, a jedan slučaj u klasi 9, daje u prosjeku najbolje rezultate. Naime, od izračunatih 30 prognoza, 23 prognoze kod (3.8) imale su odstupanja u granicama od —2,5 do 2,5°C (klase od —2 do 2). Kod (3.4) to se dogodilo u 20, a kod (2.7) u 19 slučajeva.

4. ZAKLJUČAK

Pokazalo se da najbolje rezultate daju jednadžbe regresije koje sadrže prediktore koji su međusobno nezavisni, a u najvećoj mjeri reduciraju varijancu pediktanda. Od razmatranih deset prediktora, pokazala se najboljom, što naravno ne znači da ne postoji i bolja, kombinacija prediktora RT 850/1000 mbara, relativna vlaga na 850 mbara, srednji prizemni vjetar u m/sek, relativna vlaga u 13 SEV i perzistencija.

Analiza reziduala i verifikacija rezultata je pokazala, da bi se najveća odstupanja u prognozi minimalne temperature izbjegla ako bi uzeli u račun eventualne značajnije promjene prediktora, tj. uveli nove vrijednosti prediktora, do kojih je došlo neposredno poslije 01 SEV. Najveći broj odstupanja prognozirane od stvarne minimalne temperature pada u razmjerno uski interval od —2,5 do 2,5°C.

Povećanjem niza podataka dobivaju se stabilniji regresioni koeficijenti, a time i bolji rezultati, što je i pokazala jednadžba (3.8).

Međutim, krajnji cilj — prognoza minimalne temperature za više dana unaprijed — ovisit će uvelike o točnosti karata numeričke prognoze, iz kojih bi se dobivale vrijednosti prediktora.

Tabela 6. Razdioba reziduala za mjesec siječanj 1982.
Table 6. The distribution of residuals for January 1982.

	—5	—4	—3	—2	—1	0	1	2	3	4	5	6	7	8	9
(3.4)	1	2	3	2	3	7	5	3	2	1			1		
(3.7)		2	3	1	6	5	3	4	4	1				1	
(3.8)	2		3	4	4	5	6	4		1					1

LITERATURA

Barry, R. G.,. and A. H. Perry, 1973: Synoptic climatology. Methuen and Co., London.

Belamarić, G. i M. Kisegi, 1981: Statističko-dinamička prognoza meteoroloških elemenata; prognoza minimalne temperature za Zagreb. RHMZ SRH, Centar za meteorološka istraživanja, 62. str.

Belov, P. N., 1975: Čislenije metodi prognoza pogodi. Gidrometeoizdat, Lenjingrad.

Benjamin, J. R., and C. A. Cornell, 1970: Probability, statistics and decision theory for engineers. Mc Graw-Hill, New York.

Brelsford, W. M. and R. H. Jones, 1967: Estimating probabilities. Mon. Wea. Rev., 95, 570—576.

Carter, G. M., A. L. Forst, W. H. Klein, and J. P. Dallavalle, 1978: Improved automated forecasts of maximum/minimum and 3-hourly temperatures. Preprint, Conference on Weather Forecasting and Analysis and Aviation Meteorology, Silver Spring, Md., Oct. 8 pp.

Draper, W. R., and H. Smith, 1966: Applied regression analysis, John Wiley and Sons, New York—London—Sidney.

Efroymson, M. A., 1960.: Multiple regression analysis. Mathematical Methods for Digital Computers, A. Ralston and H. S. Wilf, Eds., New York, Wiley 191—203.

Epstein, E. S., 1969: Stochastic dynamic prediction. Tellus, 21, 739—757.

Fleming, R. J., 1971: On stochastic dynamic prediction. Mon. Wea. Rev., 99, 851—872.

Glahn, H. R., 1965: Objective weather forecasting by statistical methods. The Statistician, Vol. 15, No 5, 111—141.

Hammons, G. A., J. P. Dallavalle and W. H. Klein, 1976: Automated temperature guidance based on three-month seasons. Mon. Wea. Rev., 104, 1557—1564.

Jones, R. H., 1968: A nonlinear model for estimating probabilities of K events. Mon. Wea. Rev., 96, 383—384.

Kisegi, M., 1981: Objektivna prognoza kondicionalne vjerojatnosti snježnih oborina u zapadnom području unutrašnjosti Hrvatske, Zbornik meteoroloških i hidrometeoroloških radova, 7, 25—35.

Klein, W. H., 1978: Objective forecasts of local weather bu means of model output statistics Seminars on the interpretation and use of large scale numerical forecasts products, Bracknnell, 186—220.

———, B. M. Lewis, and J. Engel, 1959: Objective prediction of five-day mean temperatures during winter. J. Meteor., 16, 672—682. and F. Lewis, 1970: Computer forecasts of maximum and minimum temperatures. J. Appl. Meteor., 9, 350—359.

———, and F. Marshall, 1973: Screening improved predictors for automated max/min temperature forecasting. Preprints 3rd Conf. on Probability and Statistics in the Atmospheric Sciences, Amer. Meteor. Soc., Boston, Mass., 36—43.

Lönnqvist, O., 1978: Automatic interpretation of forecast charts. Seminar on the interpretation and use of large scale numerical forecasts products, Bracknell, 117—137.

Lubin, A., and A. Summerfield, 1951.: A square root method of selecting a minimum set of variables in multiple regression. I the Method. Psychometrika, 16, 271—284.

Miller, R. G., 1958: The screening procedure (in »Studies in Statistical Weather Prediction«). Final Report Contract No. AF 19 (604)—1590 (Editer by Shorr), The travelers Research Center, Inc., Hortford, Conn., 86—95.

Penzar, I., 1957: Prilog prognozi mraza u našim krajevima. Rasprave i prikazi RHMZ SRH, 1, 16 str.

SUMMARY

In order to develop a statistical method for forecasting minimum temperature for January in Zagreb, this study employs the equation of multiple linear regression.

The first part of this paper presents a review of various approaches to weather forecasting using statistical methods. The basis of correlation and regression methods and the manner for Selecting the predictors are presented, and the possible errors which might appear in application are discussed.

The second part of this paper presents the selection of ten potential predictors from the data of the observatory Zagreb—Maksimir, which are: minimum temperature, relative humidity at 12 GMT on the previous day, surface pressure, 850 mbar temperature, relative humidity and wind at 850 mbar, thickness 850/1000, 700/1000 and 500/1000 mbar at 12 GMT, and the mean surface wind for the forecasting day. Statistical characteristics of regression (Table 3) show that for the minimum temperature prediction the most suitable predictors are: persistency, relative humidity at 850 mbar, mean surface wind, 850/1000 thickness and wind at 850 mbar. These predictors give the largest contribution of predictor's reduction of variance, and their coefficients of regression have the smallest relative error.

Seven regression equations are used as a combination of three to five most suitable predictors, except for eq. (3.1). All equations are determined from a five-year data series 1975—79. The verification is done for January 1982. For the combination of five mutually independent predictors, which indicated the largest contribution to the reduction of variance of the predictand, the equation (3.8) is determined from a 10-year data series 1972—81, which resulted in more stable coefficients of regression.

Analysis of residuals (Table 5) has shown that all equations give relatively high coefficients of regression, varying from 0,58 in (3.3) to 0,75 in (3.4).