ОПРЕДЕЛЯНЕ НА ДНЕВНИТЕ СУМИ НА СУМАРНАТА СЛЪНЧЕВА РАДИАЦИЯ ЧРЕЗ МЕТОДА НА МНОЖЕСТВЕНАТА ЛИНЕЙНА РЕГРЕСИЯ

Здравка Николаева, Гинка Байкушева-Димитрова
Университет "Проф. д-р Асен Златаров", факултет "Природни науки"
8010 Бургас,България, бул. "Проф. Якимов" 1,
burievazdr@btu.bg

DETERMINATION OF DAILY AMOUNT OF TOTAL SOLAR RADIATION BY THE METHOD OF MULTIPLE LINEAR REGRESSION

Zdravka Nikolaeva, Ginka Baikusheva-Dimitrova
University "Prof. Dr. Asen Zlatarov" 8010 Burgas, Bulgaria
burievazdr@btu.bg

ABSTRACT

In this paper is used the method of multiple linear regression for determination of daily amount of total solar radiation. An empirical formula of relationship between a total solar radiation and values: average maximal and minimal temperatures, average wind speed, precipitation, atmospheric pressure and horizontal visibility was developed. Statistical values were obtained with the LINEST function of EXCEL, which in addition to the coefficients of the regression equation are given: the standard errors of coefficients, coefficient of determination, standard errors of assessment, F – statistics, degrees of freedom, the regression sum of squares and residual sum of squares.

An assessment of the credibility of the high values of F – statistics and test of hypothesis testing for statistical significance of coefficients are made. The analysis shows that only three coefficients of the regression equation are useful for estimation and prognosis or were statistically significant. They are the coefficients before the average maximal and minimal temperatures and average wind speed.

In our investigation was deduced an empirical formula by the method of multiple linear regression. This provides possibility to calculate the daily amount of total solar radiation in Sofia using only data on the average maximal and minimal temperatures and average wind speed measured at certain location. The resulting coefficient of determination value is very close to unity ($R^2 = 0.952$). It means that the method of multiple linear regression is precise and correct.

Key words: *solar radiation, the method of multiple linear regression, the standard errors, coefficient of determination, F – statistics.*

ВЪВЕДЕНИЕ

Слънчевата радиация е основен климатичен и екологичен фактор, като в последните години нарастват усилията за нейното изучаване. Това се дължи на повишения интерес към слънчевата енергия като алтернативен източник на енергия [1, 2]. Информация за слънчевата радиация е необходима при проектиране на отоплителни и хладилни системи за дома, изготвянето на парници и др.

За определяне на дневните суми на сумарната слънчева радиация е използван метода на множествената линейна регресия. Изведена е емпирична формула на връзката между сумарната слънчева радиация и величините: средна максимална и минимална температури; средна скорост на вятъра; количество валежи; атмосферно налягане и хоризонтална видимост. Получени са статистически величини с функцията LINEST на EXCEL, с която освен коефициентите на регресионното уравнение са дадени и: стандартните грешки на коефициентите; коефициентът на детерминация; стандартната грешка на оценката; *F*–

статистиката; степените на свобода; регресионната сума на квадратите и остатъчната сума на квадратите.

Направена е оценка на правдоподобността на високата стойност на F–статистиката и тест за проверка на хипотези за статистическата значимост на коефициентите.

МЕТОДИ НА ИЗСЛЕДВАНЕ

За определяне на дневните суми на сумарната слънчева радиация е използван метода на множествената линейна регресия. Регресионната статистика е широко използвана техника за анализ на експериментални данни [3-7].

При линейна регресия връзката между случайните променливи X и Y има вида:

$$\widehat{y} = ax + b, \quad (1)$$

където: a и b са коефициентите на регресията. За да се определят коефициентите a и b така, че правата да минава най-близо до точките (x_i, y_i), се използва методът на най-малките квадрати.

В Excel изчисляването на коефициентие на уравнението на регресия може да се извърши по различни начини: чрез функциите LINEST, SLOPE, INTERCEPT, с инструмента Regression на Data Analysis или чрез добавяне на линия на тенденцията (Trendline) към серия от данни в диаграма.

Чрез използването функцията LINEST резултатът от пресмятанията се получава в 5 реда и 2 колони:

a - коефициент на регресия	**b** - коефициент на регресия
$\mathbf{S_{ea}}$ - стандартна грешка на a	$\mathbf{S_{eb}}$ - стандартна грешка на b
$\mathbf{R^2}$ - коефициент на детерминация	$\mathbf{S_{\varepsilon}}$ или $\mathbf{S_{ev}}$ - стандартна грешка на оценката
F - статистика	$\mathbf{d_F}$ - степени на свобода
$\mathbf{SS_{reg}}$ - регресионна сума на квадратите	$\mathbf{SS_{resid}}$ - остатъчна сума на квадратите

В общия случай функцията LINEST може да се използва за намиране на коефициентите на уравнението (множествена линейна регресия):

$$y = a1x1 + a2x2 + ... + anxn + b, \quad (2)$$

където: a_i (i = 1, 2, …, n) и b са коефициентите на регресията.

За получаване на коефициентите и другите резултати е необходимо преди извикването на функцията да се маркират клетки в 5 реда и n+1 колони. Коефициентите се получават в обратен ред: a_n, ..., a_2, a_1, b.

При регресионния анализ може да се използва и инструмента Regression на Data Analysis. Data Analysis се отваря с Tools/Data Analysis.

F и d_F в резултата може да се използва за оценка на правдоподобността високата стойност на F да е получена случайно. F може да се сравни с критичните стойности от таблици на F – разпределението или да се използва функцията FDIST. По този начин може да се прецени дали полученото регресионно уравнение е полезно за прогнозиране на функцията.

Друг тест за проверка на хипотези може да определи дали всеки коефициент a_n, ..., a_2, a_1, b е полезен за оценка на функцията. Това става чрез изчисляване на t – статистика.

РЕЗУЛТАТИ И ОБСЪЖДАНЕ

Използваните от нас данни са взети от интернет сайт [8]. Те са за сумарната слънчева радиация S, W/m^2; средната максимална T, K и минимална температура T_0, K; средната скорост на вятъра v, m/s; количеството валежи r, mm; атмосферното налягане p, hPa и хоризонталната видимост L, km. Данните са за София, 1961 – 1990 г., който е един от приетите базисни периоди (климатична норма) [9, 10]. Данните за продължителността на слънчевото греене са от [11, 12].

В повечето случаи няма данни за слънчевата радиация на дадено място и тя трябва да се оцени по наличните метеорологични данни.

За определяне на дневните суми на сумарната слънчева радиация използваме метода на множествената линейна регресия. Изследвана е зависимостта на S от величините: средната максимална T, K и минимална температура T_0, K; средната скорост на вятъра v, m/s; количеството валежи r, mm; атмосферното налягане p, hPa и хоризонталната видимост L, km.

Общият вид на полученото регресионно уравнение има вида:

$$y = a1.x1 + a2.x2 + a3.x3 + a4.x4 + a5.x5 + a6.x6 + b, \qquad (3)$$

където: a_1 – коефициентът пред T; a_2 – коефициентът пред T_0; a_3 – коефициентът пред v; a_4 – коефициентът пред r; a_5 – коефициентът пред p и a_6 – коефициентът пред L.

Получените статистически величини с функцията LINEST са представени в таблица 1. Освен коефициентите от уравнение (3) са дадени и: стандартните грешки на коефициентите (S_e); коефициентът на детерминация R^2; стандартната грешка на оценката S_{ev}; F – статистиката; d_F – степени на свобода; SS_{reg} - регресионна сума на квадратите и SS_{resid} – остатъчна сума на квадратите.

Таблица 1. Получени статистически величини за сумарната слънчева радиация с метода на множествената линейна регресия при оценка на дневните суми на сумарната слънчева радиация.

	a_6		a_5	a_4	a_3	a_2	a_1	b
	-7.00284		4.380405	0.57191	68.355	-28.7689	34.80572	-6280.0
S_e	3.837		2.905	0.629	18.445	7.582	6.099	3320.62
R^2	0.964975	S_{ev}	16.79423					
F	133.164	d_F	29					
SS_{reg}	225350.4	SS_{resid}	8179.34					

Направена е ***оценка*** на правдоподобността на високата стойност на F, чрез използването на функцията FDIST (F, v_1, v_2). За получените статистически величини: $v_1 = n - d_F - 1 = 6$; $v_2 = d_F = 29$.

Направен е и тест за ***проверка на хипотези***. Той определя дали всеки коефициент a_1, a_2, a_3, a_4, a_5, a_6. и b в уравнение (3) е полезен за оценката на y. За целта се проверява статистическата значимост на коефициентите. Получената оценка и проверка на хипотези е представена на таблица 2.

От таблица 2 се вижда, че при оценката и използване на функцията FDIST, се получава много малка вероятност за големи F стойности, получени случайно (FDIST = $9.36.10^{-20}$).

Направен е и тест за проверка на хипотези. От таблица 2 се вижда, че получените стойности за статистическите значимости на коефициентите a_1, a_2 и a_3, са по-големи от критичната стойност TINV = 2.04841:

5.70637>2.04841 ; **3.79420>2.04841** ; **3.70585>2.00758**,

а останалите статистически зависимости на другите коефициенти са по-малки от критичната стойност TINV = 2.04841.

Това показва, че ***само коефициентите a_1, a_2 и a_3 в регресионното уравнение са полезни за прогнозиране и оценка***, а това са коефициентите пред средната максимална T, K, минимална температура T_0, K и средната скорост на вятъра v, m/s.

Таблица 2. Получена оценка и тест за проверка на хипотези за сумарната слънчева радиация.

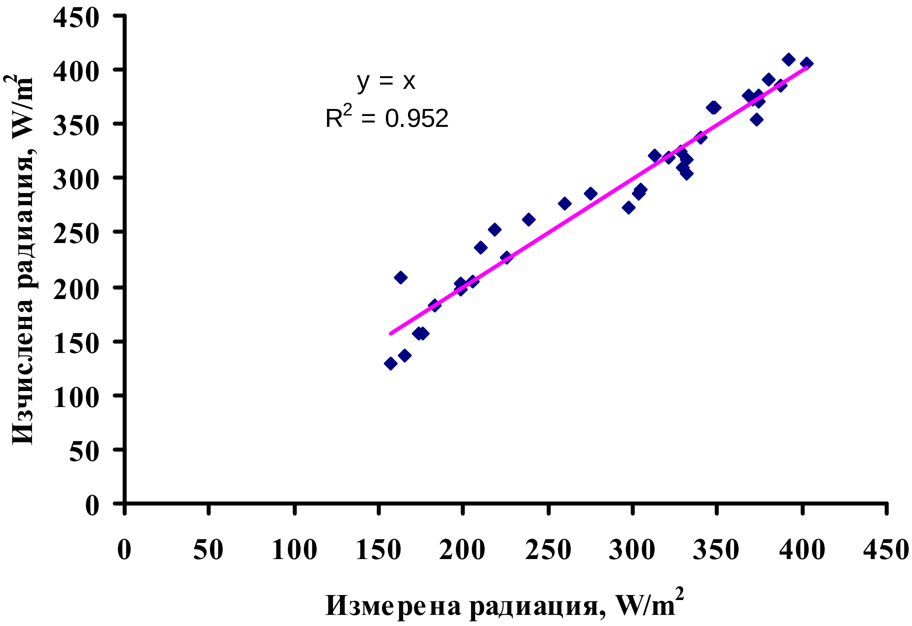
Оценка правдобността на F – статистиката	
FDIST (*F*, v_1, v_2)	$9.36.10^{-20}$
Тест за проверка на хипотези	
Критична стойност TINV	2.04841
Статистическа значимост на:	
$\boldsymbol{a_1}$	**5.70637**
$\boldsymbol{a_2}$	**3.79420**
$\boldsymbol{a_3}$	**3.70585**
$\boldsymbol{a_4}$	0.90957
$\boldsymbol{a_5}$	1.50799
$\boldsymbol{a_6}$**.**	1.82502
b	1.89123

В по-нататъшните изследвания ще използваме само тези три параметъра:

С метода на множествената линейна регресия е изведена друга емпирична формула за дневните суми на сумарната слънчева радиация *S* в София, чрез средната максимална и минимална температури и средната скорост на вятъра:

$$S = 27.733042.T - 23.65592.T0 + 31.74648.\ v - 1193.09, \qquad (4)$$

където: *T* - средна максимална температура, [K]; T_0 - средна минимална температура, [K]; *v* - средна скорост на вятъра, [m/s].



Фиг. 1. Зависимост между измерената и изчислена с регресия сумарна слънчева радиация, W/m^2

На фиг. 1 е дадена зависимостта между измерената и изчислена с регресия сумарна слънчева радиация, W/m^2. Полученият коефициент на детерминация има много голяма стойност:

$$R^2 = 0.952,$$

което показва, че получената емпирична формула (4) много добре определя зависимостта между измерената и изчислена с метода на множествената линейна регресия сумарна слънчева радиация.

ИЗВОДИ

1. Използваните данни са взети от интернет сайт [8]. Те са София, 1961 – 1990 г., който е един от приетите базисни периоди (климатична норма).

2. Направена е ***оценка*** на правдоподобността на високата стойност на F – статистиката. Получава се много малка вероятност за големи F – стойности, получени случайно (FDIST = 9.36.10^{-20}).

3. Направен е и тест за проверка на хипотези. Получените стойности за статистическите значимости на коефициентите a_1, a_2 и a_3, са по-големи от критичната стойност TINV = 2.04841: 5.70637>2.04841; 3.79420>2.04841; 3.70585>2.00758, а останалите статистически зависимости на другите коефициенти са по-малки от критичната стойност TINV = 2.04841.

4. Направената оценка и теста за проверка на хипотези показват, че само три коефициента в регресионното уравнение са полезни за прогнозиране и оценка, а това са коефициентите пред средната максимална и минимална температури и средната скорост на вятъра.

5. От новата емпирична зависимост за дневните суми на сумарната слънчева радиация в София, чрез средната максимална и минимална температури и средната скорост на вятъра, се вижда, че полученият коефициент на детерминация има много голяма стойност: $\boldsymbol{R^2 = 0.952}$. Получената формула много добре определя зависимостта между измерената и изчислена с метода на множествената линейна регресия сумарна слънчева радиация.

ЛИТЕРАТУРА

1. Волфганг, П, 1995. Слънчево електричество. Техника, София, с. 67.
2. Оценка на потенциала на слънчевата радиация в България, http://www.emde-solar.com/топлинна-карта.
3. Димитрова, Р., Ж. Михайлова. Система за обучение по статистика и статистически изчисления с Excel. http://teststat.hit.bg/.
4. Марковски, Б. и Н. Янева, 1990. Статистически анализ на физическия експеримент. Наука и изкуство, София.
5. Kim, S., 1993. Fuzzy Linear Regression with interval – valued Data. Proc. IFSA Word Congress’93, Seoul.
6. Sugeno, M., Kang G. Structure identification of fuzzy model. – Fuzzy Sets and Systems, 28, p. 15 – 53.
7. Множествена регресия, http://www.kirova.org/Mnogestvena-regresia.doc.
8. Вековни данни за времето, http://www.stringmeteo.com/synop/bg-tuti.php.
9. Александров, В., П. Симеонов, В. Казанджиев, Г. Корчев, А. Йотова, 2010. Климатични промени. НИМХ – БАН, София.
10. Alexandrov, V., M. Schneider, E. Koleva, J. Moisselin, 2003. Climate Variability and Change in Bulgaria during the 20th Century. J. Theoretical and Applied Climatology, Vol. 79, pp. 133–149.
11. Продължителност на слънчевото греене,
12. http://www.gaisma.com/en/location/sofia.html.
13. Ефемериди на Слънцето, http://astrocalendar.50webs.com/suncords.html.