جغرافیا و برنامه‌ریزی منطقه‌ای- پیش شماره پاییز و زمستان ۱۳۸۸
صص ۹۵- ۱۰۷

تحلیل رگرسیونی داده‌های اقلیمی با استفاده از ایستگاه‌های انتخابی ایران

دکتر سید احمد بابازاده[۱] / دکتر مجید جاوری[۲]

چکیده

هدف از تحلیل رگرسیونی پیش‌بینی متغیر وابسته از طریق متغیر مستقل می‌باشد. بر این اساس در تحلیل رگرسیونی محقق به دنبال پیش‌بینی است. به همین دلیل برای تحلیل‌های پیشرفته و پیش‌بینی تغییر در متغیر وابسته در صورت تغییر در متغیر یا متغیرهای مستقل باید از روش‌های تحلیل رگرسیونی استفاده کرد. وقتی رابطهٔ متغیرها بررسی می‌شود باید به این نتیجه رسید که متغیرها قابلیت تحلیل رگرسیونی را دارند. در تحلیل رگرسیونی بررسی فرض‌های خطی بودن، بهنجاری، ثابت بودن واریانس داده و مستقل بودن مشاهدات اهمیت زیادی دارد. لذا کاربرد تحلیل رگرسیونی در مطالعات اقلیمی به منظور بررسی رابطهٔ بین متغیرها و پراکنش آنها می‌باشد. مراحل اجرای تحلیل رگرسیونی داده‌های اقلیمی سنجش، تحلیل و پیش‌بینی است. در مرحلهٔ سنجش محقق باید داده‌ها را برای تحلیل رگرسیونی ارزیابی کند، سپس شرایط موجود هر ایستگاه را تحلیل کند و در نهایت رابطه و میزان تغییر و پراکنش متغیرها را پیش‌بینی نماید. با بررسی رابطه و پراکنش متغیرها می‌توان به وضعیت و شرایط حاکم بر یک ایستگاه پی برد. متناسب با روند تحلیلی مدل‌های رگرسیونی از سری‌های دمای ایستگاه‌های یزد و خرم آباد استفاده شده است.

کلیدواژه: *تحلیل رگرسیونی ، متغیر وابسته ، متغیر مستقل ، سنجش ، پیش‌بینی ، بهنجاری و مدل‌های رگرسیونی.*

مقدمه

امروزه کاربرد آمار و مخصوصاً آمار پیشرفته در مطالعات اقلیم شناسی اهمیت زیادی پیدا کرده است . کاربرد شاخص‌ها و مدل‌های آماری قابلیت کاربردی اقلیم شناسی را افزایش داده است و این قابلیت روز به روز گسترش قابل توجهی پیدا کرده است. همچنین کاربرد فنون و نرم افزارهای آماری تحقیقات اقلیمی از اهمیت زیادی برخوردار است. ماهیت موضوعات اقلیمی و داده‌های مورد استفاده در این رشته علمی ضرورت توجه بیشتر به فنون آماری را می طلبد. یکی از فنون بسیار مهم در مطالعات اقلیمی تحلیل‌های رگرسیونی است. در تحلیل رگرسیونی توجه به تغییر در متغیرهای مستقل که مقادیر متغیرهای وابسته را چگونه تحت تأثیر قرار می‌دهد، حائز اهمیت است. وجه تمایز بین

[۱] – دانشیار دانشگاه پیام نور مرکز تهران
[۲] – استادیار دانشگاه پیام نور مرکز تهران M_javari@pnu.ac.ir

متغیرهای مستقل و وابسته همیشه به‌طور کامل روشن نمی‌باشد و وابسته به اهداف تحقیق است. واژهٔ رگرسیون[۳] به معنای برگشت است و نشان می‌دهد که مقدار یک متغیر به متغیر دیگری برمی‌گردد. این واژه اولین بار توسط فرانسیس گالتن [۴] در سال ۱۸۷۷ به‌کار برده شده است (کلانتری،۱۳۸۷: ۱۷۱). با توجه تعداد متغیرها، در مطالعات اقلیمی علاوه بر رگرسیون خطی ساده از رگرسیون چندگانه استفاده می‌شود. برای کاربردی کردن روش‌های مذکور از سری دمای ۳۹ سالهٔ ایستگاه‌های انتخابی ایران استفاده شده است.

کاربرد رگرسیون خطی ساده در تحلیل دما

در رگرسیون خطی ساده هدف این است که با محاسبهٔ عامل های β_0 و β_1، مدل خطی ساده‌ای را به داده‌های برازش[۵] اختصاص داده شود. روشی که برای برازش کردن داده‌ها به‌کار می‌رود روش کمترین مربعات[۶] یا LS می‌باشد(جاوری ، ۱۳۸۴: ۱۲۴). در این روش معادلهٔ خط برازش طوری شکل می‌گیرد که مجموع مربعات توان‌های دوم انحرافات عمودی از خط برازش شده به حداقل برسد. برآورد ضرایب به روش حداقل مربعات صورت می‌گیرد. برای محاسبهٔ ضرایب می‌توان از رابطه‌های زیر استفاده کرد :

رابطه (۱)
$$\hat{\beta}_o = \bar{Y} - \hat{\beta}_1 \bar{X}$$

رابطه (۲)
$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

رابطه (۳)
$$SS_{xx} = \sum X^2 - \frac{(\sum X)^2}{N}$$

رابطه (۴)
$$SS_{xy} = \sum xy - \frac{(\sum X)\times(\sum Y)}{N}$$

رابطه (۵)
$$SS_{yy} = \sum Y^2 - \frac{(\sum Y)^2}{N}$$

β_o عرض از مبدأ، β_1 شیب خط، Y متغیر وابسته و X متغیر مستقل می‌باشد. با توجه به معادلهٔ شمارهٔ (۱) مدل رگرسیون خطی ساده را می‌توان به صورت رابطه (۶) نشان داد:

رابطه (۶)
$$Y = \beta_0 + \beta_1 x + \varepsilon$$

[3] - Regression
[4] - Francis Galton
[5]- Fit
[6] - Least Squares

در معادلهٔ بالا ε عامل خطا می‌باشد. با توجه به مدل بالا معادلهٔ خط برازش شده را می‌توان به‌صورت رابطه (۷) نوشت:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$ رابطه (۷)

مقدار خطا حاصل از تفاضل مقدار اولیه و مقدار پیش‌بینی شده می‌باشد. به طریقی از حاصل تفاضل مقادیر پیش‌بینی شده از مقدار اصلی سری مقدار خطاها یا مانده‌ها[7] به‌دست می‌آید. یعنی:

$$\varepsilon = Y - \hat{Y}$$ رابطه (۸)

در بررسی مدل رگرسیونی خطی ساده فرض می‌شود که:

۱- مانده‌ها دارای توزیع بهنجار یا نرمال هستند.

۲- واریانس سری ثابت است.

۳- متغیرها مستقل از یکدیگر هستند.

برای کاربرد مدل مذکور از سری دمای ایستگاه‌های انتخابی ایران استفاده شده است. در این راستا برای تحلیل سری‌های دمای ایستگاه‌ها متناسب با روند و نیاز تحقیقی از نرم افزارهای Minitab، SPSS، SPLUS، NCSS، STATISTICA 7 و SURFER استفاده شده است. با توجه به نتیجه خروجی تحلیل، در جداول زیر ضرائب مربوطه مشخص شده است. جدول شمارهٔ ۱ ضرایب معادلهٔ خط رگرسیون همراه با انحراف معیار برآوردها؛ ضرائب استاندارد شده و آماره‌های t و P را که برای آزمون فرض‌های آماری به‌کار می‌رود، نشان می‌دهد.

جدول شمارهٔ ۱: ضرائب رگرسیون خطی دمای ایستگاه یزد:

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18.668	.189		98.896	.000
	T	8.563E-03	.008	.169	1.041	.305

متناسب با جدول شمارهٔ ۱ معادلهٔ رگرسیون برابر است با:

$$y = 18.7 + 0.00856\, t$$ رابطه (۹)

[7] - Residual

حال اگر فرض شود که $H_o : \beta_o = 0$ و $H_o : \beta_1 = 0$ آنگاه:

آزمون ۱:

$$H_o : \beta_o = 0$$
$$H_1 : \beta_o \neq 0$$

برای این آزمون از آمارهٔ t استفاده می‌شود که مقدار آن برابر است با:

$$t_{0.05}^{(39-2)} = t_{0.05}^{(37)} \approx 1.684$$

از طرفی ناحیهٔ بحرانی عبارت است از:

$$t = 98.89 > t_{0.05}^{(37)} = 1.684$$

بنابراین فرض صفر رد می‌شود. به عبارت دیگر عرض از مبدأ معادلهٔ رگرسیونی صفر نیست. روش دیگر جهت این آزمون استفاده از آمارهٔ P-Value است. که در آن:

$$\alpha = 0.05 > P_Value = 0$$

پس نتیجه کلی این است که فرض صفر در سطح معنی‌دار پنج درصد رد می‌شود.

آزمون (۲)

$$H_o : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

و مشابه آزمون ۱:

$$|t| = t = 1.o41$$
$$t_{0.05}^{(37)} = 1.684$$

$$t = 1.041 < t_{0.05}^{(37)} = 1.684$$

$$\alpha = 0.05 < P_Value = 0.305$$

بنابراین فرض صفر در سطح معنی‌دار پنج درصد رد نمی‌شود، یعنی معادلهٔ رگرسیونی بدون شیب رگرسیونی است. کل نتیجه آزمون ها را می‌توان به صورت جدول شمارهٔ ۲ نشان داد.

جدول شماره ۲. نتایج کلی آزمون های مورد استفاده برای دمای یزد

Regression Estimation Section	Intercept	Slope
Parameter	B(0)	B(1)
Regression Coefficients	18.6677	0.0086
Lower 95% Confidence Limit	18.2853	-0.0081
Upper 95% Confidence Limit	19.0502	0.0252
Standard Error	0.1888	0.0082
Standardized Coefficient	0.0000	0.1687
T Value98.8957	1.0410	------
Prob Level (T Test)	0.0000	0.3046
Reject H0 (Alpha = 0.0500)	Yes	No
Power (Alpha = 0.0500)	1.0000	0.1736
Regression of Y on X	18.6677	0.0086
Inverse Regression from X on Y	12.8210	0.3009
Orthogonal Regression of Y and X	18.6673	0.0086

رگرسیون و تحلیل واریانس دما ایستگاه یزد:

ایدهٔ اساسی تحلیل واریانس مبتنی بر نمایش میزان کل تغییرات یک مجموعه آماری، به صورت جمع چند عبارت است که بتوان هر یک را به منشاء یا علت خاص تغییرات نسبت داد (صالحی صدقیان، ۱۳۸۱: ۸۹). اگر در رگرسیون، تغییرات Y را نتوان به تغییرات X نسبت داد آنگاه وجود رابطهٔ خطی بین متغیرها مردود است. برای انجام آزمون فرض در خصوص $\mathbf{H_o : \beta_1 = 0}$ می‌توان جدول تحلیل واریانس سری دمای یزد را به صورت جدول شمارهٔ ۳ نشان داد.

جدول شماره ۳. تحلیل واریانس سری دمای یزد

Analysis of Variance Section					
Source DF	Sum of Squares	Mean Square	F-Ratio	Probe Level	Power (5%)
Intercept 1	13841.37	13841.37			
Slope 1	0.3622045	0.3622045	1.0838	0.3046	0.1736
Error 37	12.36575	0.3342096			
Adj. Total	38	12.72796	0.3349463		
Total 39	13854.1				
s = Square Root(0.3342096) = 0.5781086					

با توجه به جدول تحلیل واریانس (ANOVA) رابطهٔ خطی بین متغیرهای Y و X (t) را در سطح معنی‌دار پنج درصد می‌توان با توجه به آزمون ها بررسی نمود.

آزمون فرض‌ها: $\mathbf{H_o : \beta_1 = 0}$ ، $\mathbf{H_1 : \beta_1 \neq 0}$

آمارهٔ آزمون: F = ۱/۰۸۴

ناحیهٔ بحرانی: $\mathbf{F_\alpha(1, n-2) = F_{0.05}(1, 37) = 4.1}$

بنابراین فرض $\mathbf{H_o : \beta_1 = 0}$ رد نمی‌شود. به عبارت دیگر دلایل کافی دال بر رابطهٔ خطی بین متغیرها وجود ندارد. با توجه به جدول شمارهٔ ۳ مجموع مربعات مدل[8] ۰/۳۶۲، مجموع مربعات خطا[9] ۱۲/۳۶۵ و مجموع مربعات کل[10] ۱۲/۲۷۸ می‌باشند. از نکات قابل توجه در جدول خلاصهٔ تحلیل واریانس شاخص ضریب تعیین می‌باشد. جدول شمارهٔ ۴ ، ضریب تعیین را نشان می‌دهد.

ضریب تعیین سری دمای یزد نشان می‌دهد که ۲/۸ درصد از تغییرات Y توسط X توضیح داده می‌شود. یعنی:

$\mathbf{R - Squared = 0.0285\% = 2.85\%}$

[8] - $SS_{Regression}$
[9] - $SS_{Residual}$
[10] - SS_{Total}

جدول شماره ۴. خلاصهٔ تحلیل واریانس سری دمای یزد

Summary Section			
Parameter	**Value**	**Parameter**	**Value**
Dependent Variable	C2	Rows Processed	40
Independent Variable	C1	Rows Used in Estimation	39
Frequency Variable	None	Rows with X Missing	1
Weight Variable	None	Rows with Freq Missing	0
Intercept	18.6677	Rows Prediction Only	0
Slope	0.0086	Sum of Frequencies	39
R-Squared	0.0285	Sum of Weights	39.0000
Correlation	0.1687	Coefficient of Variation	0.0307
Mean Square Error	0.3342096	Square Root of MSE	0.5781086

برآورد فاصلهٔ اطمینان برای متوسط پیش‌بینی از یک مدل خطی ساده

اگر در رابطهٔ‌خطی $\mathbf{Y = \beta_0 + \beta_1 x + \varepsilon}$ مقدار معینی از X^* یعنی x^* (t) در دست باشد، می‌توان مقدار Y را به‌طور متوسط با توجه به رابطهٔ زیر برآورد نمود:

رابطه (۱۰)

$$\mathbf{E(Y/X = x^*)} = b_0 + b_1 x^*$$

برای یک مقدار x^* ، برآورد Y^* برابر است با:

رابطه (۱۱)

$$Y^* = \hat{b}_0 + \hat{b}_1 x^*$$

که در آن، فاصلهٔ اطمینان به صورت زیر می‌باشد:

رابطه (۱۲)

$$\hat{b}_0 + \hat{b}_1 X^* \pm t(n-2,1-\frac{a}{2})' \hat{s}\sqrt{1+\frac{1}{n}+\frac{(X^*-\bar{X})^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}}$$

و با توجه به سری دمای ایستگاه یزد حدود اطمینان تا ۱۰ سال برابر با داده‌های جدول ۵ است.

جدول شماره ۵. حدود اطمینان دمای ایستگاه یزد تا ۱۰ سال

	Fit	SE Fit	95% CI	95% PI
1	18.6763	0.1816	(18.3082; 19.0443)	(17.4485; 19.9041)
2	18.6848	0.1746	(18.3310; 19.0386)	(17.4612; 19.9085)
3	18.6934	0.1677	(18.3536; 19.0332)	(17.4738; 19.9131)
4	18.7020	0.1609	(18.3760; 19.0280)	(17.4861; 19.9179)
5	18.7105	0.1542	(18.3980; 19.0231)	(17.4982; 19.9229)
6	18.7191	0.1477	(18.4197; 19.0185)	(17.5101; 19.9281)
7	18.7277	0.1414	(18.4411; 19.0142)	(17.5218; 19.9336)
8	18.7362	0.1353	(18.4620; 19.0104)	(17.5332; 19.9392)
9	18.7448	0.1294	(18.4825; 19.0071)	(17.5444; 19.9451)

10	18.7533	0.1238	(18.5024; 19.0043)	(17.5554; 19.9513)

تحلیل مانده‌های رگرسیونی

وقتی به بررسی رابطهٔ متغیرها پرداخته شد معمولاً باید محقّق به این نتیجه برسد که سری‌ها قابلیت تحلیل رگرسیونی دارند. محقّق باید به این نتیجه برسد که آیا رابطه‌ای خطی بین متغیرها وجود دارد و آیا توزیع متغیر وابسته بهنجار است و یا واریانس آن برای مقادیر مختلف متغیر مستقل ثابت است. بر این اساس بخش مهم تحلیل رگرسیونی بررسی فرض‌های خطی بودن، بهنجار بودن، ثابت بودن واریانس و مستقل بودن مشاهدات است (تیموری یانسری ، ۱۳۸۶: ۵۹۵). لذا اولین قدم در تحلیل رگرسیونی سنجش سری‌های مورد بررسی است. یعنی محقّق باید ابتدا سری‌ها را بسنجد که آیا شرایط تحلیل و کاربرد رگرسیونی را دارند یا نه (نیرومند، ۱۳۸۱: ۱۱۶). اگر فرض پذیرفته شد در آن مرحله است که محقّق می‌تواند از مدل‌های تحلیل رگرسیونی استفاده کند. اگر فرض‌های تحلیل رگرسیونی صادق باشد، مانده‌های تحلیل باید مشخصات زیر را داشته باشند:

۱- باید توزیع آنها بهنجار یا نرمال باشد.

۲- باید واریانس آنها برای تمام مقادیر متغیر مستقل ثابت باشد.

۳- هنگام نمایش مانده‌ها در مقابل مقادیر پیش‌بینی شده نباید الگویی مشاهده شود.

۴- باید مانده‌ها تقریباً مستقل از یکدیگر باشند.

برای نمایش فرض‌های مذکور از سری دمای یزد استفاده شده است.

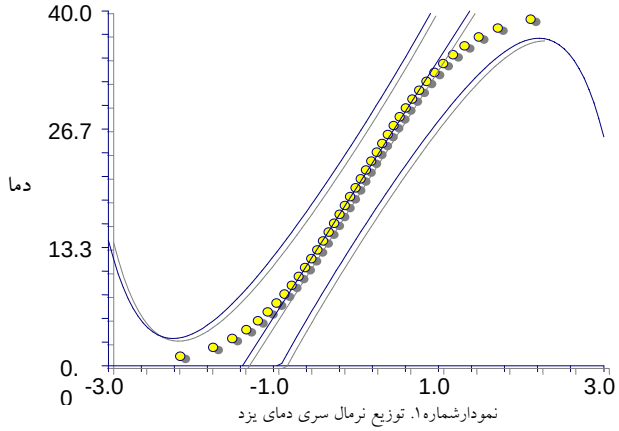
پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

پاسخ به فرض‌ها:

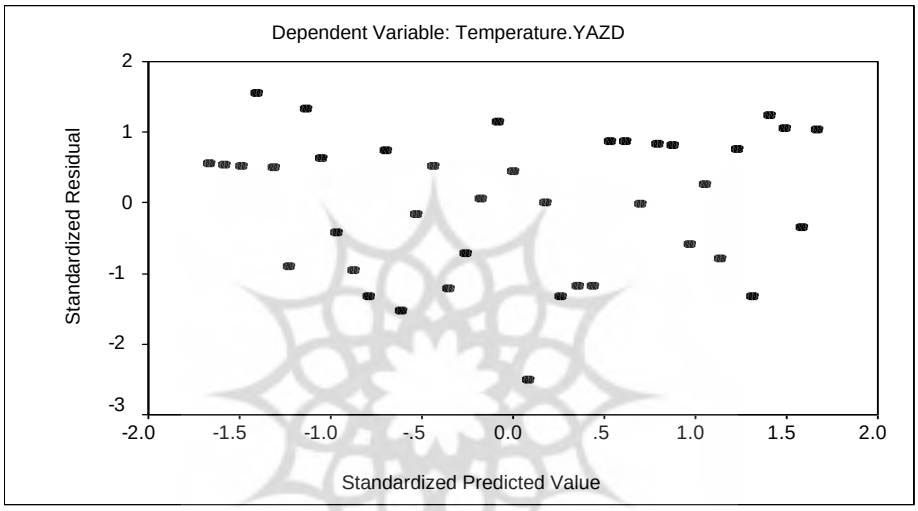
با توجه به نمودار شماره ۱، سری مانده‌های دمای یزد، وضعیت بهنجاری را نشان می‌دهد.

برای پاسخ به فرض دوم و سوم، باید واریانس آنها برای تمام مقادیر متغیر مستقل ثابت باشد و هنگام نمایش مانده‌ها در مقابل مقادیر پیش‌بینی‌شده نباید الگویی مشاهده شود. با توجه به نمودارهای شمارهٔ ۲ و۳، چون سری مانده‌ها الگوی مشخصی را در نمودار نشان نمی‌دهند بر این اساس واریانس مانده‌های سری برای متغیر مستقل تقریباً ثابت می‌باشد.

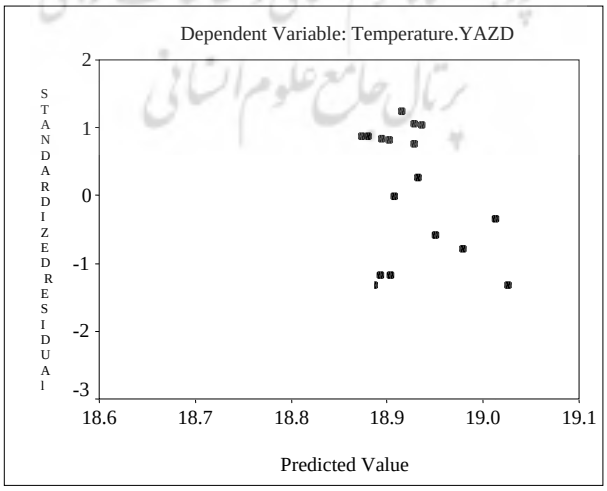
برای پاسخ به فرض چهارم ، باید مانده‌ها تقریبا" مستقل از یکدیگر باشند.



نمودارشماره۱. توزیع نرمال سری دمای یزد



نمودار شماره ۲. نمودار مانده‌های استاندارد شدهٔ دمای یزد



نمودار شماره ۳. نمودار مانده‌های استاندارد شده با مقادیر پیش‌بینی شده دمای یزد

برای بررسی استقلال مانده‌ها از همدیگر می‌توان از آزمون دوربین _ واتسون [11] استفاده کرد، اگر مقدار به‌دست آمده بین ۱/۵ تا ۲/۵ باشد می‌توان تقریباً به استقلال مانده‌ها از همدیگر پی برد. با توجه نتیجهٔ کاربرد آزمون دوربین _ واتسون می‌توان به مستقل بودن مانده‌ها پی برد. جدول شمارهٔ ۶ نتیجه آزمون فوق را نشان می‌دهد.

جدول شماره ۶. آزمون دوربین _ واتسون برای استقلال مانده‌های دمای یزد

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	مقدار آزمون دوربین- واتسون
1	.169	.028	.002	.5781	1.948

کاربرد رگرسیون چندگانه در تحلیل‌های اقلیمی

در تحلیل‌های اقلیمی برای پیش‌بینی مقدار متغیر وابسته، به بیش از یک متغیر مستقل نیاز می‌باشد. برای مثال وقتی پژوهشگری وضعیت اقلیمی یک منطقه یا یک کشور و یا یک ایستگاه را بررسی می‌کند لازم است رابطهٔ متغیرهای متعددی را بررسی کند. بر این اساس یکی از فنون آماری مهم، کاربرد رگرسیون چندگانه می‌باشد. در تحلیل رگرسیون چندگانه به جای یک عرض از مبدأ و ضریب زاویه‌ای دارای یک مقدار ثابت، دارای چند ضریب می‌باشد که هر یک متعلق به یکی از چند متغیر مستقل می‌باشد (جاوری،۱۳۸۴: ۱۳۶). این ضرایب به عنوان ضرایب رگرسیون جزئی[12] نامیده می‌شوند. ضرایب رگرسیون جزئی در نمونه برآوردی از ضرایب نامعلوم جامعه است. همچون رگرسیون خطی ساده، در رگرسیون چندگانه نیز فرض‌های زیر باید مورد توجه قرار گیرند:

۱- رابطهٔ بین متغیر وابسته و متغیرهای مستقل باید خطی باشد .

۲- برای هر یک از ترکیب‌های مقادیر متغیرهای مستقل، توزیع متغیر وابسته نرمال یا بهنجار باشد.

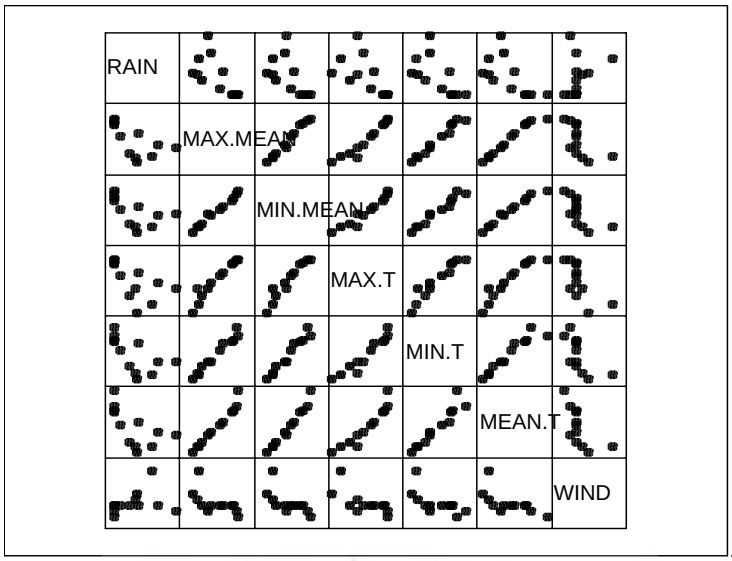
۳- واریانس مقادیر آن ثابت باشد.

قبل از آنکه ضرایب برآورد شوند باید مطمئن شد که متغیرهای مستقل به‌طور خطی با متغیر وابسته رابطه دارند. برای عملی نمودن رگرسیون خطی چندگانه از متغیرهای متعددی که در ایجاد بارش ایستگاه خرم آباد (رطوبت نسبی، دماهای حداقل مطلق، حداکثر مطلق، معدل حداکثر، معدل حداقل و سرعت باد) مؤثر می‌باشند، استفاده شده است. از ماتریس نمودارهای پراکنش بین متغیرهای مستقل و وابسته برای بررسی رابطهٔ متغیرها استفاده شده است. برای کاربرد مدل رگرسیون چندگانه می‌توان از معادلهٔ زیر استفاده کرد (رضایی ،۱۳۷۷: ۱۱۶):

[11] - Durbin_Watson
[12] - Partial Regression Coefficients

$$Y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n + e$$ ربطه (۱۳)

جدول شماره ۷. ماتریس نمودارهای پراکنش متغیرهای ایجاد بارش خرم



آباد

با توجه به جدول ماتریس نمودارهای پراکنش متغیرها، رابطهٔ خطی بین بعضی متغیرها وجود دارد. برای تبیین بهتر لازم است براساس رابطه شمارهٔ ۱۳ به تحلیل رگرسیون چندگانه پرداخت.

y = 203.73 - 16.76 x1 + 37.38 x2 - 0.46 x3 - 4.74 x4 - 7.726 x5 + 7.70 x6

بررسی مناسب بودن مدل

وقتی یک مدل رگرسیون اجرا می‌شود، اولین نکته بررسی مناسب بودن مدل است. برای این منظور لازم است به تحلیل رگرسیون چندگانه توجه نمود (صالحی صدقیان ،۱۳۸۱: ۱۴۲). جدول‌های زیر تحلیل رگرسیون چندگانه متغیرهای ایجاد بارش خرم آباد را نشان می‌دهند.

جدول شماره ۸. خلاصهٔ آمارهٔ مدل رگرسیون چندگانه

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.933	.870	.715	23.5965

جدول شماره ۹. تحلیل واریانس متغیرهای بارش خرم آباد

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18670.651	6	3111.775	5.589	.039
	Residual	2783.978	5	556.796		
	Total	21454.629	11			

جدول شماره ۱۰. ضرایب رگرسیون چندگانهٔ متغیرهای بارش خرم آباد

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	203.730	89.916		2.266	.073
	MAX.MEAN	-16.764	6.222	-4.369	-2.694	.043
	MIN.MEAN	37.378	9.211	6.785	4.058	.010
	MAX.T	-.461	2.915	-.107	-.158	.881
	MIN.T	-4.741	5.761	-.799	-.823	.448
	MEAN.T	-7.726	3.984	-1.935	-1.939	.110
	WIND	7.699	6.022	.352	1.278	.257

با توجه به جدول خلاصهٔ آماره‌های تحلیل، ضریب تعیین تحلیل نشان می‌دهد که۸۷ درصد از پراکندگی مشاهده شده ایجاد بارش خرم آباد توسط شش متغیر مستقل توجیه می‌شود و به طریقی متغیرهای مستقل رابطهٔ خطی با متغیر وابسته ایجاد بارش خرم آباد را نشان می‌دهد. از طرفی مقدار همبستگی متغیرها که برابر است با ۰/۹۳ است نشان می‌دهد که مدل رگرسیون خطی مذکور به خوبی می‌توان برای پیش‌بینی استفاده نمود. با توجه به جدول شمارهٔ ۸ آمارهٔ F ($F=\frac{MST}{MSE}=\frac{3111.775}{556.796}=5.59$) فرض صفر رد می‌شود و نتیجه گرفته می‌شود که بین متغیر وابسته (مقدار بارش خرم آباد) و متغیرهای مستقل (دماها ، رطوبت و سرعت باد) رابطه وجود دارد.

هنگامی که فقط یک متغیر مستقل در مدل رگرسیون وجود دارد، تفسیر ضریب رگرسیون آن به سادگی انجام می‌گیرد و اگر فرضیه صفر رد شد، می‌توان نتیجه گرفت که رابطهٔ خطی بین دو متغیر مستقل و وابسته وجود دارد. تفسیر این ضریب در رگرسیون چندگانه پیچیده‌تر است. اگر ضرایب رگرسیون جزئی برای متغیرها به‌دست آوریم، با بررسی ماتریسی ضرایب امکان بررسی رابطه‌ها وجود دارد. چون از یک طرف ضرایب همبستگی را نشان می‌دهد از طرفی دیگر سطوح معنی‌داری مشاهدات را نشان می‌دهد. جدول شمارهٔ ۱۱ ضرایب رگرسیونی جزئی متغیرهای مؤثر در ایجاد بارش خرم آباد را نشان می‌دهد.

ضریب همبستگی بین میزان بارش و رطوبت نسبی ۰/۷۵۸ است. بر این اساس سطح معنی‌داری مشاهده شده می‌تواند این فرضیه را رد کند که بین دو متغیر رابطهٔ خطی وجود دارد. از طرفی مشاهده می‌شود که متغیرهای مستقل با هم رابطهٔ خوبی را نشان می‌دهند.

جدول شماره ۱۱. ضرایب همبستگی جزیی متغیرهای ایجاد بارش خرم آباد

		RAIN	MAX.MEAN	MIN.MEAN	MAX.T	MIN.T	MEAN.T	WIND	HUMIDITY
RAIN	Pearson Correlat	1.000	-.643	-.554	-.606	-.610	-.601	.298	.758
	Sig. (1-tailed)	.	.012	.031	.018	.018	.019	.173	.002
MAX.MEA	Pearson Correlat	-.643	1.000	.989	.965	.979	.972	-.611	-.936
	Sig. (1-tailed)	.012	.	.000	.000	.000	.000	.017	.000
MIN.MEA	Pearson Correlat	-.554	.989	1.000	.962	.975	.982	-.674	-.901
	Sig. (1-tailed)	.031	.000	.	.000	.000	.000	.008	.000
MAX.T	Pearson Correlat	-.606	.965	.962	1.000	.957	.934	-.671	-.888
	Sig. (1-tailed)	.018	.000	.000	.	.000	.000	.008	.000
MIN.T	Pearson Correlat	-.610	.979	.975	.957	1.000	.939	-.655	-.911
	Sig. (1-tailed)	.018	.000	.000	.000	.	.000	.010	.000
MEAN.T	Pearson Correlat	-.601	.972	.982	.934	.939	1.000	-.647	-.911
	Sig. (1-tailed)	.019	.000	.000	.000	.000	.	.012	.000
WIND	Pearson Correlat	.298	-.611	-.674	-.671	-.655	-.647	1.000	.476
	Sig. (1-tailed)	.173	.017	.008	.008	.010	.012	.	.059
HUMIDIT	Pearson Correlat	.758	-.936	-.901	-.888	-.911	-.911	.476	1.000
	Sig. (1-tailed)	.002	.000	.000	.000	.000	.000	.059	.

قدرت رابطهٔ خطی بین متغیرهای مستقل توسط شاخص تولرانس[۱۳] اندازه‌گیری می‌شود. برای هر متغیر مستقل تولرانس، نسبتی از پراکندگی آن متغیر است که توسط روابط خطی آن متغیر با سایر متغیرهای مستقل موجود در مدل توجیه نمی‌شود (نیرومند،۱۳۸۱: ۱۴۰). بر این اساس برای محاسبه شاخص چند هم خطی[۱۴] در نرم افزار مینی تب از عامل افزایش واریانس[۱۵] (VIF) می‌توان استفاده کرد. جدول شماره ۱۲ شاخص مذکور را نشان می‌دهد. مقدار عامل افزایش واریانس باید بین ۵ تا ۱۰ باشد، اگر بیشتر از این حد باشد، امکان ضعف برآورد ضریب رگرسیون در اثر چند هم خطی وجود دارد. با توجه به این که تولرانس یک نسبت است، مقدار آن بین صفر و یک می‌باشد. مقدار نزدیک به یک به این معنی است که در یک متغیر مستقل بخش کوچکی از پراکنش آن با سایر متغیرهای مستقل توجیه می‌کند و مقدار نزدیک به صفر نشان می‌دهد که یک متغیر تقریباً یک ترکیب خطی از سایر متغیرهای مستقل است.

جدول شماره ۱۲. آماره‌های تولرانس و عامل افزایش واریانس سری‌های مؤثر در بارش خرم آباد

Model		Collinearity Statistics	
		Tolerance	VIF
1	MAX.MEAN	.010	101.325
	MIN.MEAN	.009	107.726
	MAX.T	.056	17.825
	MIN.T	.028	36.346
	MEAN.T	.026	38.365
	WIND	.342	2.925

[13]- Tolerance
[14] - Multicollinearity
[15] - Variance Inflation Factor

نتیجه‌گیری

در تحلیل‌های اقلیمی بررسی پراکنش و رابطهٔ بین متغیرها از اهمیت زیادی برخوردار است. بر این اساس کاربرد تحلیل‌های رگرسیونی در مطالعات اقلیمی ضرورتی اجتناب ناپذیر می‌باشد. متناسب با مدل‌های تحلیل رگرسیونی (ساده و چندگانه) مراحل تحلیل و عملیاتی کردن آنها این امکان را به محقق می‌دهد که بهتر بتواند وضعیت حاکم بر هر ایستگاه یا هر منطقه را بررسی و پیش‌بینی نماید. کاربرد مراحل سنجش، تحلیل و پیش‌بینی در تحلیل‌های رگرسیونی برای مطالعات و تحلیل‌های اقلیمی ضروری است. رعایت اولویت‌های عملیاتی مدل‌های رگرسیونی متناسب با فرایند تحقیق در مطالعات اقلیمی از اهمیت زیادی برخوردار است. بسته به موضوع تحقیق، تحلیل‌های رگرسیونی این امکان را به محقق می‌دهد که متغیرهای متعددی را از هر ایستگاه در فرایند تحقیق مورد استفاده قرار دهد. بر این اساس و با توجه به ماهیت مطالعات اقلیمی، کاربرد تحلیل‌های رگرسیونی ضروری و دارای اهمیت زیاد خواهد شد.

فهرست منابع

۱- آمار غیر پارامتری برای علوم رفتاری (۱۳۷۲) ترجمه: یوسف کریمی؛ انتشارات دانشگاه علامه طباطبایی.

۲- آمار کاربردی، همراه با برنامه‌های کامپیوتری (۱۳۷۴) ترجمه: محمد صادق تهرانیان و ابوالقاسم بزرگ نیا؛ انتشارات جهاد دانشگاهی مشهد.

۳- آمار و احتمال برای مهندسی معدن، جلد دوم (۱۳۸۶) ترجمه: مهدی تیموری یانسری؛ انتشارات دانشگاه امیرکبیر.

۴- اساس روش‌های آماری (۱۳۶۰) ترجمه: محمد صادق تهرانیان؛ انتشارات دانشگاه مشهد.

۵- افشین نیا، منوچهر (۱۳۷۲) روش‌های آماری و کاربرد آن در علوم؛ انتشارات اتا.

۶- بهبودیان، جواد (۱۳۷۸) آمار ناپارامتری؛ انتشارات دانشگاه شیراز.

۷- جاوری، مجید (۱۳۸۴) مقدمه ای بر روش‌های تحقیق در علوم انسانی با تأکید بر جغرافیا؛ انتشارات طلور.

۸- جلیلی، خشنود، خلیل (۱۳۸۰)، تحلیل آماری؛ انتشارات سرا.

۹- حسینی، یعقوب (۱۳۸۲) آمار ناپارامتری (روش تحقیق و نرم افزار SPSS 10)؛ انتشارات دانشگاه علامه طباطبایی.

۱۰- رگرسیون خطی کاربردی (۱۱۳۷۴) ترجمه: حسینعلی نیرومند؛ انتشارات دانشگاه مشهد.

۱۱- صالحی صدقیان، جمشید؛ابراهیمی، ایرج (۱۳۸۱) تحلیل آماری پیشرفته؛ نشر هوای تازه.

۱۲- فرشادفر، عزت اله (۱۳۸۰) اصول و روش‌های آماری چند متغیره؛ انتشارات دانشگاه رازی.

۱۳- منصورفر، کریم (۱۳۸۷) روش‌های پیشرفته آماری؛ انتشارات دانشگاه تهران.

14- Brien , O .L (1992) Introducing Quantitative Geography , Routledge , Landon-New York.

15- Bryman ,A& Cramer , d (1990) Quantitative Data Analysis for Social Scientists ,Routledge ,Chapman and hall.

16- Hays,W.L(1982) Statistics for the Social Sciences , New York ,Holt, Rinehart and Winston ,Inc.

17- Howell ,D.C(1989) Fundamental Statistics For the Behavioral Sciences, Pws-KENT.

18- Neter ,J & Wasserman ,W (1974) Applied Linear Statistical Models. Homewood .

19- Winer ,B.J(1971) Statistical Principles in Experimental Design, McGraw-Hill , New York. ■