جمع			12/95	13/119		24/39

$$m_3 = \frac{M^4}{M_2^2} = \frac{24.39}{13.119^2} = .142$$

چون مقدار ضریب کشیدگی بیشتر از صفر است، کشیدگی سری دمای مسجد سلیمان مثبت است. لذا با توجه به ضریب کشیدگی بدست آمده برای دمای سالانهٔ مسجد سلیمان نشان می دهد که توزیع متغیرهای دما در نقطهٔ اوج خوابیده می باشند.

محاسبهٔ کشیدگی دمای مسجد سلیمان با استفاده از نرم افزار SPSS :

Descriptive Statistics

	Skewness		*Kurtosis*	
	Statistic	*Std. Error*	*Statistic*	*Std. Error*
میانگین دمای سالانهٔ مسجد سلیمان	**-.333**	**0.512**	**0.154**	**0.992**

محاسبهٔ ضریب تغییرات متغیرهای اقلیمی :

یکی از معیارها و شاخص های پراکندگی بسیار مهم که در بیشتر مطالعات اقلیمی مورد استفاده قرار می گیرد ضریب تغییرات[1] یا پراکندگی نسبی واریانس[2] می باشد. این شاخص برای مقایسهٔ دو متغیر زمانی که میانگین یکسانی ندارند ، بیشتر بکار می رود. این شاخص چون با واحد اندازه گیری مختلف ارتباطی ندارد بنابراین می توان از آن برای مقایسهٔ پراکندگی سریها با واحدهای مختلف استفاده نمود. ضریب

1 - Coefficient of Variation
2 - Relative Variance

تغییرات از نسبت انحراف معیار به یک حد متوسط که معمولا این حد متوسط میانگین است (پاره ای موارد محققین از میانه یا نما استفاده می کنند) بدست می آید . برای محاسبهٔ ضریب تغییرات متغیرها می توان از رابطهٔ شمارهٔ استفاده نمود:

$$C.V = \frac{S}{\bar{X}} \times 100$$

موقعیت شمارهٔ :

اگر سری دمای 44 سالهٔ (1995-1952) ایستگاه های اصفهان و یزد را در دست داشته باشیم، ضریب تغییرات دوسری راباهم مقایسه کنید؟

$$C.V = \frac{.66}{22.986} \times 100 = 2.87$$ **برای سری اصفهان**

$$C.V = \frac{.583}{18.86} \times 100 = 3.09$$ **برای سری یزد**

با توجه به ضرائب تغییرات، سری دمای یزد میزان پراکندگی آن نسبت به دمای اصفهان بیشتر است .

محاسبهٔ ضریب تغییرات سری های اقلیمی با استفاده از نرم افزار NCSS:

Skewness and Kurtosis Section of temperature_yazd

			Coefficient	*Coefficient*
Parameter	*Skewness*	*Kurtosis*	*Fisher's g1*	*Fisher's g2*	*of Variation*
Value	*-0.429703378021337*		*2.30270244951713*	*-0.445020526745701*	*-0.633722857249912*	*3.09190345628066E-02*	*2.48564593301435E-02*
Std Error	*0.261801115787813*		*0.455219910457506*	*2.71691161399019E-03*			

Skewness and Kurtosis Section of temperature_esfahan

				Coefficient
Coefficient				

Parameter	Skewness	Kurtosis	Fisher's g1	Fisher's g2
of Variation				
Value	-0.54593197026115	3.53420985860109	-0.568016439835464	0.780780018824064
2.87489080202494E-02	2.07803790412486E-02			
Std Error	0.323274923688276	0.763797940739688	3.71104035508723E-03	

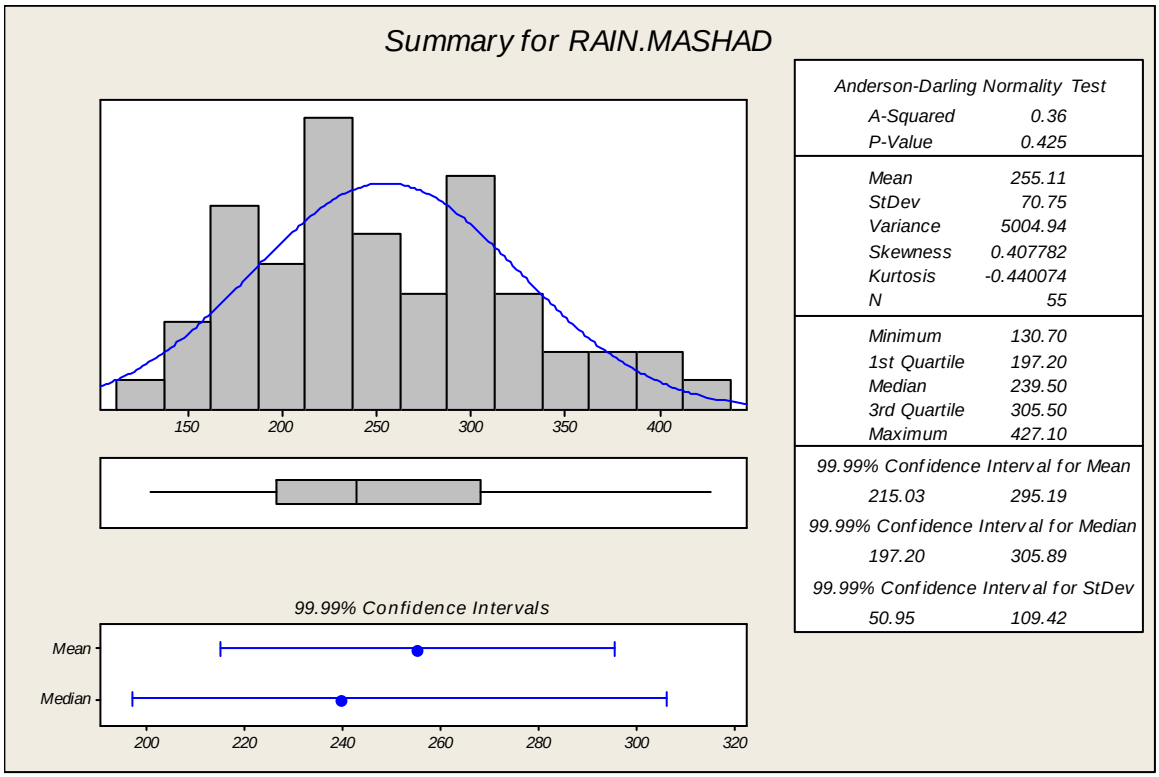
محاسبهٔ ضریب تغییرات سری های اقلیمی با استفاده از نرم افزار Minitab:

Descriptive Statistics: yazd; esfahan

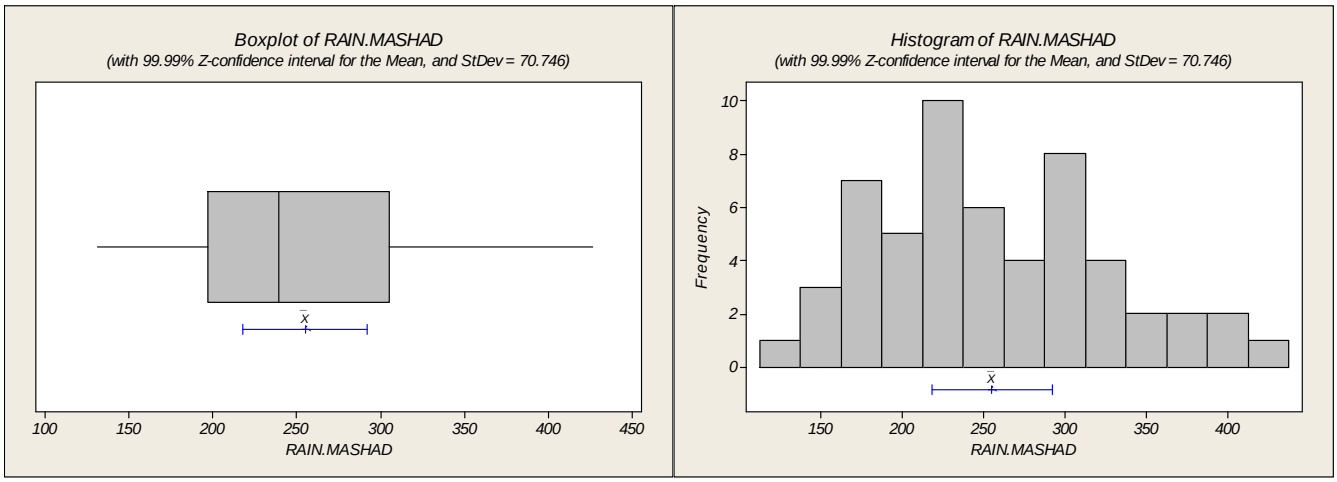
Variable	Mean	SE Mean	StDev	CoefVar	Skewness	Kurtosis
yazd	18.860	0.0879	0.583	3.09	-0.45	-0.63
esfahan	22.986	0.106	0.661	2.87	-0.57	0.78

گزاره ای بر کاربرد آماره های توصیفی در تحلیل های اقلیمی:

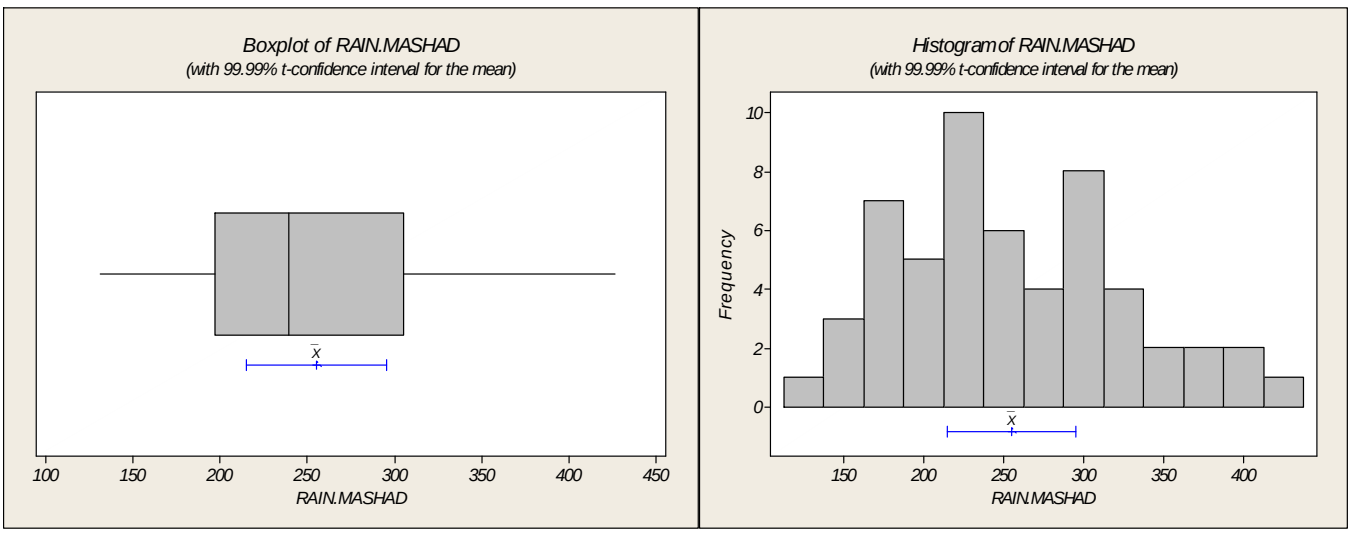
برای کاربرد و ارائهٔ گزاره های آماره های توصیفی ، از سری بارش ایستگاه های مختلف ایران استفاده شده است . بر این اساس بارش 56 سالهٔ(1951 2005) ایستگاه مشهد در قالب گزارهٔ مورد نظر مورد استفاده قرار گرفت :



با توجه به نمودار آماره های توصیفی بارش مشهد مشخص است که با توجه به ضریب آزمون بهنجاری دارین ـاندرسون (0/425) می توان استدلال نمود که فرض تحقیق ما در خصوص بهنجاری سری مشهد پذیرفته می شود . (مقدار ضریب از مقدار 0/05 بیشتر است). ضریب تعیین مقدار 0/36 را نشان می دهد . میانگین بارش 255/11 میلی متر وانحراف معیار 70/75 می باشد . میزان چولگی 0/408 نشانگر کشیدگی توزیع می باشد و مقدار عدم تقارن را نشان می دهد ومقدار مثبت آن میل به چولگی به راست را نشان می دهد. همچنین مقدار 0/44 کشیدگی پهن شدگی توزیع را نشان می دهد و مقادیر چارک اول (197/2) و چارک سوم (305/5) و میانهٔ آن (239/5) می باشد. از نکات دیگر تحلیل آماره ها مقادیر فاصلهٔ اطمینان سری بارش مشهد است . بطوری که فاصلهٔ اطمینان برای میانگین با احتمال 99/99 درصد بین 215/03 تا 295/19 میلی متر می باشد. یعنی با احتمال 99/99 درصد بارش ایستگاه مشهد از میانگین دامنه ای بین 215/03 تا 295/19 میلی متر دارد . همچنین فاصلهٔ اطمینان برای میانه با احتمال 99/99 بین 192/2 تا 305/19 میلی متر می باشد و فاصلهٔ اطمینان برای انحراف معیار با احتمال 99/99 درصد بین 50/95 تا 109/42 میلی متر می باشد. یعنی میزان انحراف از میانگین برای بارش مشهد بین 50/95 تا 109/42 میلی متر دامنه دارد. برای تحلیل از نمودارهای دیگر می توان استفاده کرد.

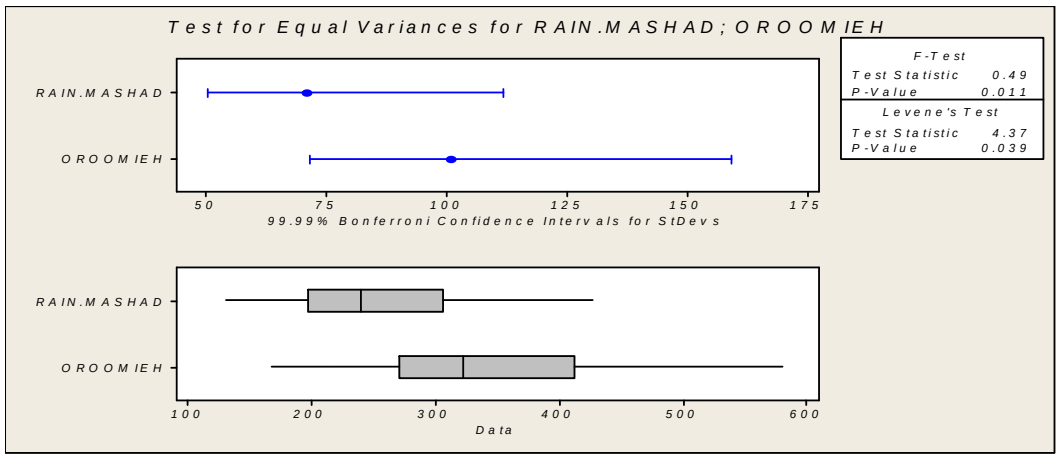


با توجه به فاصلهٔ اطمینان نمرهٔ معیار سری بارش مشهد با احتمال 99/99 درصد با درنظر گرفتن میزان انحراف معیار سری ، بارش سری بین 217/992 تا 292/219 میلی متر می باشد. همچنین میزان میانگین سری براساس نمرهٔ معیار 255/105 میلی متر می باشد و انحراف از میانگین 9/539 میلی متر می باشد. یعنی $SE.Mean = \frac{S}{\sqrt{N}} = \frac{70.746}{\sqrt{55}} = 9.539$ **. همچنین در این رابطه می توان به تحلیل آمارهٔ دیگری پرداخت .**

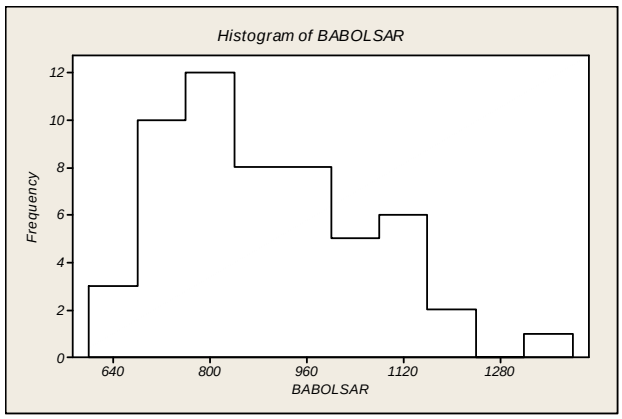


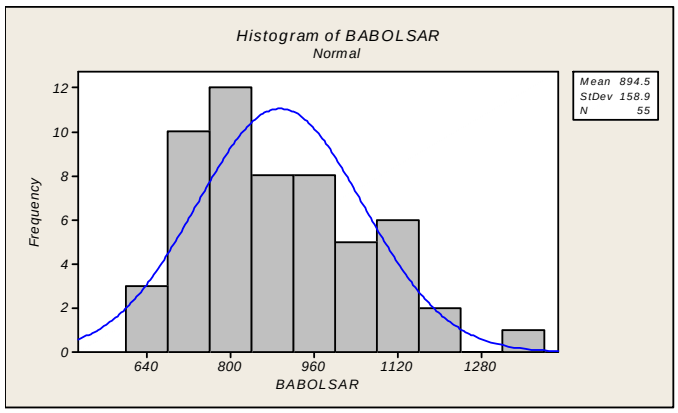
با توجه به فاصلهٔ اطمینان آزمون t سری بارش مشهد با احتمال 99/99 درصد با درنظر گرفتن میزان میانگین سری ، بارش سری بین 295/186 تا 215/025 میلی متر می باشد. همچنین میزان میانگین سری 255/105 میلی متر می باشد و انحراف از میانگین 9/539 میلی متر می باشد. از نکاتی که باید در تحلیل آماره های توصیفی در تحقیقات اقلیمی به آن توجه نمود سنجش برابری واریانس سری ها می

باشد. در این مورد از آزمون های **Levene** و **Bartlett** می توان استفاده کرد. آزمون اول نسبت به آزمون دوم حساسیت کمتری نسبت به سنجش بهنجاری سری ها دارد. بر این اساس برای سنجش سری ها از بارش ایستگاه های مشهد و ارومیه استفاده شد.

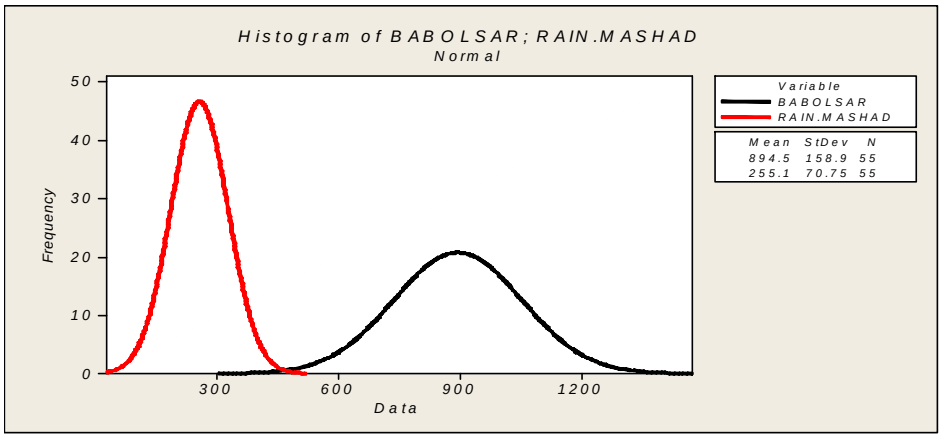


با توجه به آزمون **F** نشان می دهد سری بارش ایستگاه ها نرمال یا بهنجار می باشند . آزمون لیون غیر نرمال بودن سری را مورد سنجش قرار می دهد. به کاربرد آزمون را در قسمت های بعدی کتاب اشاره می شود. یکی دیگر از رویکرد های تحلیل آماره های توصیفی استفاده از نمودار هسیتوگرام است که شکل ، مکان، تمایل مرکزی و گستردگی توزیع را نشان می دهد. برای کاربرد آن از سری **55** سالهٔ **(2005-1951)** بابلسر استفاده شده است .

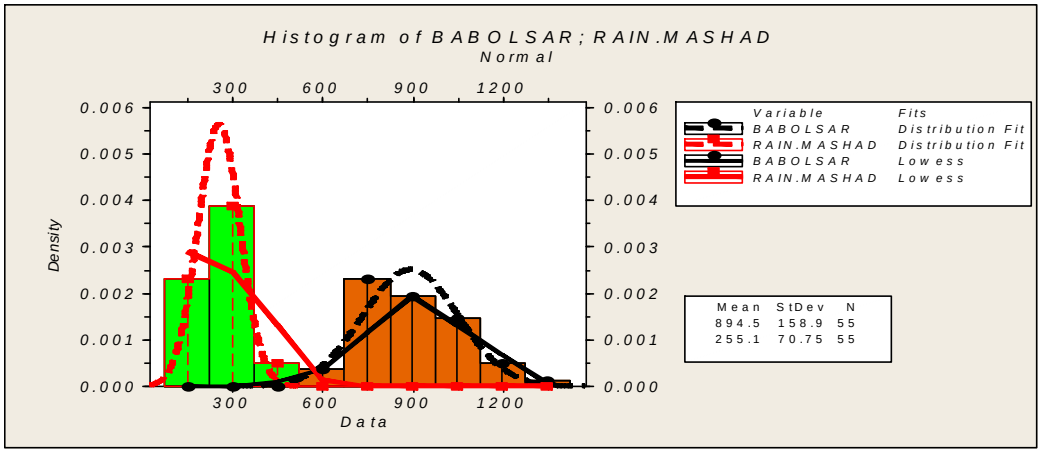




نمودارها فراوانی سری بارش و وضعیت برازش سری بارش بابلسر را نشان می دهند. نمودار دیگری که در برازش سری اهمیت دارد، نمونه ای از سری هیستوگرام است. یعنی :



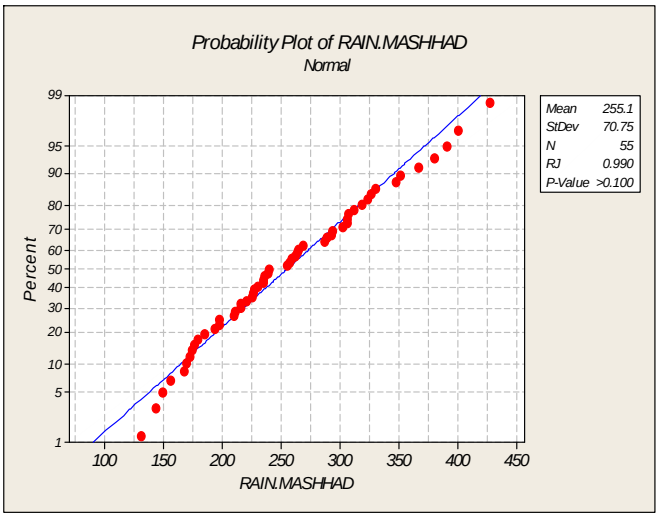
نمودار بالا میزان کشیدگی سری های بارش مشهد و بابلسر را نشان می دهد. بطوری که میزان کشیدگی بارش مشهد بیشتر از سری بابلسر است . هر چند که هر دو سری کشیدگی منفی را نشان می دهند. یعنی سری های بارش مشهد و بابلسر از اوج وضعیت نرمال فاصله دارند.یعنی :

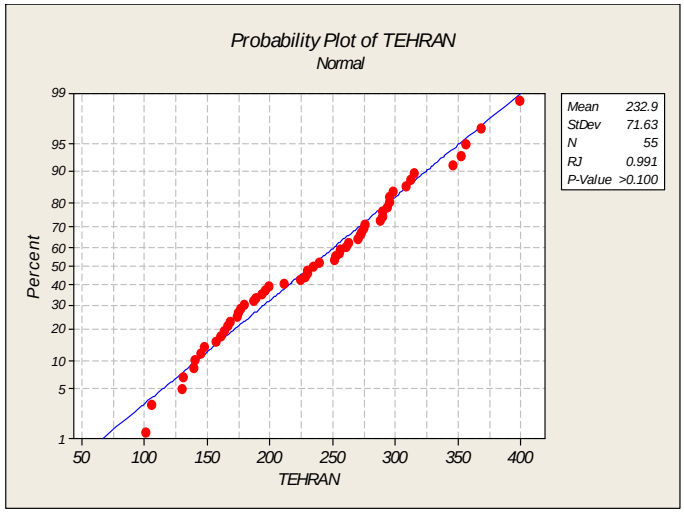


نمودار بالا وضعیت سری بارش ایستگاه های مشهد و بابلسر را نسبت به حالت برازش بهنجاری سری نشان می دهد.

بررسی بهنجاری سری های اقلیمی :

برای بررسی بهنجاری های سری های اقلیمی از آزمون های مختلفی می توان استفاده کرد . نمونه ای از روش ها، استفاده از نمودار بهنجاری است . قابلیت های مختلف نرم افزاری فراوانی وجود دارد که می تواند به این فرآیند توجه نمود . بر این اساس برای بررسی از نرم افزار های **minitab** و **NCSS** استفاده می شود. برای انجام این فرآیند از سری بارش ایستگاه های مختلف استفاده می شود.



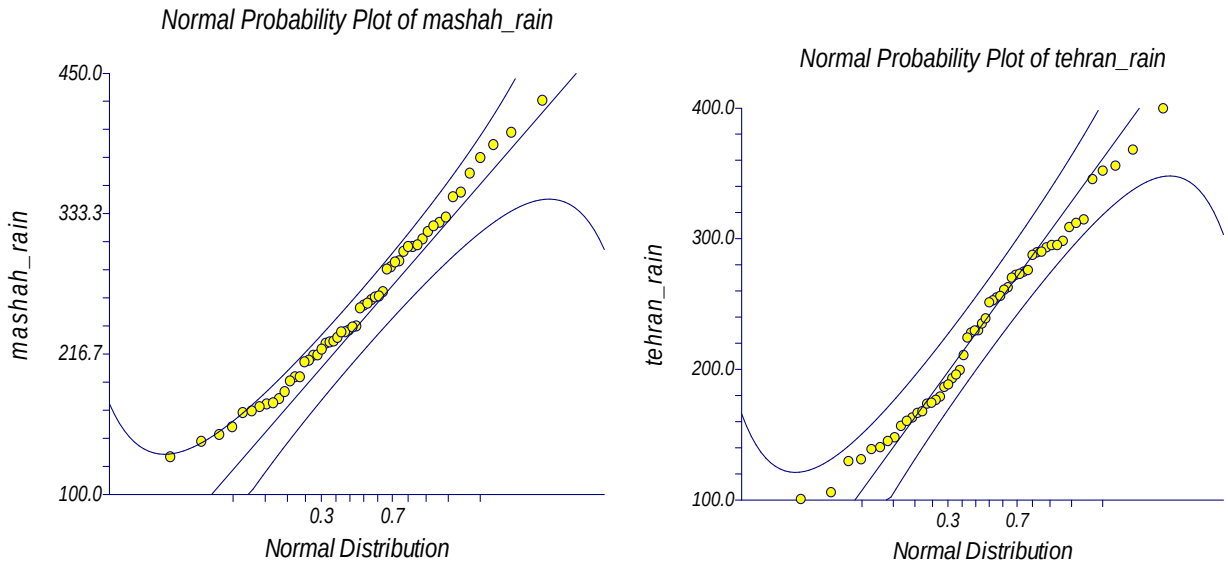


با توجه به سری های 55 سالهٔ (1951-2005) بارش مشهد و تهران از آزمون آریانا- جونیر[1] استفاده شده است . آین آزمون یک ضریب همبستگی بین سری ها و توزیع نرمال را نشان می دهد و هرچه ضریب به یک نزدیک تر باشد داده ها به توزیع نرمال نزدیک می شود. بر این اساس سری های سالانهٔ بارش مشهد و تهران به شرایط بهنجاری نزدیک هستند یا به عبارتی سری ها بهنجار می باشند. آزمون ریانا- جونیر را می توان از رابطهٔ زیر استفاده کرد:

$$RJ = \frac{\sum x_i b_i}{\sqrt{V(n-1)\sum b_i^2}}$$

X_i مشاهدهٔ مرتب شده، b_i امتیاز نرمال برای داده های مرتب شده و V واریانس نمونه است. سنجش سری ها با استفاده از نرم افزار NCSS :

[1] - Ryan-Joiner



با توجه به نمودارها ، وضعیت بهنجاری سری های بارش مشهد و تهران مشخص می باشد . در این نمودارها میل به بهنجاری سری تهران نسبت به سری مشهد بیشتر است .

کنترل آماره های توصیفی در تحلیل متغیرهای اقلیمی :

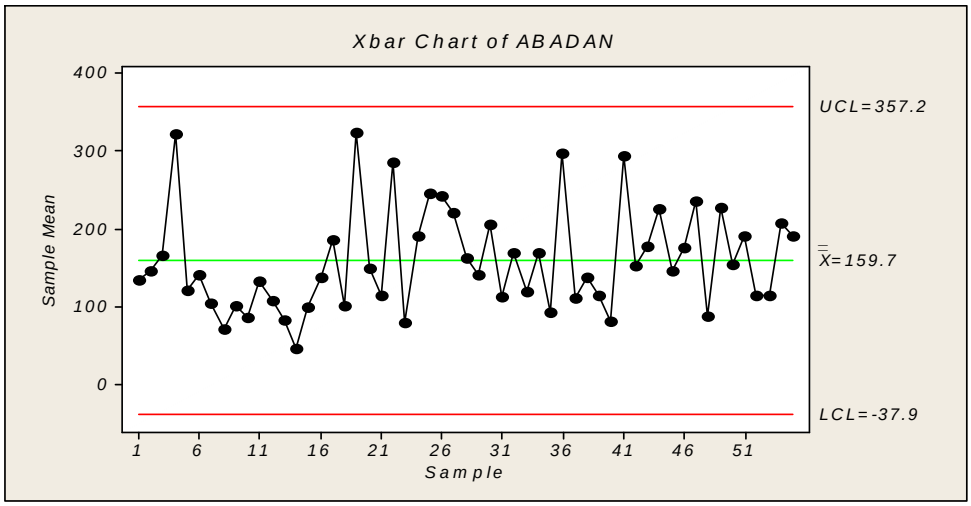
برای تحلیل کنترل سری های اقلیمی می توان آنها را با استفاده از آماره های توصیفی بررسی و سنجش قرار داد . شرایط وضعیت های اقلیمی که تحت تأثیر تغییرات ماهیتی خود می باشند، قابل کنترل هستند و شرایط اقلیمی که در قالب زمان بر اثر عوامل خارجی اقلیمی تغییر نمایند و بیشتر قالب تصادفی دارند قابل کنترل نمی باشند. بر این اساس تغییرات آماره های توصیفی در طول زمان تغییر می کنند و بیشتر این تغییرات قابل کنترل نمی باشند . لذا نمایش سریهای اقلیمی در برابر آماره های توصیفی از اهمیت زیادی برخوردار است .نمودارهای قابل کنترل دارای خط مرکز[1](CL) ، حدود کنترل[2](LCL و UCL) ونقاط ترسیمی هستند. برای تبیین کنترل آماره های توصیفی به موقعیت های زیر اشاره شده است .

موقعیت شمارهٔ :

بر این اساس سری 55 سالهٔ (2005-1951) بارش ایستگاه آبادان تحت بررسی قرار گرفت.

---

[1] - *Center Line*
[2] - *Upper Control Limit & Lower Contorl Limit*



سری بارش ایستگاه آبادان در برابر میانگین (159/7) نشان داده شده است . از طرفی حدود اطمینان بارش آبادان بین(37/9= $LCL = \overline{X} - 3S$ تا 357/2 = $UCL = \overline{X} + 3S$در نوسان می باشد. همچنین می توان میزان اریب انحراف معیار[1] را برآورد کرد. یعنی :

$$P_S = \frac{S_p}{C_4(d)}$$

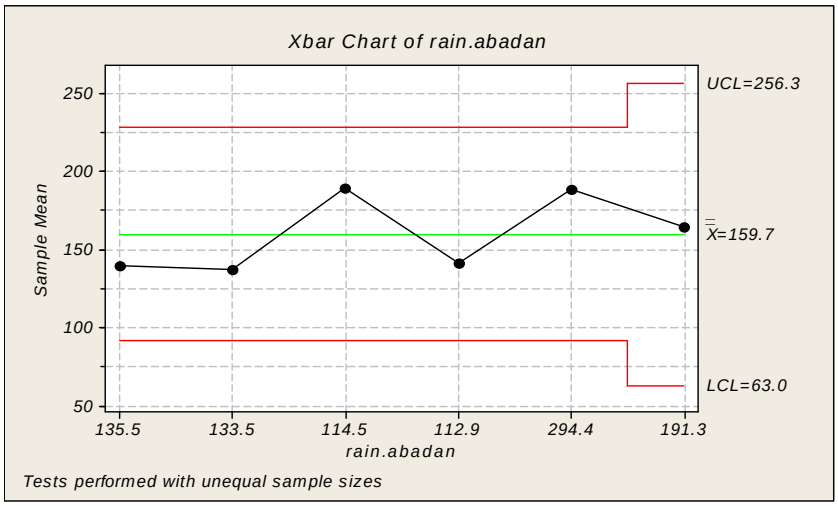
$$S_p = \sqrt{\frac{\sum_i \sum_j (X_{ij} - \overline{X}_i)^2}{\sum_i (n_i - 1)}}$$

درجه آزادی S_p نیز به صورت $d = \sum (n_i - 1)$ خواهد بود. مقدار ضریب $C_4(d+1)$ با استفادهٔ فرمول زیر قابل محاسبه است:

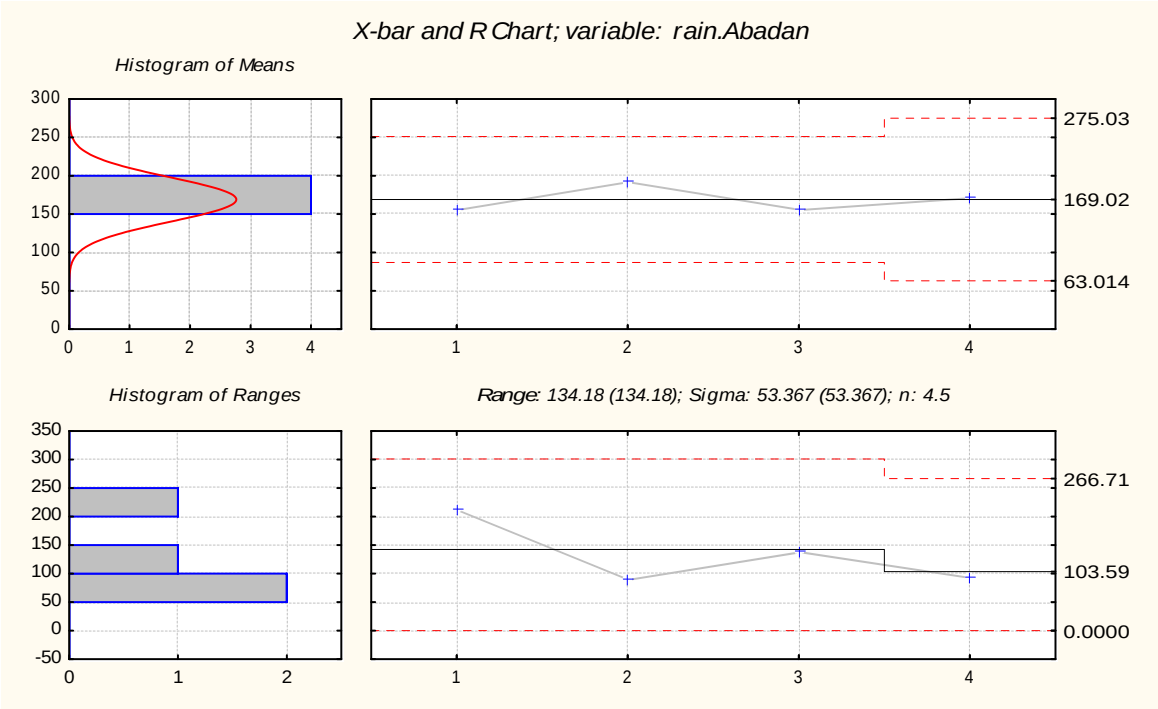
$$C_4(d+1) = \sqrt{\frac{2}{d}} \times \frac{\Gamma\frac{(d+1)}{2}}{\Gamma\left(\frac{d}{2}\right)}$$

لذا بارش آبادان را به صورت فرآیند کنترل می توان بررسی نمود:

[1]- *Pooled Standard Deviation*



کنترل بارش آبادان با استفاده از نرم افزار Stattistica7 :



برای بررسی میزان میانگین در برابر انحراف معیار یکی دیگر از شاخص های کنترل آماره های توصیفی متغیرهای اقلیمی می باشد. سری بارش آبادان در برابر انحراف معیار را با نمودار های شمارهٔ نشان داده شده است. برای محاسبهٔ میانگین انحراف معیار می توان از رابطهٔ زیر استفاده کرد:

$$\bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_i$$

اگر سری مورد نظر توزیع بهنجار باشد رابطهٔ میانگین و انحراف معیار سری را می توان به صورت رابطهٔ زیر در نظر گرفت:

$$\sigma_{\bar{X}} = \frac{\bar{S}}{C_4\sqrt{n}}$$

حدود اطمینان کنترل با استفاده از روابط زیر قابل محاسبه است :

$$LCL = \bar{X} - 3\frac{\bar{S}}{C_4\sqrt{n}}$$

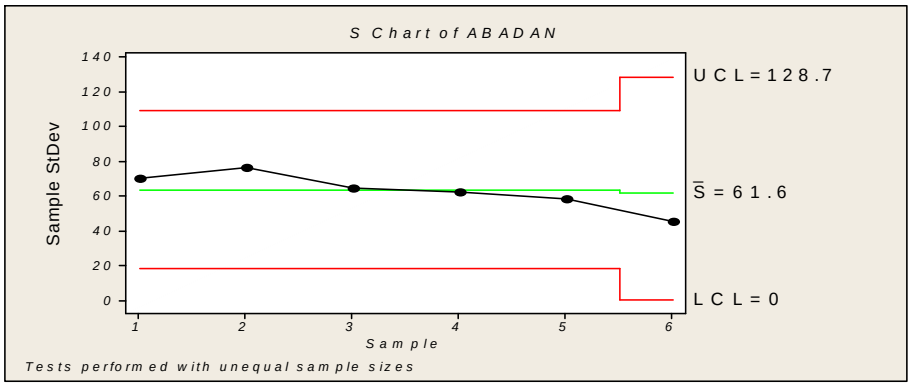
$$UCL = \bar{X} + 3\frac{\bar{S}}{C_4\sqrt{n}}$$

C_4مقدار ثابت است که فقط به اندازهٔ نمونه nبستگی دارد و مقادیر آن به ازای nهای متفاوت از جدول معین قابل استخراج است . خط مرکز نمودار S برابر با $\bar{S}$ است . برای محاسبهٔ حدود اطمینان کنترل نمودار ترسیمی لازم است تخمین دقیقی از انحراف معیار داشت. ثابت می شود انحراف معیار سری برابر با $\bar{X}\sqrt{1-C_4^2}$ است . لذا حدود اطمینان کنترل نمودار انحراف معیار برابر خواهد بود:

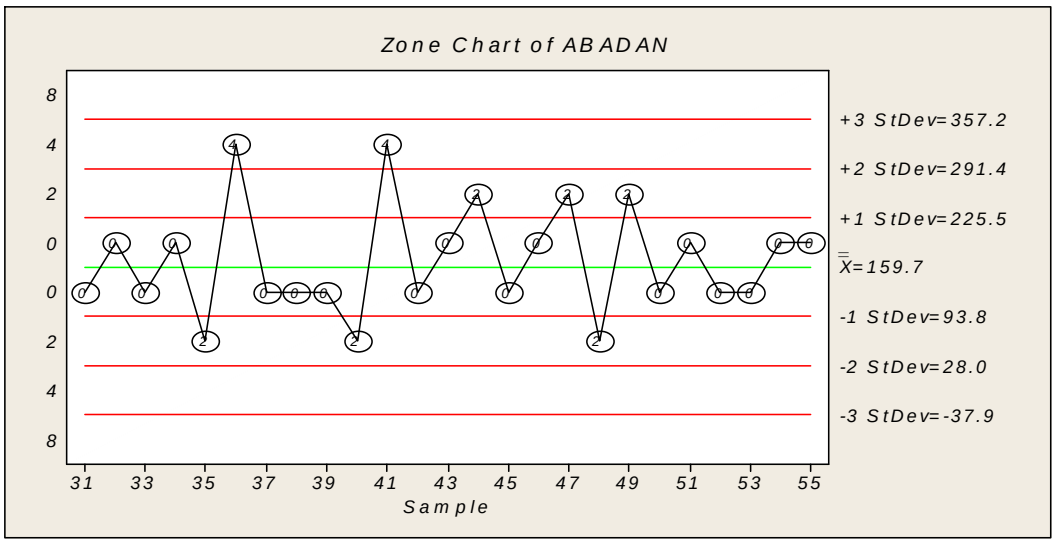
$$LCL = \bar{S} - 3\bar{X}\sqrt{1-C_4^2}$$

$$UCL = \bar{S} + 3\bar{X}\sqrt{1-C_4^2}$$

با توجه به مطالب بالا نمودار بارش آبادان در نمودار مذکور می توان بررسی نمود .آن نمودار سری بارش آبادان را با میانگین انحراف معیار 61/6 با حدود اطمینان حداکثری 128/7 نشان می دهد:



نمودار بالا سری بارش آبادان در برابر انحراف معیار و سری بارش آبادان را با حدود اطمینان حداکثری نشان می دهد.



سری بارش ایستگاه آبادان در برابر تنوع مقدار انحراف معیار مشخص شده است. بیشترین نوسان سری در این سطح منفی و مثبت یک با حدود از 93/8 تا 225/5 میلی متر در نوسان می باشد.

کنترل سری اقلیمی با استفاده از میانگین با دامنهٔ متحرک (I-MR) :

خط مرکزی این نمودار از میانگین نمونه های مستقل سری بدست می آید . برای تخمین حدود اطمینان کنترل باید محاسبهٔ دقیقی از تغییرپذیری سری بدست آورد . در این مرحله تغییرپذیری از طریق دامنهٔ متحرک دو سری متوالی بدست می آید .تخمین دامنهٔ متحرک از رابطهٔ زیر بدست می آید:

$$MR_i = \left| X_i - X_{i-1} \right|$$

اگر توزیع شرایط بهنجاری داشته باشد رابطهٔ بین میانگین و دامنهٔ متحرک برابر است با:

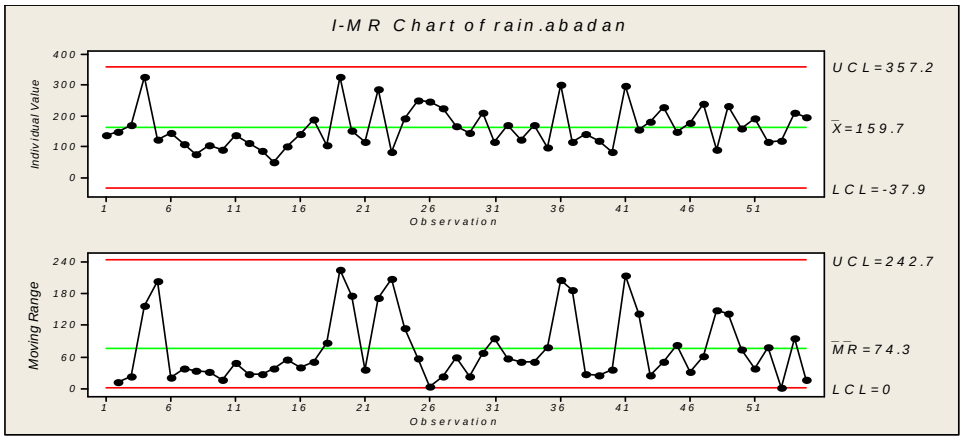
$$\hat{\sigma} = \frac{\overline{MR}}{D_2}$$

D_2 مقدارثابتی است که با توجه به تعداد سری (N) از جدول معین بدست می آید. حدود اطمینان کنترل سری برابر است با:

$$UCL = \bar{X} + 3\frac{\overline{MR}}{D_2}$$

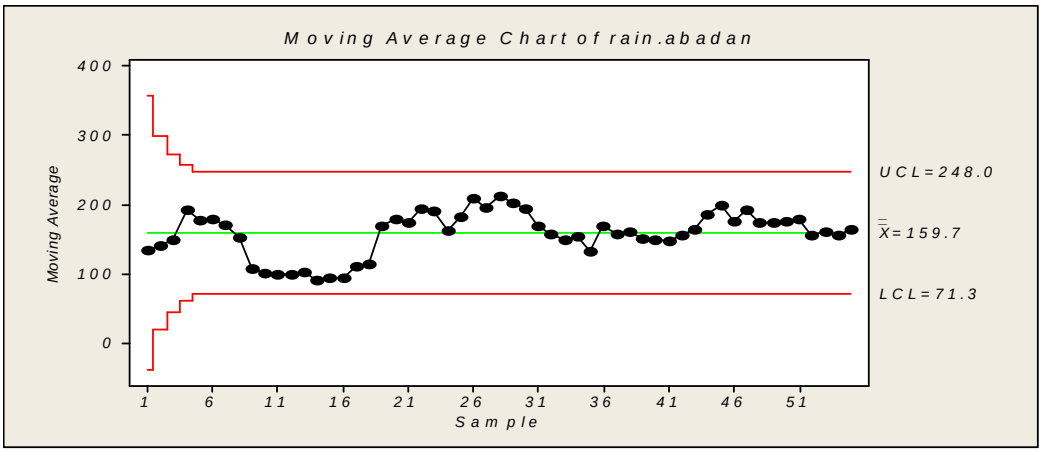
$$LCL = \bar{X} - 3\frac{\overline{MR}}{D_2}$$

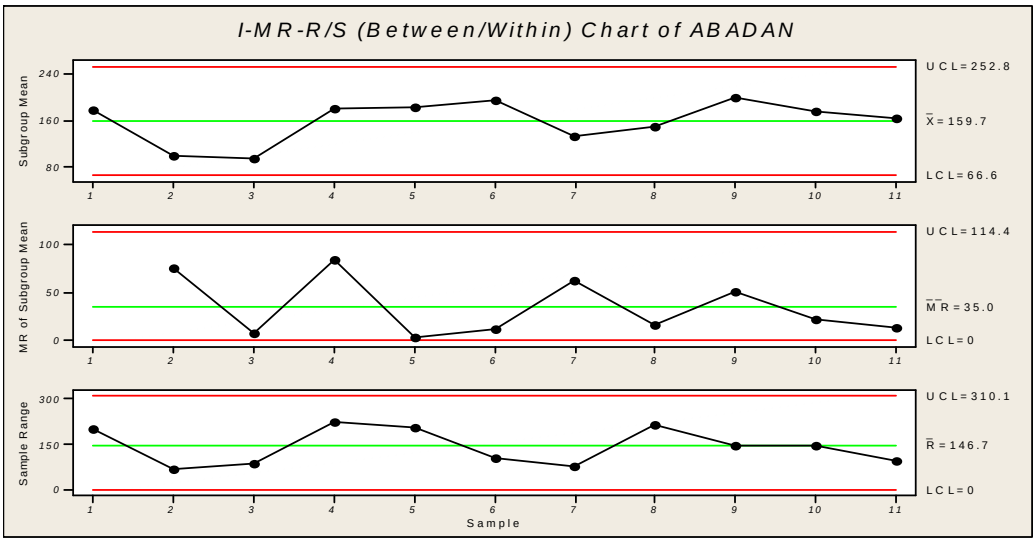
سری بارش آبادان با نمودار مذکور مورد بررسی قرار گرفت :



کنترل سری بارش آبادان به صورت دامنهٔ میانگین حداکثر اطمینان بارش را 242/7 میلی متر نشان می دهد و میانگین آن برابر با 74/3 میلی متر می باشد.

کنترل سری بارش آبادان با استفاده از میانگین متحرک (5 ساله) حدود اطمینان بین71/3 تا 248 در نوسان را نشان می دهد . از طرفی هرچه رتبهٔ میانگین متحرک افزایش یابد میل سری به میانگین نزدیک تر می شود.





سری بارش آبادان در برابر میانگین، میانگین متحرک و دامنهٔ تغییرات را نشان می دهد.

کنترل میانگین متحرک متغیرهای اقلیمی :

کنترل میانگین متحرک متغیرهای اقلیمی براساس میانگین متحرک سرهای متوالی بدست می آید . اگر سری طبقه بندی نشده باشد از روش های میانگین دامنهٔ متحرک، میانه دامنهٔ متحرک استفاده می شود و در صورتی که حجم نمونه ها بزرگتر از یک باشد می توان از روش میانگین متحرک استفاده کرد . برای تحلیل و استفاده از شاخص کنترلی می توان از روابط زیر استفاده کرد:

$$MA_i = \frac{\bar{X}_1 + \ldots \bar{X}_n}{n}$$

اگر n بزرگتر از w(طول دوره هایی که میانگین متحرک بر اساس آن محاسبه می گردد) باشد از رابطهٔ زیر می توان استفاده کرد:

$$MA_i = \frac{\bar{X}_1 + \ldots \bar{X}_{i-w+1}}{w}$$

حدود اطمینان سری اگر n مساوی یا کوچکتر از w باشد :

$$LCL = \bar{X} - K(S/n)\sqrt{1/n_i + \ldots 1/n_1}$$

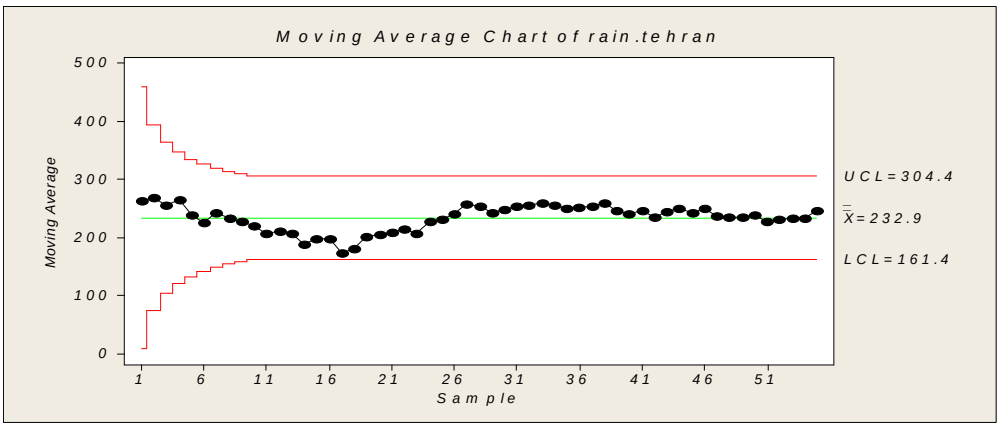
$$UCL = \bar{X} + K(S/n)\sqrt{1/n_i + \ldots 1/n_1}$$

حدود اطمینان سری اگر n بزرگتر از w باشد:

$$LCL = \bar{X} - K(S/w)\sqrt{1/n_i + \ldots 1/n_{i-w+1}}$$

$$UCL = \bar{X} + K(S/w)\sqrt{1/n_i + \ldots 1/n_{i-w+1}}$$

بر این اساس سری بارش 55 سالهٔ (1951- 2005) تهران مورد بررسی قرار گرفت :



در سری مذکور طول دوره 10 گرفته شده است . براین اساس حدود اطمینان بالایی 304/4 میلی متر و حداقل حد اطمینان 161/4 میلی متر برآورد زده شده است .

کنترل سری های اقلیمی با استفاده از میانگین متحرک وزنی :

نمودار کنترل میانگین متحرک وزنی[1] **(EWMA)** یکی از روش های مهم به منظور شناسایی تغییرات جزیی در هر مجموعه ای از متغیرهای اقلیمی است .آمارهٔ این نمودار را می توان به صورت زیر نشان داد:

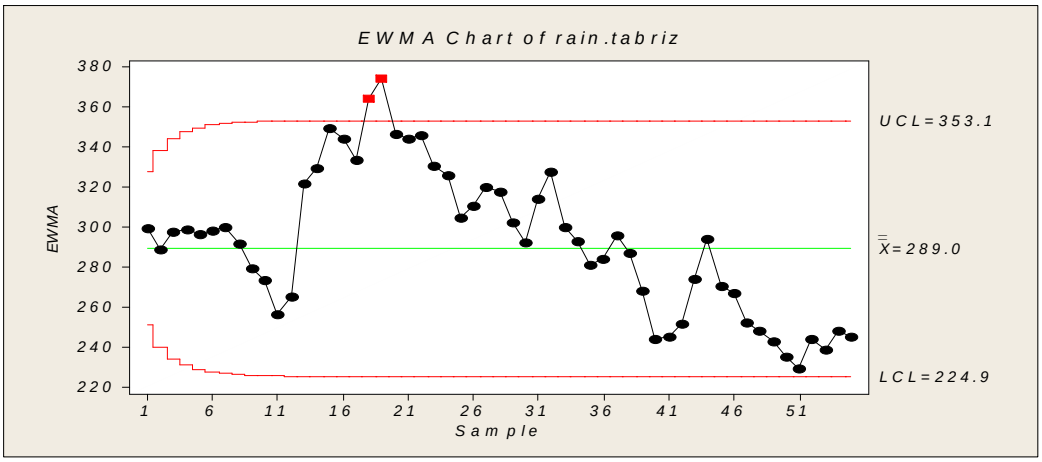
$$\bar{X}_{t=1} = (1-\lambda)\bar{X}_t + \lambda x_t$$

$\bar{X}_t$ میانگین متحرک وزنی در زمان **t** و λ ضریب وزنی برای متغیرها بطوری که $0 < \lambda < 1$ است.حدود اطمینان سری را می توان برای سری های بیش از **30** سال از روابط زیر بدست آورد:

$$LCL = \bar{X} - k(S)\sqrt{[\lambda/(2-\lambda)]}$$

$$UCL = \bar{X} + k(S)\sqrt{[\lambda/(2-\lambda)]}$$

برای تبیین بهتر سری بارش **55** سالهٔ **(1951 2005)** ایستگاه تبریز مورد بررسی قرار گرفت :



کنترل جمع تجمعی متغیرهای اقلیمی :

[1] - Exponentially Weighed Moving Average

یکی از نمودار های مهمی که می توان متغیرهای در حال کنترل را بررسی نمود نمودار جمع تجمعی[1] (CUSUM) می باشد . در این نمودار هم می توان به صورت دوطرفه ، هم یک طرفه سری های کنترل شده را نشان داد. حدود اطمینان سری را می توان از روابط زیر بدست آورد:

$$LCL = -C\frac{S}{\sqrt{N}}$$

$$UCL = C\frac{S}{\sqrt{N}}$$

در نمودار دو طرفه مقدار جمع تجمعی برای هر داده از رابطهٔ زیر بدست می آید:

$$C_t = C_{t-1} + (\bar{X}_i - T)$$

شیب ماسک از فرمول زیر بدست می آید:

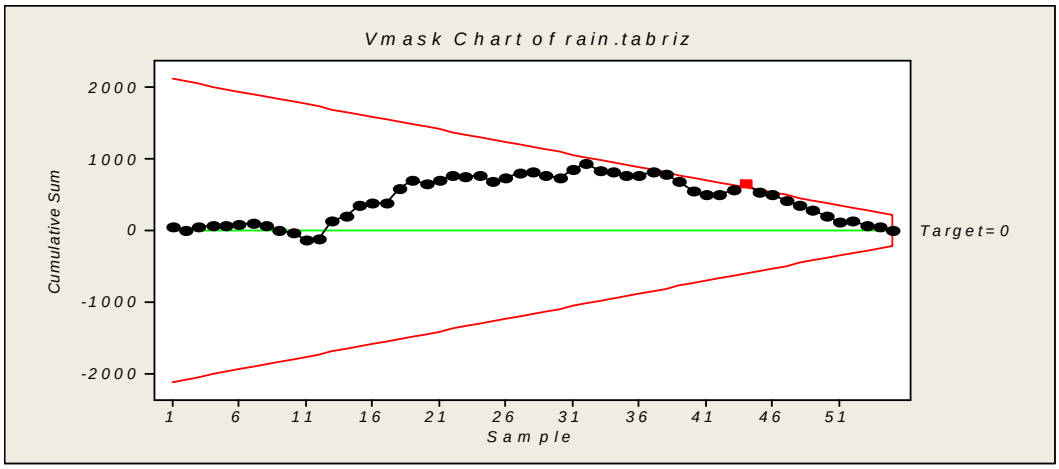
$$k\frac{S}{\sqrt{N}}$$

عرض ماسک از فرمول زیر بدست می آید:

$$2h\frac{S}{\sqrt{N}}$$

برای تبیین بهتر موضوع سری بارش تبریز مورد بررسی قرار گرفت :

[1] - Cumulative Sum

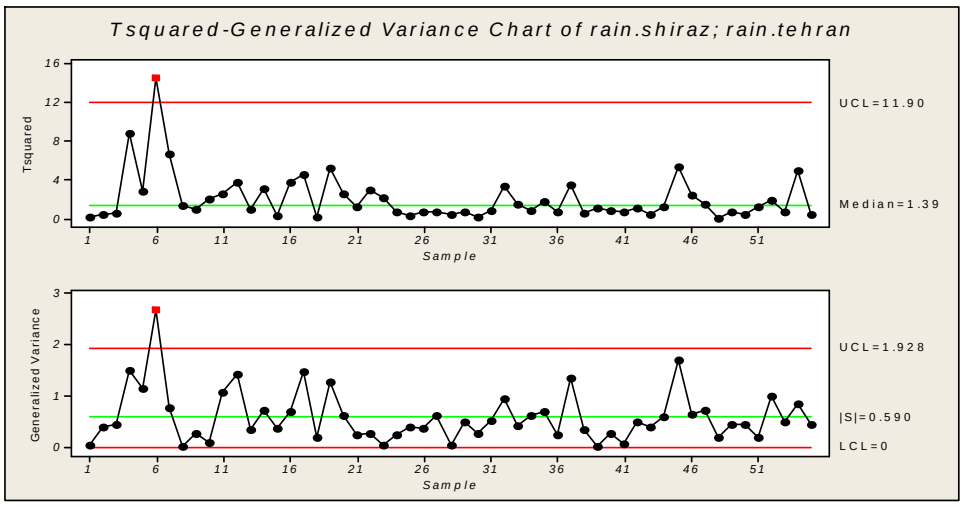


کنترل متغیرهای اقلیمی با استفاده از کنترل چند متغیره:

اکثرا" درتحلیل های اقلیمی مناطق مختلف معمولا از چند متغیر استفاده می شود. یکی از روش های کنترل ،استفاده از شیوه های کنترل چند متغیره می باشد. در استفاده از کنترل چند متغیرهٔ آماری از آمارهٔ T^2 می توان استفاده نمود. دراین آماره از میانگین نمونه ها یا سری ها، واریانس نمونه ها ، کواریانس نمونه ها استفاده می شود.مقدار آمارهٔ T^2 را می توان از رابطهٔ زیربه دست آورد:

$$T^2 = n(\bar{X} - \bar{\bar{X}})' S^{-1} (\bar{X} - \bar{\bar{X}})$$

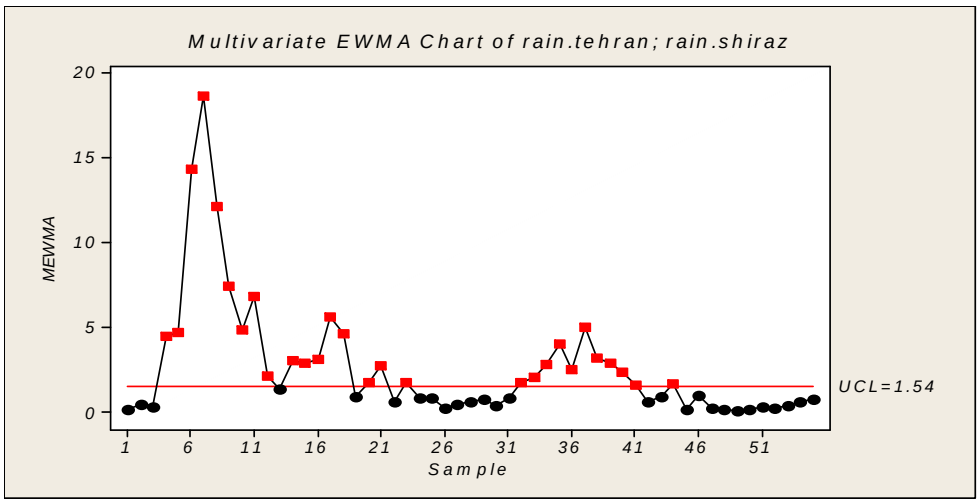
برای تبیین بیشتر از سری بارش 55 سالهٔ (1951-2005) ایستگاه های شیراز و تهران استفاده شده است. نمودار زیر کنترل چند متغیره بارش شیراز و تهران را نشان می دهد:



نمودار دیگری که در کنترل چند متغیرها بکار می رود نمودار میانگین متحرک وزنی نمایی[1] (EWMA) می باشد . در این آمارهٔ کنترل متوسط طول دوره[2] (ARL) اهمیت زیادی داردکه در واقع متوسط تعداد دوره هایی است که بعداز آن تغییرات در فرآیند را به دست آورده می شود .مقدار آماره آن را می توان از رابطهٔ زیر بدست آورد:

$$T_i^2 = z_i' \sum_{2i}^{-1} z_i$$

در این راستا بارش ایستگاه های شیراز و تهران مورد بررسی قرار گرفت:

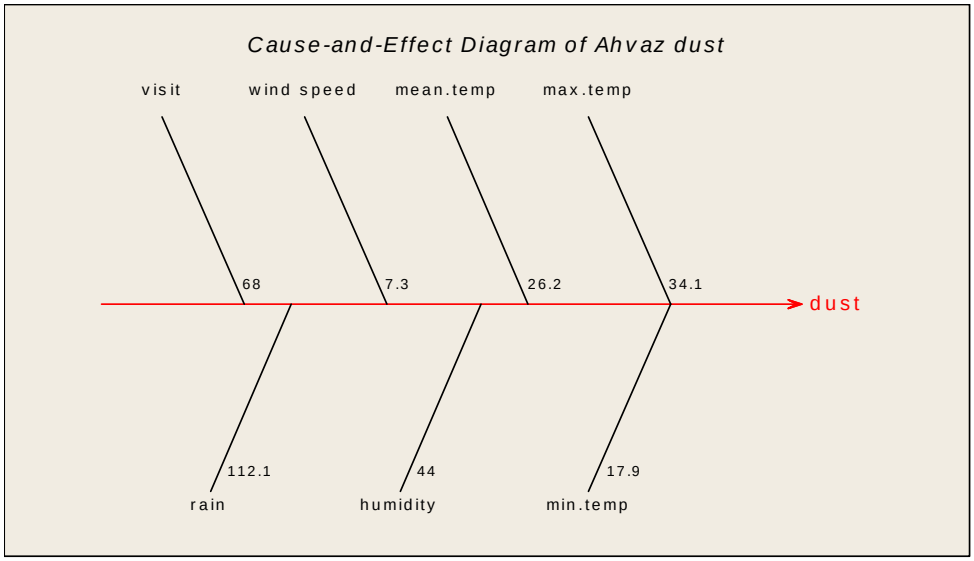


تحلیل علّی متغیرهای اقلیمی با استفاده از نمودار علت و معلول:

یک نمودار استخوانی یا علی و معلولی ، می تواند علت های بالقوه تغییرات سری اقلیمی را نمایش دهد. بدین صورت یک مشکل(معلول) در سمت راست و فهرستی از علت ها در سمت چپ دریک ساختار درختی مانند به نمایش در می آیند . شاخه های درختی اغلب به وسیلهٔ طبقه بندی اصلی علت ها به هم مرتبط می شوند . هر شاخه ای فهرستی از علت های اختصاصی هر طبقه را به همراه دارد . نمودار علت و معلولی ابزار مناسبی برای سازماندهی اطلاعات در بارهٔ علت های یک مشکل می باشد. نمودارهای علت و معلول هم بطور ساده وهم به

[1] - Exponentially weighted moving averages
[2] - Average Run Length

صورت مرکب ترسیم می شوند. بر این اساس برای فهم بهتر موضوع از متغیرهای متعددی (سری 20 سالهٔ 2005 تا 1986 دمای حداکثر، حداقل، میانگین ، سرعت باد، بارش، رطوبت نسبی و میدان دید) در تحلیل علت و معلولی منشاء بادهای گرد و غبار ایستگاه اهواز استفاده شده است :



کاربرد آزمون های آماری برای سنجش داده ها و سری های اقلیمی :

سنجش در واقع مشخص کردن ارزش هر متغیر یا مدل بر حسب قاعدهٔ خاص به منظور کمی کردن و اعتبارسنجی داده ها و مدل ها می باشد. مهمترین فرآیندی که یک اقلیم شناس باید به آن توجه نماید مشخص کردن سطح سنجش داده ها و مدل ها است . براین اساس در سنجش نمی توان قاعدهٔ عمومی را توصیه نمود . فرآیند سنجش به ما می گوید هر سری و هر مدلی را که اقلیم شناس می خواهد استفاده کند باید اول اعتبار

آن سری و آن سری را در قالب آن مدل بررسی کرد . یعنی محقق اول باید بسنجد سری اعتبار لازم برای تحلیل را دارد یا ندارد؟ اگر سری اعتبار دارد ، آن را به صورت مدل مورد نظر مطرح کند . یعنی محقق در دو مرحله عملیات سنجش را انجام می دهد . مرحلهٔ اول سنجش سری و مرحلهٔ دوم سنجش سری به صورت مدل است . یعنی سنجش به صورت داده ای و مدلی است .سطوحی که در تحقیقات مورد استفاده قرار می گیرند به عنوان مقیاس های سطوح سنجش مطرح می باشند. مقیاس های سطوح سنجش مقیا س های اسمی، رتبه ای، فاصله ای ونسبتی می باشند. متغیرها در قالب سطوح سنجش مورد استفاده قر ار می گیرد. مقیاس اسمی به متغّیر هایی تعلق می گیرد که تفکیک پذیر باشند و بتوان آنها را در طبقه ها یا دسته های جداگانه قرار داد تا تفاوت کمیت های مربوط مشخص شود . در این مقیاس ملاک طبقه بندی مبتنی بر ویژگی های مشترک افراد یا رویدادها است و ملا کهای طبقه بندی جنبهٔ ریاضی - آماری ندارد. وقتی متغّیر مستقل از چند سطح تشکیل شده باشد، متغیرمستقل اسمی نامیده می شود. برای مثال اگر هدف سطح بندی منطقه ای به صورت برخوردار و کم برخوردار از نظر امکانات در یک تحقیق جغرافیایی باشد، سطح بندی برخوردار و کم برخوردار به عنوان متغیر اسمی در طرح بالا قلمداد می شود . مقیاس رتبه ای یا ترتیبی، نشان دهندهٔ وضعیت و جایگاه یک فرد یا یک گروه یا یک چیز در داخل مجموعه ای از افراد می باشد که از طریق دادن کمیت و مقدار به تک تک آنها انجام داده می شود. در این مقیاس به تعداد افراد یا پدیده ها رتبه وجود دارد که مشخص کنندهٔ سلسله مراتب آنهاست ، مثلاً طبقه بندی ایران به چند قلمرو از نظر داشتن یا نداشتن رطوبت. بر این اساس در این مقیاس اعداد منسوب به رده ها و مقولات امکان تنظیم داده ها را با تعیین اولویت ها و ترتیب ها فراهم می کند. مقیاس فاصله ای صفت معینی را در خصوص افراد یا اشیاء یک گروه با رعایت تفاوت بین فاصلهٔ آنها، مشخص می کند. مقیاس فاصله ای نه تنها به ترتیب اشیاء توجه دارد بلکه فاصلهٔ بین آنها را نیز مشخص می سازد . علاوه بر آن در این مقیاس، مبدأ صفر وجود ندارد. مقیاس نسبتی صفت مشخصی را درخصوص افراد یا پدیده های یک گروه و رعایت تساوی بین فاصلهٔ آنها با یکدیگر مشخص می کند. در این مقیاس نه تنها مقدار صفت مورد نظر و کمی و زیادی آن نسبت به افراد دیگر در داخل یک گروه، بلکه مقدار تفاوت آنها با یکدیگر هم مشخص می شود . مقیاس نسبتی کامل ترین و دقیق ترین مقیاس اندازه گیری است. داده های مورد استفاده در تحقیقات اقلیمی قالب فاصله ای و نسبتی دارند. سطوحی که در تحقیقات مورد استفاده قرار می گیرد بعنوان مقیاس های سطوح سنجش مطرح می باشند . مقیاس های سطوح سنجش مقیاس های اسمی ، رتبه ای ، فاصله ای و نسبتی می باشند. متغیرهایی که در قالب سطوح سنجش مورد استفاده قرار می گیرد از یک نظر متغیرهای کمی و کیفی هستند . متغیرهای کمی در قالب مقیاس های فاصله ای و نسبتی مورد سنجش قرار می گیرند و متغیرهای کیفی در قالب مقیاس های اسمی و رتبه ای مورد سنجش قرار می گیرند. متغیرهای کیفی که متغیرهای مقوله ای نیز نام دارند ، متغیّر هایی هستند که مقادیر عددی به خود نمی

گیرند یا کمیت پذیر نیستند . متغّیرهای کیفی شامل ویژگی هایی از قبیل جنسیت ، مذهب ، نژاد ، زبان ، سواد ، دین و... هستند. متغیرهای کمی متغیرهایی هستند که برای اندازه گیری آنها می توان اعداد را به وضعیت های آزمودنی و قابل سنجش برطبق قاعده ای، معّین نمود .این متغیر ها را متغیرهای کمیت پذیر هم می گویند . مانند : بارش ، دما ، فشار ، سرعت باد ، رطوبت و ... متناسب با متغیرها ، مقیاس ها در قالب سطوح سنجش مورد توجه قرار می گیرند . مقیاس اسمی به متغّیر هایی تعلق می گیرد که تفکیک پذیر باشند و بتوان آنها را در طبقه ها یا دسته های جداگانه قرار داد تا تفاوت کمیت های مربوط مشخص شود. در این مقیاس ملاک طبقه بندی مبتنی بر ویژگی های مشترک افراد یا رویدادها است و ملاک های طبقه بندی جنبهٔ ریاضی _ آماری ندارد . وقتی متغّیر مستقل از چند سطح تشکیل شده باشد ، متغّیر مستقل اسمی نامیده می شود. برای مثال اگر هدف سطح بندی منطقه ای به صورت برخوردار و کم برخوردار از نظر امکانات در یک تحقیق جغرافیایی باشد ، سطح بندی برخوردار و کم برخوردار به عنوان متغیر اسمی در طرح بالا قلمداد می شود. مقیاس رتبه ای یا ترتیبی ، نشان دهندهٔ وضعیت و جایگاه یک فرد یا یک گروه یا یک چیزدر داخل مجموعه ای از افراد می باشد که از طریق دادن کمیّت و مقدار به تک تک آنها انجام داده می شود. در این مقیاس به تعداد افراد یا پدیده ها رتبه وجود دارد که مشخص کنندهٔ سلسله مراتب آنهاست . مثلا" طبقه بندی ایران به چند قلمرو از نظر داشتن یا نداشتن رطوبت . بر این اساس در این مقیاس اعداد منسوب به رده ها و مقولات امکان تنظیم داده ها را با تعیین اولویت ها و ترتیب ها فراهم می کند. مقیاس فاصله ای صفت معینی را در خصوص افراد یا اشیاء یک گروه با رعایت تفاوت بین فاصلهٔ آنها ، مشخص می کند . مقیاس فاصله ای نه تنها به ترتیب اشیاء توجه دارد بلکه فاصلهٔ بین آنها را نیز مشخص می سازد. علاوه بر آن در این مقیاس ، مبدأ صفر وجود ندارد .مقیاس نسبتی صفت مشخصی را در خصوص افراد یا پدیده های یک گروه و رعایت تساوی بین فاصلهٔ آنها با یکدیگر مشخص می کند . در این مقیاس نه تنها مقدار صفت مورد نظر وکمی و زیادی آن نسبت به افراد دیگر در داخل یک گروه ، بلکه مقدار تفاوت آنها با یکدیگر هم مشخص می شود. مقیاس نسبتی کامل ترین و دقیق ترین مقیاس اندازه گیری است . اکثرداده های مورد استفاده در تحقیقات اقلیمی قالب فاصله ای و نسبتی دارند.

سنجش داده اقلیمی متناسب با سطوح مقیاس سنجش :

یکی از متداول ترین تقسیم بندی های آماری سطوح سنجش داده ها بر اساس عاملی و ناعاملی بودن آنها می باشد . بر این اساس زمانی می توان داده ها را مورد سنجش قرار داد که ماهیت عاملی یا پارامتریک و ناعاملی یا ناپارامتریک آنها مشخص شود. لذا پیش فرض های مهمی برای کاربرد آزمون های عاملی و ناعاملی وجود دارد که باید اقلیم شناس در کاربرد به آنها توجه نماید. چنانچه متغّیر مورد مطالعه در تحلیل های کمّی دارای توزیع بهنجار باشد و آزمودنی ها بوسیلهٔ نمونه گیری تصادفی بدست آمده باشند ، برای تحلیل می توان آزمون های عاملی یا پارامتریک را بکار گرفت. چنانچه متغّیر مورد نظر توزیع بهنجاری نداشته باشد برای تحلیل باید از آزمون های ناعاملی یا ناپارامتریک استفاده نمود. آزمونهای ناپارامتری معمولا" برای متغّیرهایی که در آن مقیاس های اسمی و رتبه ای اندازه گیری شده اند ، بکار می رود . در مواردی که تعداد نمونه کوچک باشد برای متغّیرهای اندازه گیری شده در مقیاس فاصله ای نیز از آزمونهای ناپارامتریک می توان استفاده نمود. در حالیکه آزمون های پارامتریک را می توان برای متغّیر هایی که حداقل در مقیاس های فاصله ای اندازه گیری شده اند، استفاده کرد. در تحلیل های پارامتریک برای توزیع متغّیر مورد اندازه گیری در جامعه ای که نمونه ای از آن بدست آمده است ، پیش فرض هایی صورت می گیرد . اعتبار نتایج آزمون پارامتری به صادق بودن این پیش فرض ها بستگی دارد . ازجمله این پیش فرض ها بهنجار بودن ، استقلال مشاهدات و برابری واریانس می باشد . آزمونهایی که هیچ گونه فرضی در بارهٔ توزیع آنها وجود ندارد ، بصورت تحلیل های آزاد توزیع مطرح می شوند که اغلب ناپارامتریک هستند. در این مورد آزمون های فرضیه ای که استفاده می شوند می توانند یک دامنه و یا دو دامنه باشند . برای بررسی رابطهٔ سطوح مذکور می توان از ضرائب زیر استفاده کرد:

چنانچه متغّیرها ، اسمی (مانند جنسیّت ، مذهب و ...) باشند برای تعیین میزان همبستگی بین دو متغیر از ضریب لاندا (λ) و فی(ϕ) استفاده می شود[1] .

ضریب لاندا (λ) :

ضریب همبستگی لاندا برای محاسبهٔ حدود رابطهٔ بین متغّیرهای اسمی بکار می رود . این ضریب یک طرفه است و رابطهٔ بین دو متغّیر را فقط در یک جهت نشان می دهد . دامنهٔ ضریب ، بین صفر و یک متغّیر است . صفر نشان دهندهٔ عدم رابطه بین دو متغّیر است و یک نشان می دهد که بین دو متغّیر رابطهٔ کامل وجود دارد . برای محاسبهٔ ضریب لاندا از معادلهٔ زیر استفاده می شود :

$$\lambda = \frac{\sum f_i - F_d}{N - F_d}$$

[1] - جاوری , مجید : مقدمه ای بر روش های تحقیق در علوم انسانی با تأکید بر جغرافیا , انتشارات طلور , 1384.

f_i بیشتر ین فراوانی هر طبقه متغیر مستقل ، F_d بیشتر ین فراوانی کل متغیر وابسته و N تعداد کل می باشد. برای مثال در تحقیقی ، هدف ، بررسی میزان رضایت مندی کارکنان یک سازمان نسبت به خدمات دریافتی خود است ، چنانچه بعد از بررسی نتایج بصورت جدول ذیل باشد ، رابطهٔ مذکور چگونه است ؟

	نوع کارمندی			
جمع	اداری	آموزشی	خدماتی	میزان رضایت مندی
110	***40***	***50***	***20***	مثبت
100	***55***	***30***	***15***	منفی
150	***60***	***40***	***50***	بی تفاوت
360	***155***	***120***	***85***	جمع

$$\sum f_i = 50+50+60 = 160$$
$$\sum F_d = 110$$
$$\lambda = \frac{160-110}{360-110} = \frac{50}{250} = .2$$

ضریب فی (ϕ) :

برای تعیین همبستگی متغّیر های اسمی که هر کدام دو طبقه داشته باشند از ضریب فی می توان استفاده نمود . دامنهٔ این ضریب مثبت و منفی یک است . مثبت یک همبستگی کامل و منفی یک همبستگی معکوس کامل است و صفر عدم همبستگی را نشان می دهد . برای محاسبهٔ ضریب فی از معادلهٔ زیر استفاده می شود :

$$\phi = \frac{(ab-bc)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

این ضریب با آزمون خی دو در ارتباط است . $X^2 = N \times \phi$ **(3-1)** و $\phi = \frac{X^2}{N}$

مثلا اگر داده های جدول زیر را وجود داشته باشد مقدار ضریب فی آن برابر است با :

جمع	تحصیلات		درآمد
	دانشگاهی	غیر دانشگاهی	
50	***(a)40***	***(b)10***	بالا
50	***(c)10***	***(d)40***	پایین
100	***50***	***50***	جمع

$$\phi = \frac{(40\times 40)-(10\times 10)}{\sqrt{(50)(50)(50)(50)}} = \frac{1500}{\sqrt{6250000}} = .6$$

سنجش درجهٔ قوت رابطهٔ بین متغّیرهای ترتیبی :

برای سنجش رابطهٔ بین متغّیرهای ترتیبی روش های متعددی وجود دارد (روش های اسپیرمن ، تا کندال و **...**)که می توان به عنوان نمونه به ضریب یولزکیو(ضریب Q) اشاره نمود . این ضریب هم دامنه ای بین مثبت و منفی یک دارد . ضریب Q را می توان از رابطهٔ زیر بدست آورد :

$$Q = \frac{(ab) - (bc)}{(ab) + (bc)}$$

مثلا اگر سری داده های جدول ذیل وجود داشته باشند ضریب Q آن برابر است با :

رعایت قوانین	تحصیلات		جمع
	کمتر از دیپلم	بالای دیپلم	
زیاد	***100***	***80***	***180***
کم	***200***	***20***	***220***
جمع	***300***	***100***	***400***

$$Q = \frac{(80 \times 200) - (100 \times 20)}{(80 \times 200) + (100 \times 20)} = \frac{14000}{18000} = .778$$

برای سنجش رابطهٔ بین متغیرهای فاصله ای و نسبتی، می توان از ضریب همبستگی متغّیرها مانند ضریب همبستگی پیرسون استفاده کرد . نحوهٔ کاربرد ضریب همبستگی پیرسون در جلد دوم کتاب به آن اشاره شده است.

آزمون های آماری مربوط به انواع مقیاس :

با گسترش روزافزون تحقیقات اقلیمی و روش های آماری جدید اولین تکنیک های استنباطی بوجود آمده اند و برای داده هایی بکار گرفته شده اند که مفروضات فراوانی داشته اند. با توجه به مفروضات کاربرد آزمون ها در این خصوص لازم است به اختصار به سطوح سنجش و آماره های مربوط به هر سطح در قالب تحقیقات اقلیمی اشاره شود. بر این اساس در کاربرد آزمون ها باید به مقیاس، روابط آماری ، ماهیت نمونه و آماره های مربوط توجه نمود. در مطالعات اقلیمی سطوح محتوایی مطالعات آماری را می توان در قالب های سطوح تصادفی ، گرایش دار ، فصلی و دوره ای تقسیم بندی نمود . متناسب با کاربرد سطوح محتوایی آماری ، آزمون های بیشماری وجود دارد که می توان در قالب آماره های ناپارامتری مورد استفاده قرار داد . امروزه بیش از 100 آزمون آماری وجود دارد که متناسب با ماهیت سطوح قابل استفاده می باشد. فرضیه هایی که در قالب تحقیقات اقلیمی برای کاربرد آزمون های ناپارامتریک مورد استفاده قرار می گیرد فرضیه های توصیفی ، تفاوتی ، رابطه ای و متقابل می

باشند . فرضیه های توصیفی قالب یک متغیره و فرضیه های تفاوتی ، رابطه ای و متقابل دو یا چند متغیره دارند. برای مثال اگر فرضیهٔ تحقیق اقلیم شناس این باشد که با افزایش ارتفاع کوههای البرز بارش کم می شود ، این فرضیه توصیفی است و چنانچه فرضیهٔ اقلیم شناس این باشد که بارش از نوع برف نسبت به باران با افزایش ارتفاع البرز کاهش می یابد ، فرضیهٔ تفاوتی و چنانچه فرضیهٔ اقلیم شناس میزان ارتباط بارش در ارتفاعات البرز با سرعت تودهٔ هوا باشد ، فرضیهٔ مورد استفاده فرضیهٔ رابطه ای است . چنانچهٔ فرضیهٔ اقلیم شناس بررسی رابطهٔ بارش در ارتفاع البرز با توجه به تفاوت دامنه ای آن باشد ، فرضیه متقابل می باشد. از طرفی در کاربرد آزمون ها باید به مستقل و وابسته بودن داده ها توجه نمود . اگر فرضیه های تحقیق بیانگر ادعاهای محقق باشد که وی درصدد آزمون و سپس تأیید یا رد آنهاست . محقق باید شواهد قوی داشته باشد تا آنها را بپذیرد یا رد کند. فرضّیه های تحقیق در مقام آزمون بصورت مخالف یا صفر نوشته می شوند و زمانی که استنباط آماری موجب رد فرضّیهٔ صفر بشود ، فرض تحقیق پذیرفته می شود. اگر فرضیهٔ تحقیق مبین تفاوت بین دو متغّیر باشد ، در اینصورت رابطهٔ دو میانگین فوق چنین است ($\mu_1 \neq \mu_2$) ولی اگر فرضیهٔ تحقیق، مبیّن جهت تفاوت فوق نیز باشد ، یعنی اگر فرض، مبیّن بیشتر یا کمتر بودن یکی نسبت به دیگری باشد ، در این حالت رابطه بصورت($\mu_1 > \mu_2$ $\mu_1 < \mu_2$) خواهد بود . اگر احتمال رابطهٔ فوق که به صورت یک مقدار عددی از آزمون بدست آمده است ، تحت فرضیهٔ صفر برابر یا کمتر از α یعنی سطح معنی داری که غالبا" به شکل درصد($\alpha = 5\%,\ \alpha = 1\%$) مورد استفاده است ، باشد فرض صفر رد شده و فرض تحقیق قبول می شود و برعکس . سطح معنی داری در تحلیل های آزمون ها ، مبیّن پذیرش یا عدم پذیرش فرض صفر در سطح معین (۹۹٪ یا ۹۵٪ یا۹۰٪) یا (0/01 یا 0/05 یا 0/1) می باشد. همچنین درجهٔ آزادی در این قالب مورد استفاده قرار می گیرد . درجهٔ آزادی به میزان مستقل بودن مشاهدات از یکدیگر اشاره می کند . با توجه به مطالب جدول شمارهٔ 1 سطح سنجش و آماره های مربوط را نشان می دهد .

جدول شمارهٔ 1: سطح سنجش و آماره های مربوط به هر سطح:

مقیاس	آماره ها	آزمون ها
اسمی	آماره های توصیفی(مد ، فراوانی و...)	ناپارامتریک
رتبه ای	آماره های توصیفی (میانه ، صدک ها ، همبستگی رتبه ای	ناپارامتریک

	و...)	
فاصله ای	آماره های توصیفی و استنباطی(میانگین ، انحراف معیار ، همبستگی های چند گانه و...	پارامتری و ناپارامتری
نسبتی	آماره های توصیفی و استنباطی و پیشرفته	پارامتری و ناپارامتری

سنجش داده ها در قالب تغییرات اقلیم :

بسته به موضوع و عامل زمان استفاده از آزمون ها در مطالعات اقلیمی متفاوت می باشد . بر این اساس آزمون های اقلیمی و اقلیم سنجی در به صورت متفاوت مد نظر قرار می گیرد . در راستای کاربرد آزمون های اقلیمی با در نظر گرفتن ماهیت سطح سنجش سری ها آزمون های مختلفی بکار می رود. در اقلیم سنجی آماری بررسی تغییرات تصادفی ، گرایش دار ، فصلی و دوره ای اهمیت زیادی دارد .برای سنجش داده های اقلیمی از آزمون های آماری استفاده می شود. آزمون های سنجش با توجه به ماهیت موضوع و اثر زمان در مطالعات اقلیمی متفاوت می باشند .ساده ترین شکل پیش بینی در اقلیم سنجی ، سری های تصادفی می باشد . یک فرآیند تصادفی را می توان بصورت یک موضوع آماری که بر طبق قوانین احتمالی در زمان تکامل پیدا می کند ، بیان نمود . فرآیند های ایستا مؤلفهٔ مهمی از فرآیند های تصادفی را تشکیل می دهند . مشخصه های مدل سادهٔ بدون روند سطح ثابت میانگین سری و عنصر تصادف هستند. سنجش سری های تصادفی با توجه ماهیت عاملی یا ناعاملی بودن آنها با استفاده از آزمونهای عاملی و ناعاملی قابل سنجش می باشند . آزمون های ناعاملی سنجش سری های تصادفی ، آزمون های توالی ، نقاط گردش ، دنیلس ، تاوکندال و... و آزمون های عاملی سنجش سری های تصادفی ، آزمون های وان نیومن ، تابع خودهمبستگی ، تومپ ،پرایش –بوکس Q و ... هستند. شرایط آزمون ها را می توان در جدول شمارهٔ 2 مشاهده کرد.

جدول شمارهٔ2 : آزمون های عاملی و ناعاملی سنجش سری تصادفی:

نام آزمون	آزمون	قاعدهٔ آزمون	قاعدهٔ تصمیم گیری	نتیجهٔ آزمون
توالی	$Z = \frac{\lvert R - \mu_R \rvert}{\sigma_R}$	سری در سطح معین	$Re\,ject : H_0 if \ \lvert Z \rvert > Z_{\alpha/2}$	اگر مقدار نمرهٔ آزمون کمتر از مقدار نمرهٔ

		تصادفی است = H_O سری در سطح معین تصادفی نیست = H_a		جدول باشد فرض صفر رد می شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری تصادفی می باشد .
نقاط گردش	$Z = \left\| \frac{U - \mu_U}{\sigma_U} \right\|$	سری تصادفی است : H_0 سری تصادفی نیست : H_a	$Re\ ject : H_0 if \|Z\| > Z_{\alpha/2}$	اگر نمرهٔآزمون نقاط چرخش بیشتر از مقدار جدول باشد ، فرض صفر رد می شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری تصادفی است .
نشانه	$Z = \frac{V - \mu_V}{\sigma_V}$	سری بدون روند است : H_0 سری روند دارد : H_α	$Re\ ject : H_0 if \|Z\| > Z_{\alpha/2}$	اگر مقدارنمره کمتر از مقدار نمرهٔ جدول باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری تصادفی است .
دنیلس	$Z = \frac{r_s - \mu_{rs}}{\sigma_{rs}}$	سری روند ندارد : H_0 سری دارای روند است : H_α	$\mathrm{Re}\,ject : H_0 if \|Z\| > Z_{\alpha/2}$	اگر نمرهٔ آزمون از نمرهٔ جدول کمتر باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری بدون روند است.

تاو کندال	$Z_{\tau} = \frac{\tau - \mu_{\tau}}{\sigma_{\tau}}$	سری تصادفی است : H_0 سری دارای روند است : H_a	$\text{Re}ject : H_o if \lvert Z_{\tau}\rvert > Z_{\alpha/2}$	اگر نمرهٔ آزمون از نمرهٔ جدول کمتر باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری بدون روند است.
آزمون های عاملی برای سنجش تصادفی سری ها				
وان نیومن	$M = \frac{SS_{\Delta y}}{SS_{yy}}$	سری تصادفی و ایستایی است : H_o سری یاناایستا یا مانده های همبسته دارد : H_a	$\text{Re}ject : H_o if M < M_{1-\alpha/2}$	اگر نمرهٔ آزمون از نمرهٔ جدول کمتر باشد ، فرض صفر رد می شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری روند دارد.
تابع خودهمبستگی	$r_k = \frac{\sum (Y_{t-k} - \bar{Y})(Y_i - \bar{Y})}{SS_{yy}}$	$H_o : \rho_k = 0$ $H_a : \rho_k \neq 0$	$\text{Re}ject : H_o if \lvert r_k\rvert > 2/\sqrt{n}$	اگر نمرهٔ آزمون از نمرهٔ جدول کمتر باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری روند ندارد.
تومپ	$r_k = \frac{\sum (Y_{t-k} - \bar{Y})(Y_i - \bar{Y})}{SS_{yy}}$	$H_o : \rho_k = 0$ $H_a : \rho_k \neq 0$	$\text{Re}ject : H_o if \lvert r_k\rvert > 2/\sqrt{n}$	اگر نمرهٔ آزمون از نمرهٔ جدول کمتر باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری روند ندارد.

پرایس_بوکس Q	$Q_m = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k}$	$H_o : \rho_k = 0$ $H_a : \rho_k \neq 0$	$\text{Re}ject : H_o if Q_m > X_{\alpha}^2(m)$	اگر نمرهٔ آزمون از نمرهٔ جدول کمتر باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری تصادفی است.

مطالعهٔ شیوه های پیش بینی سری های زمانی الگوی های بلند مدت رشد یا عدم رشد را نشان می دهد . مقدار قابل انتظار یک سری ممکن است با گذر زمان تغییر کند . اگر نوسان سری در بلند مدت یک جهت (صعودی یا نزولی) داشته باشد ، روند نمود می یابد .آزمون های سنجش سری گرایش دار در قالب عاملی و ناعاملی متعدد هستند که برای ناعاملی دنیلس و برای عاملی آزمون t بیشتر استفاده می شود. برای آزمون دنیلس قاعده تصمیم گیری به این صورت است که اگر نمرهٔ آزمون از نمرهٔ جدول کمتر باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که سری تصادفی است . مشخصه های آزمون عاملی یا آزمون t در جدول شمارهٔ 3 مشخص شده است .

جدول شمارهٔ 3: آزمون سنجش سری گرایش دار:

نام آزمون	آزمون	قاعدهٔ آزمون	قاعدهٔ تصمیم گیری	نتیجهٔ آزمون
t	$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	$H_O : \rho = 0$ $H_a : \rho \neq 0$	$\text{Re}ject : H_o if \lvert t_r \rvert > t_{\alpha/2}$	اگر مقدار نمرهٔ آزمون کمتر از مقدار نمرهٔ جدول باشد فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری بدون روند است .

سری زمانی که معمولا" کوتاهتر از سال می باشد اغلب الگوی نوساناتی سال به سال را نشان می دهند، که حرکات فصلی خوانده می شوند . این الگوهای فصلی به صورت مدل های ضربی و جمعی مطرح می شوند . هر دورهٔ مشاهده ، یک فصل و مجموع فصل ها، طول فصل خوانده می شوند. از آزمون های ناعاملی متداول برای سنجش نوسانات فصلی ، آزمون کروسکال والیس می باشد.آزمون کروسکال - والیس به عنوان یک آزمون ناعاملی یک آزمون رتبه ای برای نمونه های مستقل به شمار می آید . همچنین این آزمون را آزمون تحلیل واریانس یک طرفه کروسکال - والیس هم می خوانند . برای محاسبهٔ آزمون مذکور می توان از معادلهٔ زیر استفاده کرد :

$$H = \frac{12}{n(n+1)}\left[\sum \frac{R_i^2}{n_i}\right] - 3(n+1)$$

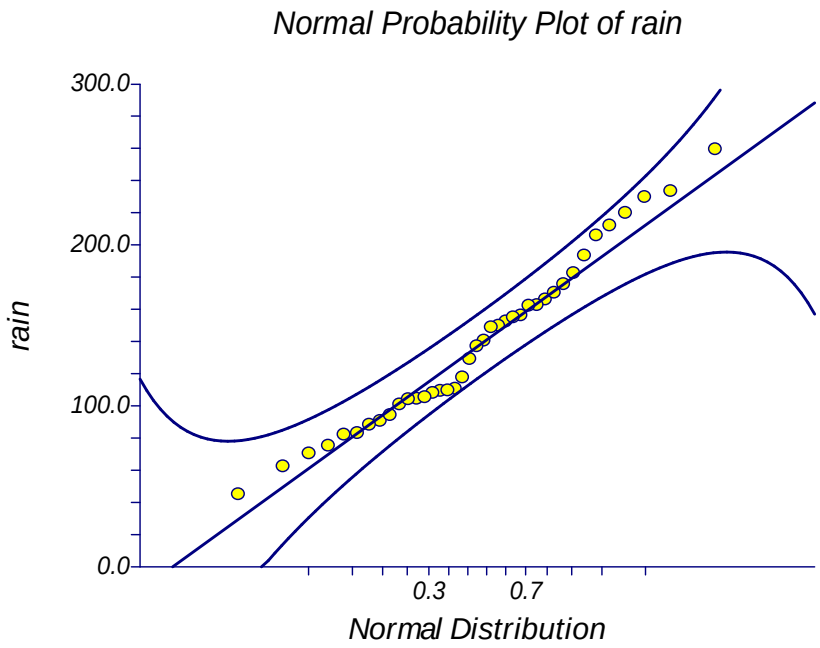
قاعدهٔ تصمیم گیری :

$$\text{Re}ject : H_o if : H > X^2_{\alpha}(L-1)$$

اگر نمرهٔ آزمون بزرگتر از نمرهٔ جدول باشد، فرض صفر رد می شود و با احتمال 95 درصد اطمینان نتیجه گرفته می شود که سری دارای نوسانات فصلی می باشد .

تغییرات دوره ای را در یک سری زمانی که خیزهای بلند مدت (غیر از روند و نوسانات فصلی) دارند می توان تحلیل و پیش بینی نمود . تحلیل تغییرات دوره ای به علت این که پایدار نیستند ، مشکل است . به هر حال تغییرات دوره ای ، تغییراتی است که غیر از شرایط روند دار و فصلی می باشد و معمولا" طولانی تر از یک سال می باشد . برای سنجش تغییرات دوره ای لازم است از آزمون های آماری استفاده کرد . یکی از شیوه های سنجش تغییرات دوره ای استفاده ازآزمون فصول لازم برای غلبه یا اثر غالب دوره ای است .آزمون های عاملی برای سنجش تغییرات دوره ای آزمون وان نیومن و تابع خودهمبستگی است که به آن اشاره شد. به هر حال برای سنجش اقلیمی هم به صورت سری و هم به صورت مدل باید از آزمون های مختلف عاملی و ناعاملی استفاده کرد. برای فهم بهتر لازم است به موقعیت شمارهٔ 8 که هدف آن سنجش بهنجاری سری بارش ایستگاه کاشان است توجه نمود .

موقعیت شمارهٔ 8 :



با توجه به بهنجار بارش سالانه ایستگاه کاشان (1967 تا 2005) می توان برای سنجش سری در قالب مدل از آزمون های عاملی یا پارامتری استفاده نمود .

کاربرد آزمون های آماری در سنجش سری های اقلیمی[1] :

با توجه به ماهیت داده ها , موضوع مورد مطالعه و سطح سنجش سری های اقلیمی آزمون های عاملی و ناعاملی مورد توجه قرار می گیرد . برای استفاده از آزمون ها باید به مقیاس سنجش , ماهیت و معیار های کاربرد آزمون ها که در قسمت قبلی اشاره شد , توجه نمود. با توجه به سطح سنجش (اسمی بودن مقیاس) آزمون های ناعاملی مد نظر قرار می گیرد . مهمترین آزمونی که در این راستا اشاره می شود آزمون های خی دو و کالماگورف – اسمیرنوف هستند , بنابراین برای استفاده از آزمون های مذکور لازم است به موقعیت های مختلف اشاره شود.

موقعیت شمارهٔ 1:

وقتی هدف تحقیق در اقلیم شناسی بررسی رابطهٔ دو متغیر اسمی باشد می توان رابطهٔ متغیر ها را در قالب جدول توافقی متناسب با رابطهٔ بین متغیر ها بسته به عدم جهت , میزان و ماهیت داده ها مورد توجه قرار داد. برای تبیین ساده تر به مثال زیر توجه نمایید:

مثال : اگر محققی در صدد بررسی ارتباط بین میزان رطوبت در نقاط مختلف ایستگاهی یک منطقه باشد و در بررسی اطلاعات جدول زیر بدست آمده باشد . ارتباط ایستگاهی آن منطقه را بسنجید.

- **کاربرد آزمون خی دو :**

مقدار رطوبت						منطقهٔ ایستگاهی	
مجموع	خیلی زیاد (بیش از 80 درصد)	زیاد(بین 80 تا 60 درصد)	متوسط (50 تا 60 درصد)	کم (30 تا 50 درصد)	خیلی کم (کمتر از 30 درصد)		میزان رطوبت

[1] - به کتاب شیو ه های تجزیه و تحلیل کمی در اقلیم شناسی (با تاکید برمدل های روند) مؤلف مراجعه شود.

	1	26	45	55	67	84	277
	2	25	36	53	69	90	273
	3	24	34	55	70	85	268
	4	20	48	56	71	89	284
	5	27	44	53	72	91	287
مجموع		122	207	272	349	439	1389

برای به دست آوردن فراوانی منتظره می توان از حاصل ضرب مجموع افقی در مجموع عمودی به مجموع کل استفاده کرد, یعنی :

F_0	f_e	$(f_o-f_e)^2$	$(f_o-f_e)^2/ f_e$
26	24/33	2/79	0/11
25	40/68	245/86	6/04
24	23/54	0/211	0/009
20	24/94	24/4	0/978
27	25/21	3/2	0/127
45	41/28	13/83	0/335
36	40/68	21/9	0/538
34	39/94	35/28	0/88
48	42/32	32/26	0/762
44	42/77	1/51	0/035
55	54/24	0/578	0/01
53	53/46	0/21	0/004
55	52/48	6/35	0/12

56	55/61	0/15	0/002
53	56/2	10/24	0/18
67	69/599	6/75	0/097
69	68/59	0/168	0/002
70	67/33	7/13	0/106
71	71/358	0/128	0/001
72	72/11	0/012	0/0001
84	87/54	12/53	0/143
90	86/28	13/83	0/16
85	84/7	0/09	0/001
89	89/76	0/578	0/006
91	90/7	0/09	0/0009
جمع			10/647

درجهٔ آزادی جدول برابر است با :

$$df = (k-1)(c-1)$$

$$df = (5-1)(5-1) = 16$$

نتیجه گیری :

چون مقدار آزمون محاسبه شده (10/647) در سطح 0/05 از مقدار ضریب بحرانی جدول (26/296) کوچکتر است , فرض صفر رد نمی شود و نتیجه گرفته می شود که رابطهٔ متغیرها براساس شانس است و واقعی نیست.

مقدار ضریب خی دو نشانگر معنی داری سری از نظر آماری است و رابطهٔ متغیر ها را مشخص نمی سازد . لذا برای مشخص کردن میزان همبستگی یا رابطه بین متغیر ها می توان از شاخص هایی مانند ضریب فی , ضریب توافق پیرسون , ضریب همبستگی چوپرو و... استفاده کرد که به بعضی از آنها در قسمت قبلی اشاره شده است . برای تصحیح آزمون خی دو می توان از ضریب تصحیح یتس [1] استفاده کرد. فرمول تصحیح یتس برابر است با :

$$X^2_{yates} = \sum \frac{\left[\left|F_o - F_e\right| - 0.5\right]^2}{F_e}$$

زمانی از فرمول تصحیح یتس می توان استفاده کرد که حجم نمونه کمتر از 4 و درجه آزادی جدول هم کمتر از 3 باشد.

- **کاربرد آزمون مان- ویتنی :**

آزمون مان – ویتنی [2] بعنوان یک آزمون ناعاملی (مشخصه های کاربرد آزمون ها یعنی بهنجار نبودن و...داشته باشد) اهمیت زیادی در آمار استنباطی در سنجش یک متغیر اسمی و یک متغیر رتبه ای دارد .از آزمون مان – ویتنی زمانی که استنباط می شود آیا بین توزیع های جامعه تفاوت وجود دارد یا خیر ؟ استفاده می شود . حتی این آزمون را می توان برای متغیرهای رتبه ای استفاده کرد. آزمون مذکور با توجه به حجم نمونه به اشکال مختلفی قابل محاسبه است . در مطالعات جغرافیایی چون نمونه های مختلفی کاربرد دارد ،لذا لازم است با توجه به حجم نمونه از روش خاصی برای محاسبهٔ آزمون استفاده کرد. نمونه های مورد نظر در محاسبهٔ آزمون را می توان به انواع زیر تقسیم نمود[3]:

- **حجم نمونه های کوچک (کمتر از *8* نمونه).**
- **حجم نمونه های متوسط(بین *9* تا *20* نمونه).**
- **حجم نمونه های بزرگ (بیش از *20* نمونه).**

روش حجم نمونهٔ کوچک :

برای تبیین روش لازم است به مثال زیر اشاره شود:

مثال :

استاد درس برنامه ریزی محیطی برای آموزش درس خود معتقد است استفاده از نرم افزار *ArcGIS* امکان تحلیل داده ها مکانی را بهتر از استفاده از نرم افزار *ArcView* فراهم می کند به همین منظور تعداد *10* نفر دانشجو را انتخاب می کند ، سپس *5* نفر از دانشجویان را به طور تصادفی با روش استفاده از *ArcGIS* و *5* نفر دیگر را با روش استفاده از *ArcView* برای تحلیل داده های مکانی آموزش می دهد . پس از پایان دوره از تمامی دانشجویان در ارتباط با دقت تحلیل داده های مکانی امتحان گرفته شده است و نتایج آن به صورت زیر بدست آمده است :

ArcGIS	***17***	***14***	***19***	***18***	***13***
ArcView	***16***	***15***	***12***	***13***	***18***

[1] - Yates Correction

[2] - Mann-Whitney Test

[3] - جاوری , مجید و... : روش تحقیق در جغرافیا , دانشگاه پیام نور .

فرض تحقیق این است که روش استفاده از *ArcGIS* (*G*) بهتر از *ArcView* (*V*)در تحلیل داده های مکانی می باشد و فرض صفر هم این است که تفاوتی بین دو روش مورد استفاده وجود ندارد .

مراحل محاسبه :

- **ابتدا نمرات را از کوچک به بزرگ مرتب کنید .**
- **مقدار مرتبهٔ (*u*) هر نمره را به طور جداگانه در جامعه حساب کنید .**
- **مقدار مرتبهٔ محاسبه شده را با مقدار بحرانی جدول مقایسه نمود و نتیجه گیری نمایید . اگر مقدار هر مرتبهٔ (*u*) محاسبه شده بزرگتر از مقدار جدول باشد فرض صفر رد نمی شود.**

متناسب با مراحل:

روش	*V*	*V*	*G*	*G*	*V*	*V*	*G*	*V*	*G*	*G*
نمره	*12*	*13*	*13*	*14*	*15*	*16*	*17*	*18*	*18*	*19*

مقدار مرتبه (*u*):

برای مثال نمرهٔ *12* در روش *V* از هیچ کدام نمره های روش *G* بزرگتر نیست ، نمرهٔ *13* روش *V* از هیچ نمرهٔ روش *G* بزرگتر نمی باشد ، نمرهٔ *15* روش *V* از دو نمرهٔ (*13* و *14*) روش *G* بزرگتر است . نمرهٔ *16* روش *V* از دو نمرهٔ (*13* و *14*) روش *G* بزرگتر است . نمرهٔ *18* روش *V* از سه نمرهٔ (*13* ، *14* و *17*) روش *G* بزرگتر است و الی آخر

ArcView(u)	*0*	*0*	*2*	*2*	*3*
ArcGIS(u)	*1*	*2*	*4*	*4*	*5*
7=جمع مرتبه های روش *V*			*16*=جمع مرتبه های روش *G*		

با توجه به مقدار کمتر مرتبه های روش های (*u*) مقدار مرتبه های کمتر برآورد شده (*u=7*) با مقدار بحرانی مرتبه (*u*) جدول *n1=5* و *n2=5* برابر با *0/155* می باشد. لذا چون مقدار آزمون (*0/155*) از مقدار بحرانی (*0/05*) بیشتر است ، فرض صفر رد نمی شود ، بنابراین با احتمال *95* درصد نتیجه گرفته می شود که بین نمره های دو روش تفاوت معنی داری وجود ندارد .

روش حجم نمونهٔ متوسط :

در این روش برای بدست آوردن مرتبهٔ (*u*) می توان از معادلات زیر استفاده کرد(گودرزی ، *1388* ، *221*):

$$u_1 = n_1 \times n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

R1 **مجموع رتبه های نمونهٔ اول و *n1* حجم نمونهٔ اول می باشد . مرتبه های نمونهٔ دوم را می توان از رابطهٔ زیر بدست آورد:**

$$u_2 = n_1 \times n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

R2 **مجموع رتبه های نمونهٔ دوم و*n2* حجم نمونهٔ دوم می باشد.**

مثال :

اگر میزان خدمات شهری در دو منطقهٔ تهران مورد بررسی قرار گیرد و میزان درصد خدمات شهری در دو منطقهٔ تهران از نظر متخصصان شهری به صورت جدول زیر تنظیم شده باشد ، مقدار آزمون مان – ویتنی چقدر است ؟

منطقهٔ *1* (*n1*)	منطقهٔ *2* (*n2*)

نمره	
55	***68***
60	***78***
78	***56***
80	***90***
90	***89***
80	***90***
100	***60***
70	***70***
80	***100***
90	***70***

مراحل عملیات :

- رتبهٔ نمرات باید محاسبه شود (پایین ترین نمره رتبهٔ اول و الی آخر) .

- مجموع رتبه های هر نمونه بدست آورده شود.

- مقادیر مرتبه ها با استفاده از معادلات آزمون محاسبه شود .

- مقدار مرتبهٔ کوچکتر محاسبه شده با مقدار جدول بحرانی مقایسه می شود . اگر مرتبهٔ (u) محاسبهٔ شده بزرگتر از مقدار مرتبهٔ بحرانی جدول باشد ، فرض صفر رد نمی شود لذا با اطمینان معین نتیجه گرفته می شود که بین نمرات یا درصد نظر متخصصان در خصوص خدمات شهری در منطقه تفاوت معنی داری وجود ندارد .

متناسب با مراحل :

رتبهٔ نمرات :

منطقهٔ 1 (n_1)		منطقهٔ 2 (n_2)	
رتبه	نمره	رتبه	نمره
1	***55***	5	***68***
4	***60***	10	***78***
10	***78***	2	***56***
12	***80***	17	***90***
17	***90***	14	***89***
12	***80***	17	***90***
20	***100***	4	***60***
7	***70***	7	***70***
12	***80***	20	***100***
17	***90***	7	***70***

جمع رتبه ها	101	جمع رتبه ها	110

$$u_1 = 10\times 10 + \frac{10(10+1)}{2} - 110 = 45$$
$$u_2 = 10\times 10 + \frac{10(10+1)}{2} - 101 = 54$$

چون مقدار کوچکتر آزمون محاسبه شده (45) بزرگتر از مقدار مرتبهٔ جدول (23) است ، لذا فرض صفر رد نمی شود و با احتمال 95 درصد نتیجه گرفته می شود بین نمرات خدمات شهری در دو منطقهٔ تهران تفاوت معنی داری وجود ندارد.

روش حجم نمونه های بزرگ :

چون سری های جغرافیایی در بیشتر موارد از 20 نمونه بیشتر می باشد ، لذا از این روش بیشتر می توان استفاده کرد. چون در روش حجم نمونه های بزرگ توزیع مان – ویتنی به توزیع بهنجاری سری نزدیک می باشد . لذا بعداز محاسبهٔ آزمون مقدار Z با عدد 1l96 (در سطح 0l05) مقایسه می شود . اگر قدر مطلق ضریب بدست آمده از مقدار ضریب بحرانی جدول کمتر یا مساوی با آن باشد ، فرض صفر رد نمی شود و با احتمال 95 درصد نتیجه گرفته می شود که تفاوت معنی داری بین نمرات دو نمونه وجود ندارد و برعکس . برای بدست آوردن آزمون به این روش از فرمول زیر می توان استفاده کرد(گودرزی ، 1388 ، 224):

$$z = \frac{\left| u - \frac{n_1 \times n_2}{2} \right|}{\sqrt{\frac{(n_1)\times(n_2)\times(n_1+n_2+1)}{12}}}$$

مثال :

اگر در تحلیل جمعیت شهر اصفهان در خصوص میزان مهاجرت دو ناحیهٔ آن تحت بررسی باشد . چنانچه تعداد افراد مهاجر در ناحیهٔ اول 40 نفر و تعداد افراد مهاجر در ناحیهٔ دوم 50 نفر باشد و میزان رضایت مندی آنها در خصوص خدمات رفاهی و اجتماعی نواحی مذکور در قالب درصد اندازه گیری شده باشد همچنین طی مراحلی مقدار مرتبهٔ (u) سری محاسبه شده باشد و برابر با 75 باشد مقدار آزمون چقدر است ؟

$$z = \frac{\left| 75 - \frac{40\times 50}{2} \right|}{\sqrt{\frac{(40)\times(50)\times(40+50+1)}{12}}} = 7.52$$

چون مقدار ضریب آزمون بدست آمده (7l52) از مقدار ضریب جدول بحرانی بزرگتر است ، لذا فرض صفر رد می شود و نتیجه گرفته می شود با احتمال 95 درصد تفاوت معنی داری بین دو نمونه وجود دارد.

کاربرد آزمون مان – ویتنی با استفاده از نرم افزار SPSS :

آزمـــون t بـــرای تحلیل مقایسهٔ میانگین نمونهٔ آماری با جامعه آماری :

ازآزمونt می توان به عنوان یک آزمون پارامتری یا عاملی برای مقایسهٔ دو میانگین استفاده نمود. توزیع آزمون مذکور با توزیع بهنجاری سری تفاوت ها و شباهت هایی دارد ، شباهت این دو توزیع در این است که میانگین هر دو توزیع صفر است و تفاوت های این دو توزیع هم در این است که توزیع بهنجاری شکلی شبیه ناقوس (زنگوله ای) دارد ولی شکل توزیع آزمون مذکور بر حسب درجه آزادی و حجم نمونه متفاوت است . هرچه درجه آزادی و حجم توزیع بزرگ تر باشد ، شکل توزیع آزمون t به توزیع بهنجاری شبیه تر

می شود. زمانی که حجم نمونه از *30* بیشتر می شود تفاوت بین توزیع آزمون مذکور با توزیع بهنجار خیلی ناچیز می شود. لذا توزیع و پراکندگی آزمون *t* بیشتر از پراکندگی توزیع بهنجاری است و این موضوع با کاهش حجم نمونه بیشتر می شود. همچنین هر چه درجهٔ آزادی و حجم نمونه بزرگ تر باشد مقدار *t* جدول برای رد فرض صفر کوچک تر می شود. از طرف دیگر توزیع بهنجاری دارای کشیدگی و میانگین صفر و انحراف معیار یک است ولی کشیدگی و انحراف معیار توزیع *t* از صفر بیشتر است. در تحلیل آمار استنباطی مناسب برای یک متغیر اسمی و یک متغیر فاصله ای یا نسبی می توان از این آزمون استفاده کرد . برای کاربرد آزمون *t* در مطالعات جغرافیایی می توان در قالب انواع یک نمونه ای[1] ، نمونه های وابسته[2] و دونمونهٔ مستقل[3] استفاده کرد.

- کاربرد آزمون *t* یک نمونه ای در تحلیل های اقلیمی :

آزمون *t* استیودنت وقتی استفاده می شود که بخواهند ببینند آیا میانگین یک متغیر تفاوت معنی داری با یک مقدار آزمون دارد یا خیر . یا اگر بخواهند بررسی کنند یا میانگین و انحراف معیار مشخص ، نمونهٔ مورد بررسی به جامعه مورد نظر تعلق دارد یا خیر . لذا برای استفاده از این آزمون باید میانگین و انحراف معیار جامعه را داشت.استفاده از این آزمون به معیارهایی همچون بهنجاری توزیع جامعه ، تصادفی بودن نمونه ها ، متغیر مستقل در سطح سنجش مقیاس اسمی و متغیر وابسته در سطح سنجش فاصله ای یا نسبی بستگی دارد و مقادیر ثوابت یعنی میانگین و انحراف معیار جامعه باید مشخص باشد . به عنوان مثال ، با بررسی تفاوت میانگین بارش ایستگاهی با میانگین بارش منطقه ای که بعداز بدست آوردن *t* و تعیین درجه آزادی با مراجعه به جدول آمارهٔ *t* می توان به سؤال بالا جواب داد ، اگر مقدارقدر مطلق *t* به دست آمده از مقدار *t* جدول بیشتر باشد ، فرض صفر رد می شود ، یعنی تفاوت بین دو میانگین معنی دار است و برعکس اگر مقدار *t* بدست آمده از مقدار بحرانی جدول کمتر باشد فرض صفر رد نمی شود و نتیجه گرفته می شود با احتمال معین بین متغیر ها اختلاف معنی داری وجود ندارد. برای محاسبهٔ آزمون مذکور می توان از فرمول زیر استفاده کرد :

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

$\bar{X}$ میانگین نمونه ، μ میانگین جامعه و *S* انحراف معیار جامعه است . مقدار درجه آزادی آن برابر با $df = n - 1$ است .بر اساس حجم نمونه و معلوم یا مجهول بودن انحراف معیار , مقایسهٔ یک گروه با یک عدد ثابت به حالت های مختلف قابل بررسی می باشد. برای تبیین بیشتر به مثال های زیر توجه نمایید:

- هر گاه انحراف معیار جامعه معلوم باشد:

مثال : اگر استدلال و فرض شود که میانگین بارش ایران *240* میلی متر در سال *1388* باشد ، برای بررسی این استدلال به طور تصادفی از میان ایستگاه های اقلیم شناسی و سینوپتیکی ایران *100* ایستگاه انتخاب شده است نتایج نشان می دهد که میانگین بارش ایران *230* میلی متر با انحراف معیار *20* میلی متر است . حال مقدار آزمون *t* آن را حساب کنید :

$$t = \frac{230 - 240}{20 / \sqrt{100}} = -5$$

$$df = 100 - 1 = 99$$

چون مقدارقدر مطلق *t* (*5*) محاسبه شده با درجه آزادی *99* و سطح معنی داری *0l05* از مقدار بحرانی جدول (*1l98*) بزرگتر است ، فرض صفر رد می شود و نتیجه گرفته می شود با احتمال *95* درصد اطمینان اختلاف معنی داری بین دو متغیر وجود دارد .

---

1 - One Sample t-test

2 - Paired Sample t-test

3 - Two Independent Sample t-test

- هرگاه انحراف معیار جامعه مجهول باشد و انحراف معیار نمونه از یک نمونهٔ بزرگ (بیش از *30*) بدست آمده باشد.

مثال :

میزان بارش در *40* ایستگاه اندازه گیری شده است که میانگین و انحراف معیار به ترتیب *200* و *15* میلی متر بوده است . بررسی کنید آیا میانگین بارش ایستگاه مورد نظر *220* میلی متر است ؟

$$H_O : \mu = 220$$
$$H_\alpha : \mu \neq 220$$

$$t = \frac{200 - 220}{15/\sqrt{40}} = -8.43$$

با توجه به قدر مظلق *t* محاسبه شده (*8l43*) با درجه آزادی *39* از مقدار ضریب بحرانی جدول *2l021*) در سطح *0l05* بزرگتر است لذا فرض صفر رد می شود و نتیجه گرفته می شود که میانگین بارش ایستگاه مورد نظر *220* میلی متر نمی باشد.

- اگر انحراف معیار جامعه مجهول بوده و انحراف معیار نمونه از یک نمونه کوچک برآورد شده باشد .

مثال : میزان بارش در *10* ایستگاه به شرح جدول زیر می باشد . فرض مساوی بودن میانگین بارش را با عدد *150* میلی متر در سطح *5* درصد بسنجید:

ایستگاه	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
بارش (میلی متر)	*120*	*150*	*200*	*125*	*100*	*220*	*120*	*135*	*140*	*110*

$$H_O : \mu = 150$$
$$H_\alpha : \mu \neq 150$$

$$\bar{X} = 142$$
$$S = 38.887$$

$$t = \frac{142 - 150}{38.887/\sqrt{10}} = -0.65$$

چون مقدار *t* محاسبه شده (*0l65*)از مقدار ضریب بحرانی جدول (*2l262*) کوچکتر است فرض صفر رد نمی شود لذا با احتمال *95* درصد اطمینان نتیجه گرفته می شود میانگین بارش ایستگاه مورد نظر *150* میلی متر است .

استفاده از نرم افزار *SPSS* جهت آزمون *t* با یک مقدار ثابت :

One-Sample Statistics

	N	*Mean*	*Std. Deviation*	*Std. Error Mean*
VAR00002	*10*	*142.0000*	*38.88730*	*12.29724*

One-Sample Test

	Test Value = 150					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
VAR00002	-.651	9	.532	-8.00000	-35.8183	19.8183

- کاربرد آزمون *t* با نمونه های وابسته در تحلیل های استنباطی:

از این آزمون در تحلیل های جغرافیایی استفاده می شود که هر نمونه در دو ضعیت متفاوت مورد سنجش قرار گیرد. یعنی میانگین نمونه در دو مرحلهٔ متفاوت معنی دار مورد سنجش و بررسی قرار داده می شود. در این آزمون سنجیده می شود که آیا میانگین نمونه در دو نوبت تفاوت معنی داری با یکدیگر دارد یا خیر؟ برای استفاده از این آزمون باید به معیارهای آن یعنی ، بهنجاری جامعه ، تصادفی بودن انتخاب نمونه ها ، متغیر های مستقل(مقیاس اسمی) و متغیر وابسته(فاصله ای یا نسبی) بودن و در هر نمونهٔ آماری دو اندازه گیری وجود داشته باشد ، توجه نمود. برای استفاده از این آزمون می توان از رابطهٔ زیر استفاده کرد[1]:

$$t = \frac{\sum d}{\sqrt{\frac{N\sum d^2 - (\sum d)^2}{N-1}}}$$

d تفاضل نمره ها و N حجم نمونه می باشد. برای تبیین بیشتر موضوع به مثال زیر توجه شود:

مثال :

بررسی شود میانگین رطوبت نسبی در ایستگاه های مختلف استان فارس در طی سال های *1370* تا *1375* با میانگین رطوبت نسبی در همان ایستگاه ها در طی سال های *1375* تا *1380* تفاوت معنی داری دارد یا خیر ؟ لذا *15* ایستگاه به طور تصادفی در ایران انتخاب شده است که نتایج و مراحل محاسبهٔ آزمون در جدول زیر مشخص شده است :

نام ایستگاه ها	میانگین درصد رطوبت نسبی در دورهٔ اول در ایستگاه ها	میانگین درصدرطوبت نسبی در دورهٔ دوم در ایستگاه ها	$d = x_1 - x_2$	d^2
1	***45***	***55***	***-10***	***100***
2	***35***	***35***	***0***	***0***
3	***30***	***45***	***-15***	***225***
4	***55***	***45***	***10***	***100***

[1] - جاوری ، مجید و... (1389) : روش تحقیق در جغرافیا ، انتشارات دانشگاه پیام نور

5	60	50	10	100
6	55	60	-5	25
7	70	75	-5	25
8	45	45	0	0
9	65	40	25	625
10	50	65	-15	225
11	40	55	-15	225
12	50	35	15	225
13	70	35	35	1225
14	55	65	-10	100
15	65	60	5	25
جمع			25	3225

$$t = \frac{25}{\sqrt{\frac{15\times 3225-(25)^2}{15-1}}} = 0.428$$

$$df = 15-1=14$$

چون مقدار t محاسبه شده (0/428) در سطح معنی داری 0/05 از مقدار ضریب بحرانی جدول (2/145) کوچکتر است ، لذا فرض صفر رد نمی شود و با احتمال 95 درصد اطمینان نتیجه گرفته می شود بین میانگین درصد رطوبت نسبی دو دوره در ایستگاه های مورد نظر اختلاف معنی داری وجود ندارد .

- آزمون t برای مقایسهٔ میانگین دو گروه مستقل :

وقتی بین نمونه ها ارتباطی وجود نداشته باشد آن نمونه ها یا گروه ها مستقل هستند. فرض صفر در آزمون t با دو نمونهٔ مستقل بدین معنی است که تفاوتی بین مقادیر دو نمونه در جامعه وجود ندارد و فرض تحقیق آن دلالت بر وجود اختلاف معنی داری دو نمونه دارد . معیار های کاربرد این آزمون بهنجاری جامعه ، تصادفی بودن انتخاب نمونه ها ، متغیر وابسته در سطح مقیاس فاصله ای یا نسبی بودن و برابر واریانس دو نمونه می باشد. برای استفاده از این آزمون می توان از رابطهٔ زیر استفاده کرد:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{\sum(X_1-\bar{X}_1)^2+\sum X_2-\bar{X}_2)^2}{n_1+n_2-2}\right]\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}$$

$$df = n_1+n_2-2$$

$\bar{X}_1$ $\bar{X}_2$ **به ترتیب میانگین نمونهٔ دوم و اول n_1 و n_2 به ترتیب حجم نمونهٔ اول و دوم است . برای تبیین موضوع به مثال زیر اشاره می شود.**

مثال : جغرافیدانی برای تحقیق خود درصدد بررسی تفاوت معنی داری بین میانگین سرعت تردد حمل و نقل شهری است ، برای اثبات موضوع تحقیق ایستگاه های سرعت ترددی را به طور تصادفی انتخاب نموده است که نتایج آن در جدول زیر تنظیم شده است ، حال

سؤال این است که آیا می توان اثبات کرد بین میانگین سرعت تردد در محورهای مختلف حمل و نقل شهری تفاوت معنی داری وجود دارد یا خیر؟

میانگین سرعت تردد در محور اول				میانگین سرعت تردد در محور دوم			
ایستگاه ها	سرعت	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	ایستگاه ها	سرعت	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	10	-3/375	11/391	1	10	-3/278	10/745
2	12	-1/375	1/891	2	11	-2/278	5/189
3	13	-0/375	0/141	3	12	-1/278	1/633
4	14	0/625	0/391	4	13	-0/278	0/077
5	15	1/625	2/641	5	14	0/722	0/521
6	16/25	2/875	8/266	6	13	-0/278	0/077
				7	15	1/722	2/965
				8	14	0/722	0/521
				9	17/5	4/222	17/825
جمع	80/25	0	24/721		119/5	0	39/553

$$\bar{X}_1 = \frac{80.25}{6} = 13.375$$
$$\bar{X}_2 = \frac{119.5}{9} = 13.278$$
$$S_1^2 = \frac{24.721}{6} = 4.12$$
$$S_2^2 = \frac{39.553}{9} = 4.39$$

تقریبا واریانس نمونه ها (S^1 و S^2) برابر است . لذا مقدار آزمون برابر است با :

$$t = \frac{13.375 - 13.278}{\sqrt{\left[\frac{24.721 + 39.553}{6+9-2}\right]\left[\frac{1}{6} + \frac{1}{9}\right]}} = \frac{0.097}{1.172} = 0.083$$
$$df = 6 + 9 - 2 = 13$$

چون مقدار t محاسبه شده (0/08) در سطح معنی داری 0/05 با درجه آزادی 13 از مقدار ضریب بحرانی جدول (2/160) کوچکتر است فرض صفر رد نمی شود و با احتمال 95 درصد اطمینان نتیجه گرفته می شود اختلاف معنی داری بین دو محور سرعت تردد وجود ندارد .

کاربرد آزمون ها برای مقایسهٔ چند گروه از متغیر های اقلیمی :

در بیشتر موارد اقلیم شناسان با بیش از چند گروه ها از متغیرها روبرو می باشند. آنها می خواهند تغییرات را بین متغیرهای مختلف بررسی کنند. برای مثال اقلیم شناس می خواهد بررسی کند بین 5 ایستگاه کدام یک از نظر اثرکذاری برای میزان بارش دخالت داشته است. برای تحلیل چنین موضوعی لازم است از روش تحلیل واریانس استفاده کند. در تحلیل واریانس می توان میزان انحراف کل در مجموع متغیرها را به عوامل دیگر نسبت داد . برای مثال در مقایسهٔ مقدار عملکرد تولید یک گیاه در برابر عناصر مختلف اقلیمی چگونه انحراف دارد. فرض های لازم در کاربرد تحلیل واریانس را باید محقق به آن توجه نماید . این فرض های عبارتند از :

- **نرمال یا بهنجار بودن**
- **برابری واریانس ها**
- **استقلال میانگین و واریانس ها**
- **جمع پذیری تحلیل واریانس ها**

تحلیل واریانس یک طرفه :

در صورتی که متغیرها در چند گروه وجود داشته باشد و بر اساس یک ویژگی مشخص طبقه بندی شده و مجموع مربعات نیز به دو جزء بین گروه ها و خطای بررسی ها تفکیک شوند ، برای مقایسه گروه ها تحلیل واریانس یک طرفه به کار می رود. فرض کنید تعداد ایستگاهی (n) از کل ایستگاه های یک منطقه (k) با واریانس های برابر ، به طور تصادفی استخراج شده اند از آزمون F می توان استفاده کرد. قاعدهٔ آزمون به صورت زیر مورد نظر است :

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$
$$H_\alpha : \mu_1 \neq \mu_{2\neq} \ldots \neq \mu_k$$

آزمون :

منبع تغییر	درجه آزادی	مجموع مربعات	میانگین مربعات	
بین گروه ها	K-1	$N\sum(\bar{X}_{i1}-\bar{X}_T)^2$	$MSR=\frac{N\sum(\bar{X}_{i1}-\bar{X}_T)^2}{K-1}$	$F=\frac{MSR}{MSE}$
داخل گروه ها	N-K	$\sum_i\sum_J(X_{ij}-\bar{X}_{t1})^2$	$MSE=\frac{\sum_i\sum_J(X_{ij}-\bar{X}_{t1})^2}{N-K}$	
جمع	N-1	$\sum_i\sum_J(X_{ij}-\bar{X}_T)^2$		

مثال :

در جدول زیر میزان بارش یک منطقه را در چهار سطح رطوبت نسبی نشان داده شده است . آیا بین سطوح مختلف رطوبت نسبی اختلاف معنی دار وجود دارد ؟

سطح رطوبت نسبی(درصد)	10	30	50	60
میزان بارش (سانتی متر)	2	3	4	5

	4	3	3	4
	3	4	3	4
	2	3	5	6
جمع ستونی	11	13	15	19
میانگین ستونی	2/75	3/25	3/75	4/75
مجموع متغیرها	58			

مراحل محاسبه آزمون :

- **تعیین مقدار تصحیح تحلیل واریانس**
- **محاسبهٔ مقادیر مربعات کل ، تیمار(رگرسیون) و مانده ها**
- **محاسبهٔ میانگین های مربعات**
- **محاسبهٔ مقدار آزمون F**
- **نتیجه گیری**

انجام مراحل :

$$CF = \frac{X_T^2}{n} = \frac{58^2}{16} = 210.25$$

$$SST = \sum_j^i (X_{ij}^2 - CF) = (2^2 + 4^2 + ...4^2 + 6^2) = 228 - 210.25 = 17.75$$

$$SSR = \frac{\sum_J x_{1j}^2}{n} - CF = \frac{11^2 + 13^2 + 15^2 + 19^2}{4} - 210.25 = 219 - 210.25 = 8.75$$

$$SSE = SST - SSR = 17.75 - 8.75 = 9$$

منبع تغییر	درجه آزادی	مجموع مربعات	میانگین مربعات	F
بین گروه ها	4-1=3	8/75	2/917	3/889
داخل گروه ها	16 -4=12	9	0/75	
جمع	16 -1=15	17/75		

تصمیم گیری (نتیجه گیری) :

با توجه درجه آزادی مثال (3) صورت و درجه آزادی مخرج کسر (12) مقدار ضریب بحرانی جدول برابر با 3/49 برای سطح اطمینان 5 درصد می باشد . چون مقدار آزمون محاسبه شده (3/889) از مقدار ضریب جدول بحرانی (3/49) بیشتر است ، لذا فرض صفر رد می شود و نتیجه گرفته می شود که بین درصد رطوبت نسبی و میزان بارش اختلاف معنی داری وجود دارد .

محاسبهٔ آزمون F با نرم افزار SPSS :

ANOVA

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Between Groups	*8.750*	*3*	*2.917*	*3.889*	*.037*
Within Groups	*9.000*	*12*	*.750*		
Total	*17.750*	*15*			

تحلیل واریانس دو طرفه :

در تحلیل واریانس دوطرفه بررسی اثر دو عامل در ایجاد تغییرات مورد توجه قرار می گیرد . برای مثال تأثیر رطوبت نسبی و دما در میزان بارش هر منطقه نمونه ای از تحلیل واریانس دو طرفه می باشد . در تحلیل واریانس دوطرفه اثر متقابل متغیرها مد نظر می باشد . این بدین معنی نیست که دو عامل مورد بررسی مستقل نیستند . بر این اساس می توان در تحلیل ها و سنجش سری ها از تحلیل واریانس دو طرفه استفاده کرد.

فصل پنجم : کاربرد همبستگی و رگرسیون در تحلیل های اقلیمی :

کاربرد همبستـــــــگی ها در تحلیل های اقلیمی[1] :

در مطالعات اقلیمی بررسی رابطهٔ دو متغیر از طریق توزیع نقاط از اهمیت زیادی برخوردار است . شاخص آماری که رابطهٔ دو متغیر را در توصیف می کند ، ضریب همبستگی می باشد.این ضریب در مطالعات اندازهٔ ارتباط بین دو متغیر را نشان می دهد. در تحقیقات اقلیم شناسی این شاخص نشان می دهد که تغییرات بوجود آمده در دو متغیر اقلیمی تا چه میزان با هم ارتباط دارند. هرچه بین تغییرات دو متغیر هماهنگی بیشتر باشد ، قدرت ارتباط و همچنین ضریب همبستگی بین آنها بیشتر خواهد بود. انواع مختلفی همبستگی وجود دارد که بسته به ماهیت داده ها کاربرد آنها متفاوت می باشد. درمطالعات اقلیمی وقتی ضریب همبستگی جهت توصیف رابطهٔ خطی استفاده می شود ضریب همبستگی مقادیری از صفر تا یک خواهد داشت . برای مثال رابطهٔ خطی بین میزان بارش با مقدار ارتفاع هدف تحقیق می باشد. رابطهٔ بین این دو متغیر به صورت مثبت (افزایشی دو متغیر) منفی (افزایش یک متغیر به ازای کاهش متغیر دیگر) مد نظر قرار می گیرد. به هر حال هرگاه هدف بررسی ارتباط بین دو متغّیر باشد ، باید از همبستگی سری ها استفاده کرد . دامنهٔ ضریب همبستگی $\pm$ یک است . به طوری که :

- ضریب همبستگی منفی به معنی رابطهٔ معکوس میان دو متغّیر است .
- ضریب همبستگی مثبت به معنی رابطهٔ مستقیم بین دو متغّیر است .
- ضریب همبستگی صفر به معنی عدم رابطه بین دو متغّیر می باشد .

با توجه به موضوع تحقیق ، تفسیر ضریب همبستگی در تحقیقات متفاوت است . برای محاسبه ضریب همبستگی متناسب با فرض تحقیق و متغّیر های مورد استفاده باید از نوع خاصی ضریب همبستگی استفاده کرد . جدول شمارهٔ ***4- 1*** روش های مختلف محاسبهٔ همبستگی رادر بین متغّیرهای مختلف نشان می دهد .

جدول شمارهٔ ***4- 1*** : روش های مختلف محاسبهٔ همبستگی بین متغّیرهای مختلف:

روش	نماد	متغّیر اول	متغّیر دوم	ملاحظات
همبستگی پیرسون	r	پیوسته	پیوسته	دقیق ترین روش با کمترین خطای معیار
رتبه ای	R_{ho}	رتبه ای	رتبه ای	غالبا برای تعداد آزمودنی های کمتر از ***30*** نفر استفاده می شود.
کندال	τ	رتبه ای	رتبه ای	وقتی تعداد آزمودنی ها کمتر از ***10*** نفر است استفاده می شود.
دو رشته ای	r_{bis}	دو ارزشی	پیوسته	مقدار آن گاهی از یک بیشتر می شود

[1] - برای اطلاع بیشتر به کتاب مقدمه ای بر روش های تحقیق در علوم انسانی با تأکید بر جغرافیا از مؤلف مراجعه شود.

		ساختگی		
دو رشته ای گسترده	r_{wbis}	دو ارزشی ساختگی	پیوسته	زمانی بکار می رود که محقق به محاسبهٔ نمره های ازمودنی هایی علاقمند است که نمره های آنها دو ارزشی در حد خیلی بالا قرار دارد .
دو رشته ای نقطه ای	r_{pbis}	دو ارزشی واقعی	پیوسته	ضریب غالبا" کمتر از ضریب دورشته ای است
تتراکوریک	r_t	دو ارزشی ساختگی	دو ارزشی ساختگی	زمانی بکار می رود که بتوان دو متغّیر را در یک نقطهٔ معین نصف کرد.
فی	ϕ	دو ارزشی واقعی	دو ارزشی واقعی	برای محاسبهٔ همبستگی درونی سؤال ها بکار می رود
توافقی	C	به صورت دو یا چند طبقه	به صورت دو یا چند طبقه	در برخی از شرایط با همبستگی گشتاوری(پیرسون) یکسان است .
نسبی	η	پیوسته	پیوسته	به منظور تعیین همبستگی غیر خطی بکار می رود.

برای محاسبهٔ ضریب همبستگی از روش های مختلفی استفاده می شود که در اینجا به چند روش آن اشاره خواهد شد .

ضریب همبستگی پیرسون :

هرگاه داده های بدست آمده با استفاده از مقیاس های فاصله ای و نسبتی اندازه گیری شده باشند ، برای محاسبهٔ رابطهٔ بین متغّیرها از ضریب همبستگی پیرسون که بعنوان ضریب همبستگی از نوع پارامتری یا عاملی قلمداد می شود می توان استفاده کرد . برای محاسبهٔ ضریب همبستگی پیرسون (داده های خام) از معادلهٔ ذیل استفاده می شود :

$$r_{xy} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (y)^2\right]}}$$

مثال :

ضریب همبستگی داده های جدول ذیل را بدست آورید :

X	***Y***
1	***3***
3	***5***
2	***2***
0	***4***
4	***1***
5	***6***

متناسب با فرمول :

x	***y***	***xy***	x^2	y^2
1	***3***	***3***	***1***	***9***

3	**5**	**15**	**9**	**25**
2	**2**	**4**	**4**	**4**
0	**4**	**0**	**0**	**16**
4	**1**	**4**	**16**	**1**
5	**6**	**30**	**25**	**36**
$\sum$ **15**	**21**	**56**	**55**	**91**

$$r_{xy} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (y)^2\right]}} = \frac{6\times 56 - (15)(21)}{\sqrt{(330-225)(546-441)}}$$

$$= 0.2$$

به هر حال میزان ضریب همبستگی را نسبت واریانس ها می نامند که نسبت واریانس مشترک بین متغیرها را به کل واریانس می توان در نظر گرفت . یعنی :

$$r_{xy} = \frac{\left(Cov_{xy}\right)^2}{V_x^2 V_y^2}$$

$\left(Cov_{xy}\right)^2$ واریانس مشترک بین متغیرهای $\boldsymbol{x}$ و$\boldsymbol{y}$ و $V_x^y V_y^2$ کل واریانس متغیر ها می باشد. چنانچه نمرات استاندارد شده را داشته باشیم ، محاسبهٔ ضریب همبستگی آن از طریق معادلهٔ ذیل بدست می آید :

$$r_{xy} = \frac{\sum z_x . z_y}{N}$$

$\boldsymbol{Z}$ ها نمرات استاندارد شدهٔ $\boldsymbol{x}$ و $\boldsymbol{y}$هستند . برای استاندارد کردن نمرات همان گونه که قبلا" اشاره شد می توان از رابطهٔ زیر استفاده کرد :

$$z = \frac{x-\mu}{\sigma}$$

مثال :

ضریب همبستگی داده های جدول ذیل را با استفاده از روش استاندارد بدست آورید :

X	**Y**
1	**3**
3	**5**
2	**2**
0	**4**
4	**1**
5	**6**

با توجه به معادلهٔ :

$$\bar{X} = \frac{\sum x_i}{N} = \frac{15}{6} = 2.5$$

$$\bar{Y} = \frac{\sum y_i}{N} = \frac{21}{6} = 3.5$$

$$\sigma_X = \sqrt{\frac{\sum x^2}{N} - (\frac{\sum x}{N})^2} = \sqrt{\frac{55}{6} - (\frac{15}{6})^2} = \sqrt{9.17 - 6.25} = 1.71$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{N} - (\frac{\sum y}{N})^2} = \sqrt{\frac{91}{6} - (\frac{21}{6})^2} = \sqrt{15.17 - 12.25} = 1.71$$

سپس :

$$z_x = \frac{x - \mu_x}{\sigma_x}$$

$$z_y = \frac{y - \mu_y}{\sigma_y}$$

x	y	z_x	Z_y	$z_x.z_y$
1	**3**	**-0/88**	**-0/29**	**0/255**
3	**5**	**0/29**	**0/88**	**0/2552**
2	**2**	**-0/29**	**-0/88**	**0/2552**
0	**4**	**-1/46**	**0/29**	**-0/4234**
4	**1**	**0/88**	**-1/46**	**-0/2848**
5	**6**	**1/46**	**1/46**	**2/136**
جمع				**1/189**

$$r_{xy} = \frac{\sum z_x . z_y}{N} = \frac{1.189}{6} = 0.1982 \approx 0.2$$

ضـــریب هـمبستـــگی رتبــه ای اســپیــرمـــن :

وقتی مقیاس اندازه گیری داده ها رتبه ای باشد ، از ضریب همبستگی اسپیرمن برای محاسبه رابطهٔ متغّیرها استفاده می شود . برای محاسبهٔ ضریب همبستگی اسپیرمن از معادلهٔ زیر استفاده می شود :

$$r_{ho} = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

D تفاوت بین رتبه ها است .

مثال :

اگر داده های جدول زیر را که نتیجهٔ رتبه دانشجویان در دو درس است ، داشته باشیم ، ضریب همبستگی آنها را محاسبه کنید ؟

شماره افراد	رتبه در درس اقلیم	رتبه در درس آمار
1	**1**	**2**
2	**2**	**3**
3	**3**	**1**
4	**4**	**5**
5	**5**	**4**

متناسب با فرمول :

شماره افراد	رتبه در درس اقلیم	رتبهٔ در درس آمار	تفاضل رتبه ها(D)	D^2
1	1	2	-1	1
2	2	3	-1	1
3	3	1	2	4
4	4	5	-1	1
5	5	4	1	1
جمع				8

$$r_{ho} = 1 - \frac{6\sum D^2}{N\ (N^2 - 1)} = 1 - \frac{6 \times 8}{5(25-1)} = 1 - \frac{48}{120} = 0.6$$

ضریب تعیین :

ضریب همبستگی فقط بالا و پایین بودن ضریب را نشان می دهد . برای اینکه بتوان فهمید چند درصد از واریانس متغیر x ناشی از واریانس متغّیر y است لازم است از ضریب تعیین استفاده شود . اگر بر حسب درصد بخواهیم ضریب تعیین را حساب کنیم می توان از رابطهٔ ذیل استفاده کرد :

$$D = (r_{xy})^2 \times 100$$

اگر مثال بالا را در نظر بگیرم ، $D = (.6)^2 \times 100 = 36$ (**36**)درصد پراکندگی بین دو متغّیر مشترک است . قابل اشاره است که ضریب تعیین هیچگاه صفر نخواهد بود و دامنهٔ ضریب تعیین بین صفر و یک می باشد هر چقدر مقدار ضریب تعیین به یک نزدیکتر باشد میزان وابستگی متغیر وابسته با متغیر یا متغیرهای مستقل بیشتر است و برعکس .

معنی دار بودن ضریب همبستگی :

وقتی بخواهیم معنی دار بودن ضریب همبستگی را بررسی کنیم می توان از فرمول زیر استفاده کرد :

$$t = r\sqrt{\frac{N-2}{1-r^2}}$$

نتیجهٔ بدست آمدهٔ t با توجه به درجهٔ آزادی (N-2) با مقدار t در جدول آماری t مقایسه می شود و در صورت بزرگ بودن مقدار محاسبه شده از مقدار جدول ، نتیجه گرفته می شود که همبستگی بین دو متغّیر در سطح اطمینان (0/05 و 0/01) معنی دار است .

کاربرد رگرسیون در تحلیل های اقلیمی[1] :

ماهیت موضوعات اقلیمی و داده های مورد استفاده در این رشته علمی ضرورت توجه بیشتر به فنون آماری را می طلبد . یکی از فنون بسیار مهم در مطالعات اقلیمی تحلیل های رگرسیونی است . در مطالعات ااقلیمی بعضی مواقع هدف پیش بینی یک متغیر بوسیلهٔ متغیر دیگر است . لذا برای پیش بینی یک متغیر توسط متغیر دیگر باید به یک رابطهٔ منظم و نظامدار بین آن دو متغیر توجه نمود . بر این اساس در تحلیل

[1] - به مقالهٔ **تجزیه و تحلیل داده های اقلیمی با استفاده از تحلیل رگرسیونی ، از مؤلف مراجعه شود.**

رگرسیونی توجه به تغییر در متغیر های مستقل که مقادیر متغیرهای وابسته را چگونه تحت تأثیر قرار می دهد ، حائز اهمیت است. وجه تمایز بین متغیرهای مستقل و وابسته همیشه کاملا" روشن نمی باشد و آن بستگی به اهداف تحقیق دارد . واژۀ رگرسیون[1] به معنای برگشت است و نشان می دهد که مقدار یک متغیر به متغیر دیگری بر می گردد . این واژه اولین بار توسط فرانسیس گالتن [2] در سال 1877 به کار برده شده است. با توجه تعداد متغیر ها ، در مطالعات اقلیمی علاوه بر رگرسیون خطی ساده از رگرسیون چندگانه استفاده می شود . رگرسیون یا تحلیل کمترین مربعات یک شیوۀ آماری برای برآورد رابطۀ تابعی بین یک متغیر وابسته و یک یا چند متغیر مستقل یا پیش بینی کننده [3]است . در تحلیل رگرسیون متغیرهای توضیحی ممکن است غیر قابل اندازه گیری باشند که به آنها متغیرهای کیفی گفته می شود . یکی از راههای واردکردن متغیرهای کیفی درمدل رگرسیون استفاده از متغیرهای مجازی[4] است . لذا در تحلیل رگرسیون به تعیین روابط بین متغیر ها (رگرسیون توصیفی) و تحلیل روابط بدست آمده (رگرسیون استنباطی) توجه می شود . شکل عمومی رگرسیون متغیر وابسته و متغیرهای مستقل ممکن است یک سری زمانی باشند و یا هم نباشند . در تحلیل این روش باید به فرض های زیر توجه داشت :

1- رابطۀ بین متغیرهای وابسته و مستقل خطی باشد .
2- خطاها غیر همبسته باشند .
3- خطاها توزیع بهنجار داشته باشند .
4- میانگین خطاها صفر باشد .
5- واریانس خطاها ثابت باشد .

در تحلیل رگرسیون می توان از نمودارهای احتمال بهنجار ، نمودارهای پراکنش مانده ها در برابر سری پیش بینی شده ، نمودار پراکنش مانده ها در برابر متغیرهای مستقل و آزمون های آماری استفاده نمود . مدل رگرسیون زمانی صحیح خواهد بود که مانده ها در اطراف صفر بطور تصادفی توزیع شده باشند . وقتی در یک مدل رگرسیون از یک متغیر مستقل استفاده می شود ، مدل مذکور را مدل رگرسیون خطی ساده[5] گویند و وقتی بیش از یک متغیر مستقل در یک مدل استفاده می شود ، به آن مدل ،مدل رگرسیون چندگانه[6] گفته می شود . لازم است برای بررسی رگرسیونی رابطۀ بین سری ها و طول سری های بررسی شود . برای تحلیل رگرسیونی داده های اقلیمی باید به مراحلی همچون بیان مسئله ، انتخاب متغیرها ، جمع آوری سری یا داده ها ، تشخیص و تعیین مدل ، انتخاب روش برازش ، مناسب بودن مدل و کاربرد مدل برای

---

[1] - ***Regression***
[2] - ***Francis Galton***
[3] - ***Predictor Variable***
[4] - ***Dummy Variable***
[5] - ***Simple Linear Regression***
[6] - ***Multiple Regression***

تحلیل داده های اقلیمی باید توجه نمود.در تعیین مدل های رگرسیونی باید به الگوهای خطی و غیر خطی توجه نمود. مهمترین جنبهٔ کاربرد تحلیل رگرسیون در مطالعات اقلیم شناسی بررسی و مدل سازی رابطهٔ بین متغیر ها می باشد. برای بررسی رابطهٔ سری ها لازم است از آزمون های فرض شیب سری ها در زیر استفاده نمود :

$$t = \frac{\hat{\beta}_t \sqrt{SS_{xx}}}{S_\varepsilon} \qquad (1\text{-}4)$$

$$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (2\text{-}4)$$

$$F = \frac{SSR}{SSE\ /(n-2)} \qquad (3\text{-}4)$$

قواعد آزمون ها :

$$\text{Re}\,ject : H_o if\ |t| > t_{\alpha/2}$$

$$\text{Re}\,ject : H_o if\ |t_r| > t_{\alpha/2}$$

$$\text{Re}\,ject : H_o if : F > F_\alpha$$

اگررابطهٔ قابل ملاحظه ای بین متغیر ها وجود داشته باشد ، برآورد حدود اطمینان برای شیب خطی لازم و ضروری است ، برای محاسبهٔ حدود شیب خطی از رابطهٔ زیر استفاده می شود :

$$\hat{\beta}_1 \pm t_{\alpha/2} \times \frac{S_\varepsilon}{\sqrt{SS_{xx}}} \qquad (4\text{-}4)$$

4-2- کاربرد رگرسیون خطی ساده :

در رگرسیون خطی ساده هدف این است که با محاسبهٔ عامل های β_0 و β_1 ، مدل خطی ساده ای را به داده های برازش اختصاص داده شود. روشی که برای برازش کردن داده ها بکار می رود روش کمترین مربعات یا **LS** می باشد . در این روش معادلهٔ خط برازش طوری شکل می گیرد که مجموع مربعات توان های دوم انحرافات عمودی از خط برازش شده به حداقل برسد. برآورد ضرایب به روش حداقل مربعات صورت می گیرد . برای محاسبهٔ ضرایب می توان از معادلات زیر استفاده کرد :

$$\hat{\beta}_O = \bar{Y} - \hat{\beta}_1 \bar{X} \qquad (5\text{-}4)$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \qquad (6\text{-}4)$$

$$SS_{xx} = \sum X^2 - \frac{(\sum X)^2}{N} \qquad (7\text{-}4)$$

$$SS_{xy} = \sum xy - \frac{(\sum X) \times (\sum Y)}{N} \qquad (8\text{-}4)$$

$$SS_{yy} = \sum Y^2 - \frac{(\sum Y)^2}{N} \qquad (9\text{-}4)$$

β_o عرض از مبدأ ، β_1 شیب خط ، Y متغیر وابسته و X متغیر مستقل می باشد . با توجه معادلات بالا مدل رگرسیون خطی ساده را می توان به صورت معادلهٔ زیر نشان داد :

$$Y = \beta_0 + \beta_1 x + \varepsilon \qquad (10\text{-}4)$$

در معادلهٔ بالا ε عامل خطا می باشد . با توجه به مدل بالا معادلهٔ خط برازش شده را می توان بصورت معادلهٔ زیر نوشت :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad (11\text{-}4)$$

مقدار خطا حاصل از تفاضل مقدار اولیه و مقدار پیش بینی شده می باشد. به طریقی از حاصل تفاضل مقادیر پیش بینی شده از مقدار اصلی سری مقدار خطاها یا مانده ها[1] بدست می آید .

[1] - ***Residual***

در بررسی مدل رگرسیونی خطی ساده فرض می شود که :

1 -مانده ها دارای توزیع بهنجار یا نرمال اند .
2 -واریانس سری ثابت است .
3 -متغیر ها مستقل از یکدیگرند .

جدول شمارۀ4- 1 ضرایب معادلۀ خط رگرسیون همراه با انحراف معیار برآوردها ؛ ضرائب استاندارد شده و آماره های ***t*** و ***P*** که برای آزمون فرض های آماری بکار می رود را نشان می دهد

جدول شمارۀ 4-1 :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 معادلۀ رگرسیون برابر است با :

$$y = 18.7 + 0.00856\ t$$

حال اگر فرض شود که $H_o : \beta_o = 0$ و $H_o : \beta_1 = 0$ انجام شود ، لذا :

- آزمون *1* :

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

برای این آزمون از آمارۀ t استفاده می شود . 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 می شود . به عبارت دیگر عرض از مبدأ معادلهٔ رگرسیونی صفر نیست . روش دیگر جهت این آزمون استفاده از آمارهٔ *P-Value* است . که :

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

پس نتیجه کلی این است که فرض صفر در سطح معنی دار پنج درصد رد می شود.

- آزمون *2* :

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

مشابه آزمون *1* :

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

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

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

بنابراین فرض صفر در سطح معنی دار پنج درصد رد نمی شود ، یعنی معادلهٔ رگرسیونی بدون شیب رگرسیونی است . کل نتیجه آزمون ها را می توان به صورت جدول شمارهٔ4- 2 نشان داد .

جدول شمارهٔ4- 2 : 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 Value	*98.8957*	*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*

4-3- رگرسیون و تحلیل واریانس سری:

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. اگر در رگرسیون ، تغییرات $\boldsymbol{Y}$ را نتوان به تغییرات $\boldsymbol{X}$ نسبت داد آنگاه وجود رابطهٔ خطی بین متغیرها مردود است. برای انجام آزمون فرض در خصوص $H_0 : \beta_1 = 0$ می توان جدول تحلیل واریانس سری دمای یزد را به صورت جدول شمارهٔ4- 3 نشان داد .

جدول شمارۀ4- 3 : تحلیل واریانس سری دمای یزد :

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Probe Level	Powe (5%
Intercept	1	13841.37	13841.37			
Slope	1	0.3622045	0.3622045	1.0838	0.3046	0.173
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***) رابطۀ خطی بین متغیرهای ***Y*** و ***X*** **(*t*)** را در سطح معنی دار پنج درصد می توان با توجه به آزمون ها بررسی نمود .

- آزمون فرض ها :

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

- آمارهٔ آزمون :

$$F = 1/084$$

- ناحیهٔ بحرانی :

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

بنابراین فرض $H_O : \beta_1 = 0$ رد نمی شود . به عبارت دیگر دلایل کافی دال بر رابطهٔ خطی بین متغیرها وجود ندارد . با توجه به جدول شمارهٔ 3 مجموع مربعات مدل[1] 0/362 ، مجموع مربعات خطا[2] 12/365 و مجموع مربعات کل[3] 12/278 می باشند. از نکات قابل توجه در جدول خلاصهٔ تحلیل واریانس شاخص ضریب تعیین می باشد . جدول شمارهٔ 4 ، ضریب تعیین را نشان می دهد .

جدول شمارهٔ4- 4 : خلاصهٔ تحلیل واریانس سری دمای یزد :

---

[1] - $SS_{Regression}$

[2] - $SS_{Residual}$

[3] - SS_{Total}

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*	*0.578108*

در تحلیل رگرسیون ضریب تعیین تعدیل شده $R^2_{adj} = 1 - \frac{SS_{res}/(n-p)}{SS_T/(n-1)}$ وجود دارد که در ارزیابی و مقایسه مدل هـای رگرسـیون از آن می توان استفاده کرد .ضریب تعیین کلی $R^2 = \frac{SS_R}{SS_T}$ که مشخص می کند مدل مورد نظر تا چه حدی می توان متغیر وابسته را تخمین بزنـد را یا به طریقی معنی دار بودن مدل رگرسیون را مشخص می سازد ، را می توان استفاده کرد.دامنـهٔ ضـریب تعیـین $0 \le R^2 \le 1$ اسـت . ایـن ضریب مناسب بودن مدل را مشخص نمی کند بلکه درصد تغییرات متغیر وابسته است که بوسیلهٔ متغیر مستقل تبیین می شود . اگر در تحلیـل رگرسیون نقاط تکراری باشد ضریب تعیین یک نمی شود ، چون مدل تحلیل رگرسیون نمی تواند میزان تغییرات مربوط به مانده هـا یـا خطـا سری را تعیین کند . وقتی که مدل رگرسیون صحیح باشد مقدار ضریب تعیین بستگی به دامنهٔ تغییرات متغیر مسـتقل دارد ، یعنـی هـر چقـدر مقدار پراکنش متغیر مستقل بیشتر باشد مقدار ضریب تعیین بیشتر خواهد بود و برعکس . از طرفی ضریب تعیین دلالت بـر بـالا بـودن مقـدار شیب خط رگرسیون نمی باشد . چون مقدار ضریب تعیین مقدار شیب خط رگرسیون را اندازه گیری نمی کنـد .از طرفـی ضـریب تعیـین بـا ضریب همبستگی رابطه دارد یعنی توان دوم ضریب همبستگی ضریب تعیین است و مقدار آن با داشتن ضریب تعیین بین مقدار خطا و متغیـر وابسته برابر با $\sqrt{1-R^2}$ است . هرچه مقدار همبستگی بین متغیرها بیشتر باشد پراکنش متغیرها به خط رگرسیون نزدیکتـر مـی شـوند . اگـر قدر مطلق همبستگی کامل باشد همهٔ نقاط بر روی خط رگرسیون قرار می گیرند . با توجه به دامنهٔ ضریب تعیین مقدار آن را بصورت درصـد

بیان می کند با توجه به مثال سری دمای یزد ، ضریب تعیین سری دمای یزد نشان می دهد ه 2/8 درصد از تغییرات Y توسط X توضیح داده می شود و به طریقی تقریبا"مقدار 97.15 درصد باقی مانده معلول عوامل دیگر است که در تحلیل رگرسیون لحاظ نشده است. یعنی :

$$R-Squared = 0.0285\% = 2.85\%$$

4-4- برآورد فاصلهٔ اطمینان برای متوسط پیش بینی از یک مدل خطی ساده:

اگردر رابطهٔخطی $Y = \beta_0 + \beta_1 x + \varepsilon$ مقدار معینی از X یعنی x^* **(t)** در دست باشد ، می توان مقدار Y را به طور متوسط با توجه به رابطهٔ زیر برآورد نمود :

$$E(Y / X = x^*) = \beta_0 + \beta_1 x^* \qquad (12\text{-}4)$$

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

$$Y^* = \hat{\beta}_0 + \hat{\beta}_1 x^* \qquad (13\text{-}4)$$

متناسب با رابطهٔ بالا فاصلهٔ اطمینان به صورت زیر می باشد:

$$\hat{\beta}_0 + \hat{\beta}_1 X^* \pm t(n-2,1-\frac{\alpha}{2}) \times \hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X^*-\bar{X})^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}} \qquad (17\text{-}4)$$

با توجه به سری دمای ایستگاه یزد حدود اطمینان تا 10 سال برابر است :

	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)

4-5- تحلیل مانده ها رگرسیونی :

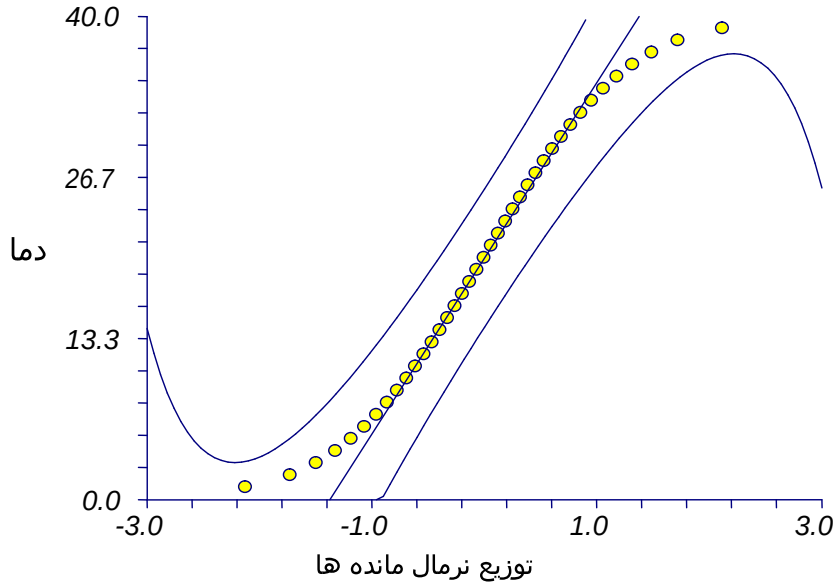
وقتی به بررسی رابطهٔ متغیرها پرداخته شد معمولا" باید محقّق به این نتیجه برسد که سری ها قابلیت تحلیل رگرسیونی دارند .محقّق باید به این نتیجه برسد که آیا رابطه ای خطی بین متغیرها وجود دارد و آیا توزیع متغیروابسته بهنجار است و یا واریانس آن برای مقادیر مختلف متغیر مستقل ثابت است. بر این اساس بخش مهمی تحلیل رگرسیونی بررسی فرض های خطی بودن ، بهنجار بودن ، ثابت بودن واریانس و مستقل بودن مشاهدات است . لذا اولین قدم در تحلیل رگرسیونی سنجش سری ها مورد بررسی است . یعنی محقّق باید ابتدا سری ها را بسنجد که آیا شرایط تحلیل و کاربرد رگرسیونی را دارند یا خیر ؟ اگر فرض پذیرفته شد در آن مرحله است که محقّق می تواند از مدل های تحلیل رگرسیونی استفاده کند. اگر فرض های تحلیل رگرسیونی صادق باشد ، مانده های تحلیل باید مشخصات زیر را داشته باشند :

1- توزیع آنها بهنجار یا نرمال باشد .
2- واریانس آنها برای تمام مقادیر متغیر مستقل ثابت باشد.
3- هنگام نمایش مانده ها در مقابل مقادیر پیش بینی شده نباید الگویی مشاهده شود.
4- مانده ها تقریبا" مستقل از یکدیگر باشند .

برای نمایش فرض های مذکور از سری دمای یزد استفاده شده است .

- پاسخ به فرض اول ، توزیع بهنجاری مانده های سری :

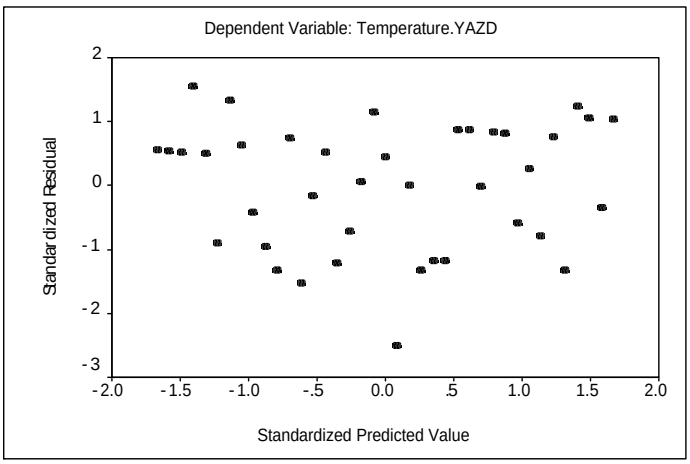
نمودارشمارهٔ *4-1* : توزیع نرمال سری دمای یزد



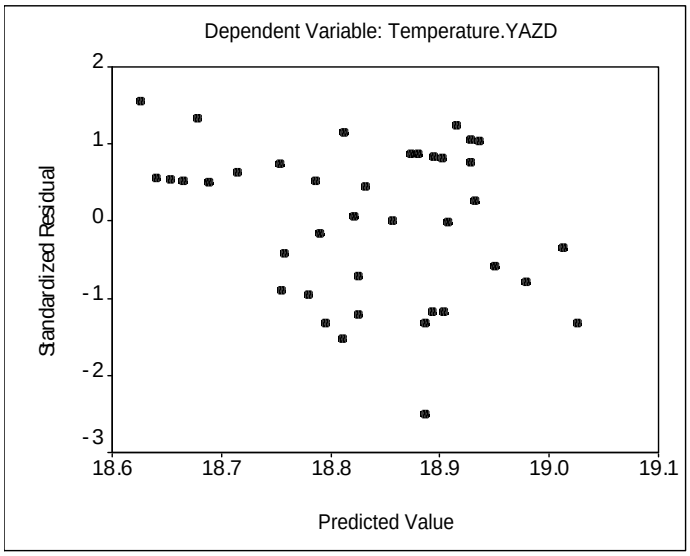
با توجه به نمودار بالا ، سری مانده های دمای یزد وضعیت بهنجاری را نشان می دهد .

- پاسخ به فرض دوم و سوم، باید واریانس آنها برای تمام مقادیر متغیر مستقل ثابت باشد و هنگام نمایش مانده ها در مقابل مقادیر پیش بینی شده نباید الگویی مشاهده شود. :

نمودار شمارهٔ ***4- 2*** : نمودار مانده های استاندارد شدهٔ دمای یزد :



نمودار شمارۀ4- 3: نمودار مانده های استاندارد شده با مقادیر پیش بینی شده دمای یزد:



با توجه به نمودارهای شمارۀ 4- 2و4-3 ، چون سری مانده ها الگوی مشخصی را در نمودار نشان نمی دهند بر این اساس واریانس مانده های سری برای متغیر مستقل تقریبا" ثابت می باشد.

- پاسخ به فرض چهارم ، باید مانده ها تقریبا" مستقل از یکدیگر باشند :

برای بررسی استقلال مانده ها از همدیگر می توان از آزمون دوربین _ واتسون [1] استفاده کرد ، اگر مقدار بدست آمده بین ***1/5*** تا ***2/5*** باشد می توان تقریبا" به استقلال مانده ها از همدیگر پی برد. با توجه به نتیجهٔ کاربرد آزمون دوربین _ واتسون می توان به مستقل بودن مانده ها پی برد. جدول شمارهٔ ***5*** نتیجه آزمون فوق را نشان می دهد.

جدول شمارهٔ***4 - 5*** : آزمون دروبین _ واتسون برای استقلال مانده های 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

4-6- کاربرد رگرسیون چندگانه برای تحلیل سری های اقلیمی :

در تحلیل های اقلیمی برای پیش بینی مقدار متغیر وابسته به بیشتر از یک متغیر مستقل نیاز می باشد . برای مثال وقتی وضعیت اقلیمی یک منطقه یا یک کشور و یا یک ایستگاه را بررسی می کند لازم است رابطهٔ متغیر های متعددی را بررسی کند ، بر این اساس یکی از فنون آماری مهم ، کاربرد رگرسیون چندگانه می باشد . در تحلیل رگرسیون چندگانه به جای یک عرض از مبدأ و ضریب زاویه ای دارای یک مقدار ثابت ، دارای چند ضریب می باشد که هر یک متعلق به یکی از چند متغیر مستقل می باشد.این ضرایب به عنوان ضرایب رگرسیون جزئی[2] نامیده می شوند. ضرایب رگرسیون جزئی در نمونه برآوردی از ضرایب نامعلوم جامعه است. همانند رگرسیون خطی ساده ، در رگرسیونی چندگانه باید به فرض های زیر باید مورد توجه قرار گیرد:

1- رابطهٔ بین متغیر وابسته و متغیرهای مستقل خطی باشد .
2- برای هر یک از ترکیب های مقادیر متغیرهای مستقل ، توزیع متغیر وابسته نرمال یا بهنجار باشد.
3- واریانس مقادیر آن ثابت باشد.

قبل از آنکه ضرایب برآورد شود باید مطمئن شد که متغیرهای مستقل به طور 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

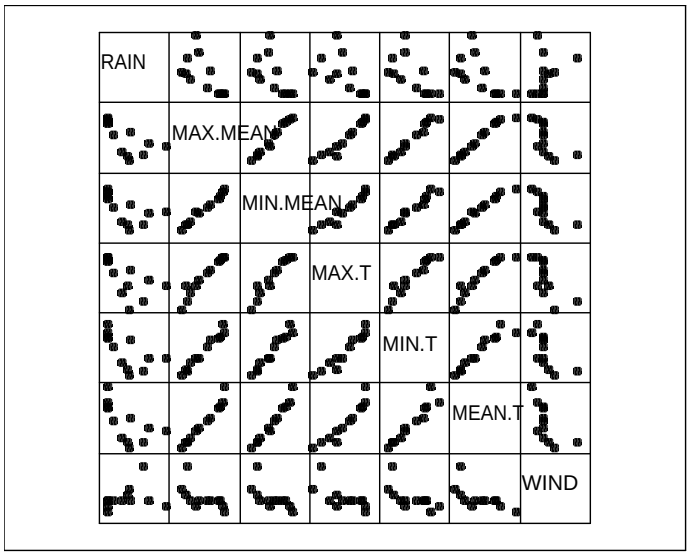
[1] - *Durbin_Watson*

[2] - ***Partial Regression Coefficients***

حداکثر ، معدل حداقل و سرعت باد) مؤثر می باشند ، استفاده شده است. از ماتریس نمودارهای پراکنش بین متغیرهای مستقل و وابسته برای بررسی رابطهٔ متغیرها استفاده شده است . برای کاربرد مدل رگرسیون چندگانه می توان از معادلهٔ زیر استفاده کرد:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \qquad (18\text{-}4)$$

جدول شمارهٔ4- 6 :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 معادلهٔ شمارهٔ(4- 18) به تحلیل رگرسیون چندگانه پرداخت .

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

4- 7 - بررسی مناسب بودن مدل :

وقتی یک مدل رگرسیون اجرا می شود اولین نکته این است که بررسی شود مدل مناسب است یا نه . برای این منظور لازم است به تحلیل رگرسیون چندگانه توجه نمود . جداول زیر تحلیل رگرسیون چندگانه متغیرهای ایجاد بارش خرم آباد را نشان می دهند .

جدول شمارهٔ **4- 7**: خلاصهٔ آمارهٔ مدل رگرسیون چندگانه :

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

جدول شمارهٔ **4- 8** : 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 شمارهٔ **4- 9** : 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 آماره های تحلیل ، ضریب تعیین تحلیل نشان می دهد که87 درصد از پراکندگی مشاهده شده ایجاد بارش خرم آباد توسط شش متغیر مستقل توجیه می شود و به طریقی متغیرهای مستقل رابطهٔ خطی با متغیر وابسته ایجاد بارش خرم آباد را نشان می دهد . از طرفی مقدار همبستگی متغیرها که برابر است با 0/93 است نشان می دهد که از مدل رگرسیون خطی مذکور به خوبی می توان برای پیش بینی استفاده نمود. با توجه به جدول شمارهٔ 8 آمارهٔ F ($F=\frac{MST}{MSE}=\frac{3111.775}{556.796}=5.59$) فرض صفر رد می شود و نتیجه گرفته می شود 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 می گیرد و اگر فرضیه صفر رد شد ، می توان نتیجه گرفت که رابطهٔ خطی بین دو متغیر مستقل و وابسته وجوددارد . تفسیر این ضریب در رگرسیون چندگانه پیچیده تر است . اگر ضرایب رگرسیون جزیی برای متغیرها بدست آید ، با بررسی ماتریسی ضرایب امکان بررسی رابطه ها وجود دارد . چون از یک طرف ضرایب همبستگی را نشان می دهد از طرفی دیگر سطوح معنی داری مشاهدات را نشان می دهد. جدول شمارهٔ10 ضرایب رگرسیونی جزیی متغیرهای مؤثر در ایجاد بارش خرم آباد را نشان می دهد.

جدول شمارهٔ4- 10 : ضرایب همبستگی جزیی متغیرهای ایجاد بارش خرم آباد:

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

ضریب همبستگی بین میزان بارش و با رطوبت نسبی 0/758 است . بر این اساس سطح معنی داری مشاهده شده می تواند این فرضیه را رد کند که بین دو متغیر رابطهٔ خطی وجود دارد. از طرفی مشاهده می شود که متغیرهای مستقل با هم رابطهٔ خوبی را نشان می دهند . قدرت رابطهٔ خطی بین متغیرهای مستقل توسط شاخصی اندازه گیری می شود که تولرانس[1] نام دارد. 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 نمی شود . بر این اساس محاسبه شاخص چند هم خطی[2] در نرم افزار مینی تب از عامل افزایش واریانس[3] (VIF) می توان استفاده کرد . جدول زیر شاخص مذکور را نشان می دهد .مقدار عامل افزایش واریانس باید بین 5 تا 10 باشد ، اگر بیشتر از این حد باشد ، امکان ضعف برآورد ضریب رگرسیون در اثر چند

---

[1] - Tolerance

[2] - ***Multicollinearity***

[3] - ***Variance Inflation Factor***

هم خطی وجود دارد . با توجه به این که تولرانس یک نسبت است ، مقدار آن بین صفر و یک می باشد . مقدار نزدیک به یک به این معنی است که در یک متغیر مستقل بخش کوچکی از پراکنش آن با سایر متغیرهای مستقل توجیه می کند و مقدار نزدیک به صفر نشان می دهد که یک متغیر تقریبا" یک ترکیب خطی از سایر متغیرهای مستقل است.

جدول شمارهٔ ***4- 11*** : آمارهٔ های تولرانس و عامل افزایش واریانس سری های مؤثر در بارش خرم آباد:

		Collinearity Statistics	
Model		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

تحلیل عاملی[1]:

در برخی تحقیقات به دلایل مختلف با حجم زیادی از متغیرها روبرو هستیم. برای تحلیل دقیق تر و رسیدن به نتایج علمی تر و در عین حال عملیاتی تر، پژوهشگر به دنبال کاهش حجم متغیرها (حذف متغیرهای اضافی) و تشکیل ساختار جدیدی(شناسایی و بررسی روابط پنهان بین متغیرها) برای آنها است و به این منظور از روش تحلیل عاملی استفاده می کند.

[1] - برای اطلاع بیشتر به کتاب مقدمه ای بر روش های تحقیق در علوم انسانی با تأکید برجغرافیا از مؤلف مراجعه شود.

نخستین کار درباره تحلیل عاملی توسط چارلز اسپیرمن (1904) صورت گرفت، به گونه ای که وی را « پدر» این روش شناخته اند. بعد از او کارل پیرسن (1910)، روش «محورهای اصلی» را پیشنهاد کرد و هُتلینگ (1933) آن را به گونه کاملتری توسعه داد

در حقیقت، هدف تشخیص این عامل های مشاهده ناپذیر بر پایه مجموعه ای از متغیرهای مشاهده پذیر است. *عامل*، متغیر جدیدی است که از طریق ترکیب خطی نمره های اصلی متغیرهای مشاهده شده بر پایه فرمول زیر برآورد می شود:

$$F_j=\sum W_{ji}X_i=W_{j1}X_1+W_{j2}X_2+...+W_{jp}X_p$$

که در آن W ها بیانگر ضرایب نمره عاملی و P معرف تعداد متغیرها است.

امید این است که با تعداد کمی از این عاملها (یعنی ترکیبهای خطی نمره های اصلی متغیرهای مشاهده شده)، بتوان تقریباً همه اطلاعاتی را که توسط مجموعه بزرگتری از متغیرها به دست می آید، در برگرفته و در نتیجه توصیف ویژگی های هر مشاهده را ساده ساخت.

تحلیل عاملی از جمله روشهای چند متغیره است که در آن ،متغیرهای مستقل و وابسته مطرح نیست زیرا این روش جزء تکنیکهای هم وابسته محسوب میگردد و کلیه متغیرها نسبت به هم وابسته اند. در این روش متغیرها در عامل هایی قرار می گیرند ، به طوری که از عامل اول به عامل های بعدی درصد واریانس کاهش می یابد ، از این رو متغیرهایی که در عامل اولی قرار می گیرند ، مؤثرترین می باشد.

بنا بر آنچه گفته شد، تحلیل عاملی تکنیکی است که کاهش تعداد زیادی از متغیرهای وابسته به هم را به صورت تعداد کوچکتری از ابعاد پنهان یا مکنون (عاملها) امکان پذیر می سازد بطوریکه در آن کمترین میزان گم شدن اطلاعات وجود داشته باشد. هدف اصلی آن، خلاصه کردن داده هاست.این روش به بررسی همبستگی درونی تعداد زیادی از متغیرها می پردازد و در نهایت آنها را در قالب عاملهای عمومی محدودی دسته بندی و تبیین می کند. بنابراین ارزش تحلیل عاملی این است که طرح سازمانی مفیدی به دست می دهد که می توان آن را برای تفسیر انبوهی از رفتار با بیشترین صرفه جویی در سازه های تبیین کننده، به کار برد.

تحلیل عاملی ،برخلاف رگرسیون چندگانه ،تحلیل تشخیصی یا همبستگی کانونی (که در آنها تعداد زیادی متغیر مستقل و یک یا چند متغیر وابسته وجود دارد)روشی هم وابسته است که در آن کلیه متغیرها به صورت همزمان مد نظر قرار میگیرند.
نکته قابل توجه این است که بین روش های تحلیل چند متغیره تفاوت اساسی وجود دارد که در بعضی روش ها متغیرها و در بعضی روش ها واحدهای آزمایشی مورد بررسی قرار می گیرد . از جمله روش هایی که روابط بین متغیرها را مورد بررسی قرار می دهد می توان به روش های تحلیل عاملی ، تحلیل مؤلفه های اصلی ، تحلیل رگرسیون اشاره نمود . تحلیل خوشه ای ، تحلیل واریانس چند متغیره و تحلیل ممیزی جزو گروه روش هایی هستند که روابط بین واحد های آزمایشی را مورد بررسی قرار می دهند.

مفاهیم کلیدی روش تحلیل عاملی:

- **اشتراک:میزان واریانس مشترک بین یک متغیر با سایر متغیرهای به کار گرفته شده در تحلیل عاملی. این اندازه ای است از اینکه چه میزانی از واریانس داده ها از یک متغیر مشخص به وسیله تحلیل تبیین می شود .در تحلیل عوامل اصلی اشتراکات اولیه همه مساوی یک هستند . بعد از استخراج عامل استخراج اشتراکات برای هر متغیر فقط بر اساس عوامل استخراج شده محاسبه می شود . هر چه مقدار اشتراک استخراج شده برای یک متغیر بیشتر باشد ، مقدار بیشتری از واریانس به وسیله عوامل استخراج شده تبیین شده است . اشتراک از طریق بارهای عاملی محاسبه می شود.**
- **مقدار خاص:میزان واریانس تبیین شده بوسیله هر عامل . مقدار ویژه اندازی است که تعیین می کند چه مقدار واریانس در کل داده ها به وسیله یک عامل تبیین می شود. در تحلیل عاملی مقدار ویژه برابر1می باشد ولی ما می توانیم در بسته آماری این مقدار را زیاد کنیم . در تحلیل عاملی مولفه های اصلی آنها هستند، که مقدار ویژه آنان بیشتر از 1 باشد. ولی اگر این مقدار کمتر از 1 باشد،**

عاملهای مورد نظر از لحاظ آماری معنی دار نیستند و باید از تحلیل کنار گذاشته شوند .هرکدام از بارهای عاملی را به توان 2 برسانیم و با هم جمع کنیم (مجموع مجذورات بارهای عاملی).

- **عامل:عبارت است از ترکیب خطی متغیرهای اصلی که خلاصه شده اند.**
- **بار عاملی: عبارت است از همبستگی بین متغیرهای اصلی و عوامل.**
- **ماتریس عاملی:جدولی است که بارهای عاملی کلیه متغیرها را در هر عامل، نشان میدهد.**
- **چرخش عاملی:فرایندی است که برای تعدیل محور عامل به منظور دستیابی به عاملهای ساده و بامعنی .** فرآیندی برای تعدیل محور عامل به منظود دستیابی به عاملهای معنی دار وساده است. محورهای مختصات عاملها را به دور مبدا چرخش داده است تا اینکه موقعیت جدیدی را بدست آورد . دونوع چرخش داریم :
- **چرخش متعامد (عاملها مستقل از یکدیگر هستند).**
- **چرخش متمایل (عاملها بایکدیگر همبستگی دارند).**
- **نمره عاملی: تحلیل عاملی متغیرهای اصلی را در تعداد محدودی از عاملها خلاصه میکند.**
- **طرح سنگ ریزه : نموداری مفید از مقادیر ویژه کلیه عواملی است که در ابتدا مورد توجه قرار گرفته اند و می توان از آن برای تصمیم گیری در مورد تعداد عواملی که باید استخراج شوند استفاده کرد.**

در یک مدل تحلیل عاملی، متغیرهای مشاهده شده $y_1,y_2,\dots,y_p$بصورت ترکیبات خطی تعداد کمتری از متغیرهای تصادفی غیرقابل مشاهده ($f_1,f_2,\dots,f_m$ یا عامل ها) بیان می شوند. بنابراین تحلیل عاملی با تحلیل مؤلفه اصلی تفاوت دارد. در تحلیل مؤلفه اصلی، مؤلفه های اصلی بصورت ترکیبات خطی از yها تعریف می شوند. بنابراین بطور کلی می توان گفت تفاوت تحلیل عاملی و مؤلفه اصلی در اهداف آنهاست؛ یعنی هدف اصلی در تحلیل مؤلفه های اصلی کاهش بعد از یک فضای بزرگتر به یک زیرفضا می باشد که بیشتر تغییرات در مشاهدات اولیه را در خود حفظ کند، در حالیکه تحلیل عاملی با فرض وجود تعداد معینی از عامل های غیر قابل مشاهده، بدنبال یافتن ضرایب مربوط به ترکیب خطی از پاسخ ها می باشیم.

تحلیل عاملی می تواند به چهار پرسش عمده پاسخ دهد:

1. برای تبیین الگوی روابط بین متغیرها به چند عامل مختلف نیاز است؟
2. ماهیت این عوامل چیست؟
3. عاملهای نظری چگونه می توانند دادهای مشاهده شده راتبیین کنند؟
4. چه مقدار از واریانس هر متغیر مشاهده شده اساسا تصادفی یا یگانه است؟

بنابراین،هدف تحلیل عاملی کشف ساده ترین الگواز میان الگوهای مربوط به روابط میان متغیرهاست.این روش به دنبال درک این مطلب است که آیا متغیرهای مشاهده شده را می توان برپایه تعداد کمتری متغیر(عامل)به گونه وسیع واساسی تبیین کرد.از این رو ،هدف های عمده تحلیل عاملی را می توان به شرح زیر بیان کرد:

1. کاهش تعداد زیادی متغیربه تعداد کمتری عامل به منظور مدل سازی.زیرا اگر تعدادمتغیرها زیاد باشد،مدل سازی برای همه اندازه ها به گونه انفرادی امکان پذیر نیست.به همین دلیل ،تحلیل عاملی بامدل سازی های مربوط به تکنیک معادلات ساختاری یکپارچه شده وبه تولید متغیرهای مکنون که از طریق مدل سازی معادله ساختاری به وجود می آید،کمک میکند.
2. گزینش یک پاره تست از مجموعه زیادی سوال که دارای بیشترین همبستگی با مولفه های اصلی باشد.

3. تولید مجموعه ای از عوامل به عنوان متغیرهای ناهمبسته به گونه ای که بااصطلاح همخطی بودن چند گانه در روش رگرسیون چند متغیری نزدیک باشد.
4. رواسازی یک مقیاس یا شاخص از طریق تعیین بار مواد سازنده مقیاس برروی عامل های استخراج شده.
5. تهیه تست های چندگانه ای که تنها یک عامل را بسنجد و اجرای تست های کمتری را ممکن سازد.
6. تعیین خوشه هایی از آزمودنی ها.
7. تعیین گروه ها از طریق مشخص کردن مجموعه افرادی که در یک خوشه قرار دارند.

تصمیم گیری در تحلیل عاملی:

هدف اصلی تحلیل عاملی تلخیص تعداد زیادی از متغیرها در تعداد محدودی از عاملها می باشد، بطوریکه در این فرایند کمترین میزان گم شدن اطلاعات وجود داشته باشد. با توجه به هدف تحلیل عاملی محقق سوالاتی از خود می پرسد چه نوع متغیرهای باید در تحلیل به کار گرفته شود . در پاسخ به این سوال باید گفت که هر متغیری مرتبط با مسئله تحقیق را می توان در تحلیل به کار گرفت.بر این اساس مفروضاتی برای استفاده از تحلیل عاملی باید در نظر گرفت که به شرح زیر می باشد :

1- متغیر ها باید حداقل در سطح اندازه گیری تربیتی باشند.
2- متغیر ها باید توزیع بهنجار داشته باشند .
3- رابطهٔ بین متغیرها باید تاحدی خطی باشد.
4- حداقل داده ها شرکت داده شده در تحلیل نباید از 100 مورد کمتر باشد.

- معنی داری ماتریس همبستگی

یکی از روشهای انتخاب متغیرهای مناسب برای تحلیل عاملی استفاده از ماتریس همبستگی است که اساس روش تحلیل عاملی برای انتخاب متغیرها به عاملهای متفاوت استفاده از همبستگی بین متغیرها اما از نوع غیر علی استوار است. البته آمارهای دیگری وجود دارد که محقق از طریق انها نیز قادر به تعیین و تشخیص مناسب بودن داده ها برای تحلیل عاملی می باشد از جمله این روشها ، مقدارکفایت نمونه یا شاخص KMO[1]می باشد که مقدار آن همواره بین 0و1 می باشد و در صورتی که این مقدار کمتر از 50/. باشد داده ها برای تحلیل عاملی مناسب نخواهد بود . یعنی:

$$KMO_i = \frac{\sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} a_{ij}^2}$$

در فرمول r_{ij} ضریب همبستگی بین متغیرهای i و j(دو عنصر بردار x هستند) و a_{ij} ضریب همبستگی جزیی (شرطی یا ناقص) بین متغیرهای i و jاست. اگر همبستگی خطی و قوی بین متغیر ها وجود داشته باشد مقدار a_{ij} کم و در نتیجه مقدار KMO نزدیک به یک خواهد بود . کایزر مقدار KMO را در سال 1974 به صورت زیر تقسیم نموده است :

[1] - Kaiser –Meyer – Olkin

- اگر $KMO \geq 0.9$ باشد تجزیه به عامل ها بسیار مفید می باشد.
- اگر $0.8 \leq KMO < 0.9$ باشد تجزیه به عامل ها خوب می باشد.
- اگر $KMO < 0.5$ باشد تجزیه به عامل ها ب مفید نمی باشد.

از نکات دیگری که در این رابطه مدنظر قرار می گیرد اندازه کفایت نمونه گیری یا **MSA**[1] می باشد. به مانند روش قبل اندازه کفایت نمونه گیری متغیر **i**ام از رابطه زیر بدست می آید:

$$MSA_i = \frac{\sum_{j \neq i} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} a_{ij}^2}$$

اگر مقدار MSA_i زیاد باشد ، تجزیه به عامل ها مفید است و بر عکس غیر مفید و باید حذف شود.

و از سوی دیگر برای اطمینان از داده ها برای تحلیل عاملی مبنی بر اینکه ماتریس همبستگی که پایه تحلیل عامل قرار می گیرد در جامعه برابر صفر است یا خیر باید از آزمون بارتلت[2] استفاده کنیم . این آزمون که به عنوان آزمون کرویت بارتلت معروف است برای ارزشیابی وضعیت ماتریس همبستگی بین متغیرها استفاده می شود . خروجی این آزمون ، آماره خی دو ، درجه آزادی و سطح معنی داری را نشان می دهد که در صورت معنی داری فرض واحد بودن ماتریس همبستگی رد می شود . یعنی اینکه آیا ماتریس کورولاسیون یک ماتریس واحد است یا خیر (کلیه عناصر روی قطر برابر بک و عناصر خارج از قطر صفر هستند) ؟ برای انجام این آزمون داده باید دارای توزیع بهنجار یا نرمال چند متغیره باشند. اگر این مقدار کمتر از **05/.** باشد داده ها با جامعه مورد معنی دار است. فرضیات این آزمون به صورت زیر می باشد:

$$H_O : \rho = I$$
$$H_\alpha : \rho \neq I$$

در اینجا **I** ماتریس واحد است . اگر فرض صفر رد شود نتیجه گرفته می شود که تجزیه به عامل ها مفید است و اگر فرض صفر رد نشود یعنی تجزیه به عامل ها مفید نمی باشد و آن را رد می کنیم . همان طوری که اشاره شد آماره این آزمون بر مبنای توزیع خی دو با درجه آزادی $\frac{P(P-1)}{2}$ می باشد. لذا :

---

[1] - Measure of Sampling Adequacy

[2] - Bartlett

$$Q = -(N - \frac{2P+1}{6})\sum_{i=1}^{P} Ln\, l_i$$

l_i مقادیر ویژه ماتریس همبستگی می باشد . اگر Q بزرگتر از X^2_α باشد فرض صفر رد می شود و نتیجه گرفته می شود تجزیه به عامل ها مفید است و برعکس فرض صفر رد نمی شود و نتیجه گرفته می شود تجزیه به عامل ها مفید نمی باشد. روش دیگری که در این رابطه استفاده می شود استفاده از ضریب تعیین همبستگی چند گانه می باشد که اگر مقدار آن کوچک بود باید آن متغیر را از مجموع متغیر ها حذف کرد.

- **حجم نمونه**

در رابطه با حجم نمونه نیز باید تاکید کرد که تعداد حجم نمونه نباید کمتر از 50 مورد باشد و ترجیحا حجم نمونه را به بیش از 100 مورد افزایش داد. به عنوان قاعده کلی تعداد نمونه باید حدود چهار یا پنج برابر تعداد متغیرهای مورد استفاده باشد.

- **انتخاب نوع ماتریس همبستگی**

بعد از اطمینان داشتن به داده ها برای تحلیل عاملی، اولین تصمیم در بکارگیری تحلیل عاملی، محاسبه ماتریس همبستگی است. برای اینکار باید مشخص شود که آیا هدف،محاسبه همبستگی بین متغیرهاست یا بین پاسخگویان. اگر هدف تحقیق تلخیص متغیرها باشد در این صورت از همبستگی بین متغیرهامحاسبه شود که این روش یکی از تکنیک های عمومی و پرکاربرد در مطالعات می باشد که به تحلیل عاملی نوع R معرو ف است. اما تحلیل عاملی ممکن است برای ماتریس همبستگی بین پاسخگویان نیز بکار گرفته شود این نوع تحلیل را تحلیل نوع Q می نامند . این نوع تحلیل عاملی شاید بدلیل مشکل بودن کمتر مورد استفاده قرار گیرد و بجای آن از روشهای نظیر تحلیل خوشه ای یا گروهبندی سلسله مراتبی برای طبقه بندی پاسخگویان یا موارد استفاده می شود.

- **انتخاب مدل عاملی**

در تحلیل عاملی مدلهای مختلفی وجود دارد که از میان آنها دو روش تحلیل مولفه های اصلی و تحلیل عاملی مشترک از پر کاربرد ترین این روشهاست. انتخاب هریک از مدلها به هدف محقق بستگی دارد.مدل تحلیل مولفه های اصلی زمانی مورد استفاده قرار می گیرد که هدف محقق تلخیص متغیرها و دستیابی به تعداد محدودی عامل برای اهداف پیش بینی شده باشد و در مقابل تحلیل عاملی مشترک زمانی بکار می رود که هدف شناسایی عاملها یا ابعادی باشد که به سادگی قابل شناسایی نیستند.

- **روش استخراج عامل ها**

علاوه بر انتخاب مدل تحلیل، محقق باید مشخص کند که عاملها چگونه باید استخراج شوند. برای استخراج عاملها دو روش وجود دارد. عاملهای متعامد و عاملهای متمایل. در روش متعامد، عاملها به شیوه انتخاب می گردند که محورهای عاملی در حالت 90 درجه قرار می گیرند و این بدین معناست که هر عامل، مستقل از سایر عاملها می باشد. بنابراین،همبستگی بین عاملها، بطورقراردادی صفر تعیین می

گردد. مدل عاملی متمایل پیچیده تر از مدل عاملی متعامد می باشد. در واقع در این روش فرآیند تحلیلی کاملا رضایت بخش بدست نمی آید. در این روش عاملهای استخراج شده دارای همبستگی می باشند.انتخاب اینکه چرخش عاملها بر اساس متعامد و یا متمایل باشدباید بر اساس نیازهای محقق و مسئله تحقیق وجود دارد انجام گیرد.

- **انتخاب نهایی عامل ها**

زمانیکه در خصوص ماتریس همبستگی، مدل عاملی و روش استخراج، تصمیم مناسب اتخاذ گردید، زمینه برای استخراج عاملهای اولیه چرخش نیافته فراهم می گردد. با بررسی ماتریس چرخش نیافته محقق می تواند به جستجوی روشهای تلخیص داده ها و تعیین عاملهای استخراجی بپردازد، اما تعیین نهایی تعداد عاملها پس از دستیابی به ماتریس عاملی چرخش یافته امکان پذیر می باشد.

انواع تحلیل عاملی

تحلیل عاملی دو نوع می باشد:

1) تحلیل عاملی اکتشافی[1] :

در این تحلیل محقق درصدد کشف ساختار زیربنایی مجموعه نسبتاً بزرگی از متغیرها است بدون وجود هیچ تئوری اولیه ای. هر متغیری ممکن است با هر عاملی ارتباط داشته باشد.

2) تحلیل عامل تأییدی[2] :

پیش فرض اساسی محقق، آن است که هر عاملی با زیرمجموعه خاصی از متغیرها ارتباط دارد. حداقل شرط لازم این است که محقق در مورد تعداد عاملهای مدل قبل از انجام تحقیق، پیش فرض معینی داشته باشد.

مراحل انجام تحلیل عاملی

1. **انتخاب و جمع آوری داده ها و تشکیل ماتریسی از ضرایب همبستگی متغیرها،**
2. **استخراج عامل ها از ماتریس ضریب همبستگی،**
3. **استخراج عامل و چرخش عامل ها به منظور به حداکثر رساندن رابطه بین متغیرها و عاملها**
4. **محاسبه بار عاملی (نمره عامل ها) برای تعیین عاملهای مورد نظر**

نکته 1: معمولاً عاملهایی انتخاب می شوند که بار عاملی آنها بیشتر از 0/4 باشد.

نکته 2: برخی از صاحبنظران معتقدند در تحلیل عاملی، تعداد مشاهدات (Cases) حداقل باید 10 برابر تعداد متغیرها باشد. برخی حتی اعتقاد به حداقل 20 برابر تعداد متغیرها دارند.

تبیین مراحل تحلیل عاملی :

---

[1] - Exploratory Factor Analysis
[2] - Confirmatory Factor Analysis

اگر سری جدول زیر (جدول شماره 1) را براساس داده های مربوط به شش متغیر در روی ده ایستگاه مد نظر قرار گیرد:

نام متغیر نام ایستگاه	X1	X2	X3	X4	X5	X6
1	9/28	158	67	309	42	20
2	12/02	188	88	404	52	16
3	10/98	184	88	395	56	16
4	11/76	198	96	360	52	16
5	11/38	198	92	361	54	20
6	10/30	168	73	365	50	16
7	11/65	189	79	293	48	22
8	11/14	184	83	337	48	18
9	12/5	177	82	381	52	18
10	12/3	174	82	406	56	16

بردار داده ها:

در جدول شماره 1 تعداد متغیر ها P و تعداد واحدهای آزمایش N در نظر گرفته می شود. در جدول مذکور تعداد متغیرها p=6 و تعداد واحدهای آزمایش N=10 می باشد. هر ایستگاه را می توان بوسیله یک بردار سطری یا ستونی مشخص کرد. مثلا " ایستگاه اول و پنجم را می توان به صورت بردار سطری نوشت و عناصر این دو بردار را مشاهدات مربوط به اولین و پنجمین واحد آزمایش را نشان داد:

$$X_1' = \begin{bmatrix} 9.28 & 158 & 67 & 309 & 42 & 20 \end{bmatrix}$$
$$X_5' = \begin{bmatrix} 11.38 & 198 & 92 & 361 & 54 & 20 \end{bmatrix}$$

به طور کلی هر واحد آزمایشی را به صورت یک بردار سطری با مرتبه 1×P می توان نوشت . عناصر این بردار X_r' متغیرهای تصادفی هستند. یعنی :

$$X_r' = \begin{bmatrix} x_{r1} & x_{r3} & \dots & x_{rj} & \dots & x_{rp} \end{bmatrix}$$

ترانهاد[1] یک بردار سطری به صورت بردار ستونی است و در تحلیل های چند متغیره اغلب بردار ستونی بکار برده می شود. یعنی :

[1] - Transpose

$$X_r = \begin{bmatrix} x_{r1} \\ x_{r2} \\ \vdots \\ x_{rj} \\ \vdots \\ x_{rp} \end{bmatrix}$$

ماتریس داده ها :

در جدول شماره 1 سطرها واحدهای آزمایشی (10 ایستگاه)و ستون ها متغیرها (6 متغیر) هستند و ماتریس داده ها از مرتبه 6 × 10 می باشد. لذا ماتریس داده را می توان به صورت زیر نوشت :

$$X = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ r \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & \dots & X_j & \dots & X_P \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \vdots & & \vdots & & & \\ x_{r1} & x_{r2} & \cdots & x_{rj} & \cdots & x_{rp} \\ \vdots & & & & & \\ x_{N1} & x_{N2} & \cdots & x_{Nj} & \cdots & x_{NP} \end{bmatrix} \end{array}$$

X تا X_p به ترتیب معرف متغیرهای انداره گیری شده در تمام واحد های آزمایشی هستند و x_{rj} معرف اندازه متغیر jام در واحد آزمایشی rام است.

پارامترهای جامعه شامل بردار میانگین[1] ، ماتریس واریانس – کوواریانس[2] و ماتریس همبستگی[3] می باشد. در توزیع بهنجاری یا نرمال بودن یک متغیر میانگین جامعه (μ) نشان داده می شود و به این ترتیب تعریف می شود $\mu = E(x)$ یعنی میانگین تمام افراد جامعه در مورد یک صف . در توزیع نرمال چند متغیره میانگین بردار متغیرهای تصادفی X با μ نشان داده می شود و به این ترتیب تعریف می شود:

1 - Mean Vector
2 - Variance-Covariance Matrix
3 - Correlation Matrix

$$\mu = E(x) = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

در توزیع نرمال یک متغیره واریانس جامعه با نماد σ^2 نشان داده می شودو به صورت $\sigma^2 = E\left[(x-\mu)^2\right]$ تعریف می شود. درحالیکه در توزیع نرمال چند متغیره ، ماتریس واریانس –کواریانس با نماد $\sum$ مشخص شده و به این ترتیب به صورت زیر تعریف می شود:

$$\sum = Cov(X) = E\left[(X-\mu)(X-\mu)'\right] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1P} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2P} \\ \vdots & \vdots & & \\ \sigma_{P1} & \sigma_{P2} & \cdots & \sigma_{PP} \end{bmatrix}$$

در این ماتریس واریانس – کوواریانس عناصر قطری σ_{11} تا σ_{PP} واریانس و عناصر غیر قطری کوواریانس را مشخص می کند. هر عنصر از این ماتریس را می توان به تنهایی به صورت زیرنوشت:

عناصر قطری $\sigma_{ii} = Var(x_1) = E(x_i - \mu_i)^2 \qquad i = 1,2...,p$

عناصر غیر قطری $\sigma_{jj} = Cva(x_i, x_j) = E\left[(x_i - \mu_i)(x_j - \mu_j\right] \qquad i \neq j = 1,2...,p$

ضریب همبستگی بین دو متغیر X_iو X_j را با استفاده از رابطه زیر می توان بدست آورد:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

مقادیر ρ_{ij} برای هر $i \neq j$ در فاصلهٔ مثبت و منفی یک می باشد. اگر دو متغیر توزیع مستقل با هم داشته باشددر این صورت $\rho_{ij} = 0$ و σ_{ij} صفر می باشد. ماتریس همبستگی برای یک بردار تصادفی را به صورت می توان نشان داد:

$$P = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1P} \\ \rho_{21} & 1 & \vdots & \rho_{2P} \\ \vdots & & 1 & \\ \rho_{P1} & \rho_{P2} & \cdots & 1 \end{bmatrix}$$

ماتریس واریانس –کوواریانس و ماتریس همبستگی ماتریس های متقارن هستند. یعنی عناصر غیر قطری آنها دو به دو به طور متناظر مساوی است مثلا" $\sigma_{13} = \sigma_{31}$ یا $\sigma_{12} = \sigma_{21}$ است.

داده های استاندارد شده :

در بسیاری موارد متغیرهای انداره گیری شده مقیاس های متفاوتی دارند . برای ازبین بردن تفاوت مقیاس های داده ها لازم است داده ها را استاندارد نمود. فرم استاندارد شده متغیر X_{ij} را با Z_{ij} نشان داده می شود و به صورت زیر می توان تعریف نمود:

$$Z_{ij} = \frac{x_{rj} - \mu_j}{\sqrt{\sigma_{jj}}} \qquad \begin{matrix} r = 1,2,\ldots,N \\ j = 1,2,\ldots p \end{matrix}$$

ماتریس داده های استاندارد شده به صورت زیر می باشد:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & \cdots & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & \cdots & \cdots & z_{2p} \\ \vdots & \vdots & & & & \\ z_{r1} & z_{r2} & \cdots & z_{rj} & \cdots & z_{rp} \\ \vdots & \vdots & & & & \\ z_{N1} & z_{N2} & \cdots & \cdots & \cdots & z_{NP} \end{bmatrix}$$

مشکلی را که استاندارد کردن به وجود می آورد اینست که تفاوت واریانس متغیرها را نیز از بین می برد.

توصف هندسی بردارها :

یک بردار دو بعدی p=2 را بصورت یک نقطه در محور مختصات دو بعدی می توان نشان داد . بیان هندسی بردار $X_1 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ **به این ترتیب است که همیشه عنصر اول روی محور افقی و عنصر دوم روی محور عمودی جدا می شود . بنابراین 5 واحد روی محور افقی و 2 واحد روی محور عمودی جد ا می شود. حال اگر بردار دیگری به صورت** $X_2 = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$ **در نظر گرفته شود چون حاصل ضرب دو بردار برابر با** $X_1' X_2 = \begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -5 \end{bmatrix} = 5(2) + 2(-5) = 0$ **بنابر این دو بردار متعامد و بر هم عمودند.**

تبدیل های خطی :

بیشتر روش های آماری چند متغیره بر اساس تبدیل های خطی بنا شده اند. یک تبدیل خطی به طور خلاصه عبارت است از حاصل جمع موزون چند متغیر به این ترتیب که هر متغیر نخست در یک عددثابت(وزن) ضرب می شود و سپس حاصلضرب ها با هم جمع می شوند. در حالت کلی یک تبدیل خطی به صورت زیر نوشته می شود:

$$Y_r = \alpha_1 x_{1r} + \alpha_2 x_{2r} + \ldots + \alpha_p x_{pr} \qquad r = 1,2,\ldots,N$$

به عبارت دیگر بردار سطری وزن ها $\alpha' = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_P \end{bmatrix}$ **که هر عنصر آن وزن متغیر متناظر است در بردار ستونی متغیر های تصادفی** $X = \begin{bmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{pr} \end{bmatrix}$ **ضرب می شود.**

مثلا":

نام متغیر نام ایستگاه	X1	X2	X3	X4	X5	X6	رتبه Y
وزن	α_1=1	α_2=2	α_3=2	α_4=3	α_5=1	α_6=2	نمره تبدیل شده
1	9/28	158	67	309	42	20	1468.28
2	12/02	188	88	404	52	16	1860.02
3	10/98	184	88	395	56	16	1827.98
4	11/76	198	96	360	52	16	1763.76
5	11/38	198	92	361	54	20	1768.38
6	10/30	168	73	365	50	16	1669.30
7	11/65	189	79	293	48	22	1518.65
8	11/14	184	83	337	48	18	1640.14
9	12/5	177	82	381	52	18	1761.50
10	12/3	174	82	406	56	16	1830.30
میانگین	11/33	363/6	166	1083/3	51	35/6	1710/831
واریانس	0/944	651/378	299/556	13350/9	18	19/378	17935/708

بر این اساس می توان برای هر ایستگاه رتبه تعیین نمود . لذا رتبه اول ایستگاه شماره 2 و رتبه آخر ایستگاه شماره 1 می باشد. میانگین و واریانس تبدیل های خطی جدول بالا به صورت زیر می باشد:

$$\bar{Y} = 1(11.33) + 2(363.6) + 2(166) + 3(1083.3) + 1(51) + 2(35.6) = 4442.63$$

با توجه به واریانس محاسبه شده به فرض مستقل بودن متغیرها ، واریانس تبدیل خطی برابر است با:

$$\sigma_Y^r = 1^2(0.944) + 2^2(651.378) + 2^2(299.556) + 3^2(13350.9) + 1^2(18) + 2^2(19.378) = 124057.9$$

مراحل اجرای تحلیل عاملی:

1. **انتخاب و جمع آوری داده ها و تشکیل ماتریسی از ضرایب همبستگی متغیرها،**
2. **استخراج عامل ها از ماتریس ضریب همبستگی،**
3. **استخراج عامل و چرخش عامل ها به منظور به حداکثر رساندن رابطه بین متغیرها و عاملها**
4. **محاسبه بار عاملی (نمره عامل ها) برای تعیین عاملهای مورد نظر**

1. انتخاب و جمع آوری داده ها و تشکیل ماتریسی از ضرایب همبستگی متغیرها،

ماتریس همبستگی ماتریسی مربعی از ضرایب همبستگی متغیرها با یکدیگر است.
برای اینکار باید مشخص شود که آیا هدف، محاسبه همبستگی بین متغیرهاست یا بین پاسخگویان.به عنوان مثال ممکن است داده هایی از 70 نفر را در مورد 10 سوال مربوط به ازدواج گردآوری شده باشد این امکان وجود دارد که همبستگی بین هریک از 10 متغیر و یا بین هرکدام از پاسخگویان محاسبه شود.

اگر هدف مطالعه، تلخیص متغیرها باشد ،در اینصورت باید همبستگی بین متغیرها محاسبه شود این روش به تحلیل عاملی نوع R معروف است.اگر تخلیل عاملی برای ماتریس همبستگی بین پاسخگویان به کار برده شود،در اینصورت روش فوق را روش تحلیل عاملی ،نوع Q میگویند.
تحلیل عاملی نوع Q به دلیل مشکل بودن کمتر،مورد توجه قرار گرفته و بیشتر از روش تحلیل خوشه ای برای طبقه بندی پاسخگویان استفاده میشود.
به عنوان مثال اگر پاسخگویان بوسیله شماره مشخص شوند نتایج تحلیل عاملی از نوع Q ممکن است نشان دهد که افراد شماره های 10،5،2و14 شباهت بیشتری به هم دارند و چون این افراد بار بیشتری بر یک عامل دارند(یعنی همبستگی بین آنها زیاد است) بنابراین افراد فوق میتوانند در یک گروه قرار گیرند.
یکی از روشهای انتخاب متغیرهای مناسب برای تحلیل عاملی استفاده از ماتریس همبستگی است.از آنجا که اساس روش تحلیل عاملی بر همبستگی بین متغیرها اما از نوع غیرعلی استوار است بنابراین در استفاده از این روش باید ماتریس همبستگی بین متغیرها نیز محاسبه گردد .توصیه میشود متغیرهایی که با هیچ متغیری همبستگی لازم را نداشته باشند از تحلیل حذف گردند .
البته آماره های دیگری نیز وجود دارند که محقق از طریق آنها نیز قادر به تعیین و تشخیص مناسب بودن داده ها برای تحلیل آماری میباشد. از جمله این روشها روشKMO میباشد.

متغیرهایی که برای تحلیل عاملی انتخاب می شوند باید در سطح سنجش آنها ترتیبی بوده و بین آنها همبستگی وجود داشته باشد و با تشکیل ماتریس همبستگی ، متغیرهایی که باهیچ یک از متغیرهای دیگری همبستگی ندارند. ، حذف شوند. در واقع همبستگی بین متغیرها باید محصول عامل دیگری ، یعنی عامل مشترک سومی باشد. در این رابطه برای اینکه سنجش شود که ماتریس داده ها برای تحلیل عاملی مناسب است یا نه باید به نکات زیر توجه نمود:

- **تعیین ترکیب ماتریس دادها .**
- **حجم گروه نمونه باید بزرگتر باشد.**

- **اندازه همخوانی .**
- **استقلال اندازه ها .**
- **معنی دار بودن ماتریس (همبستگی بیشتر).**
- **مناسب بودن شاخص کفایت نمونه .**
- **آزمون کرویت بارتلت فرض صفر را رد کند.**
- **کم بودن ضریب ضد تصویر[1](منفی بودن ضریب همبستکی تفکیکی).**

روش KMO:

مقدار آن همواره بین 0و 1 در نوسان است در صورتی که KMO کمتر از 0.5 باشد داده ها برای تحلیل عاملی مناسب نخواهد بود و اگر مقدار آن بین 0 .5 تا 0.69 باشد داده ها متوسط بوده و اگر مقدار این شاخص،بزرگتر از 0.7 باشد همبستگی های موجود در بین داده ها برای تحلیل عاملی مناسب خواهند بود.

روش بارتلت:

یکی دیگر از روشهای تشخیص مناسب بودن داده ها میباشد آزمون بارتلت ،این فرضیه را که ماتریس همبستگی مشاهده شده متعلق به جامعه ای با متغیرهای نابسته است ، می آزماید. برای اینکه یک مدل عاملی ،مفید و دارای معنا باشد لازم است متغیرها همبسته باشند.پس فرضیه آزمون بارتلت به اینصورت است:

H0 : داده ها ناهمبسته اند.

H1 :داده ها همبسته اند.

پس مطلوب آن است که فرض صفر رد شود. آگر فرض صفر رد نشود مطلوبیت تحلیل عاملی زیر سوال میرود.و باید درباره انجام آن تجدید نظر کرد.

به همین دلیل است که قبل از قبل از تحلیل عاملی بایستی به تشکیل ماتریس همبستگی بین متغیرها اقدام کرد.

2. استخراج عامل ها از ماتریس ضریب همبستگی:

در فرآیند استخراج عامل ها توجه به دو نکته زیر ضروری است :

- **روش استخراج عامل ها**
- **معیار تعیین عامل ها**

- **روش استخراج عامل ها:**

سه نوع واریانس با توجه به هدفی که ما در تحلیل عاملی داریم قابل تعریف است:

1- واریانس مشترک: آن بخش از واریانس است که با سایر متغیرهای لحاظ شده در تحلیل سهیم میباشد.

2- واریانس خاص: واریانسی است که تنها به متغیر Xi مربوط است .

[1] - Anti- Image

3- واریانس خطا: ناشی از بی اعتباری در داده های جمع آوری شده و یا شانس و تصادف در اندازه گیری پدیده هاست.

روش های مختلفی در استخراج عامل ها وجود دارد که عبارتند از :

1. **روش مؤلفه های مبناء یا اصلی .**
2. **روش تحلیل عاملی اشتراک .**
3. **روش کمترین مربعات ناموزون .**
4. **روش کمترین مربعات تعمیم یافته.**
5. **روش حداکثر درست نمایی.**
6. **روش عامل یابی الفا**
7. **روش عامل یابی تصویر**

متداول ترین روش ها روش مولفه های اصلی و تحلیل عاملی اشتراک می باشد که در تحلیل مؤلفه های اصلی عامل ها همه واریانس هر متغیر ، ازجمله واریانس مشترک با سایر متغیرهای مجموعه (کوواریانس) و نیز واریانس خاص متغیر را توجیه می کنند. در تحلیل مؤلفه های اصلی به تعداد متغیرها ، مؤلفه وجود دارد ولی در تحلیل عامل مشترک عامل ها فقط واریانس مشترک با متغیرهای دیگر مجموعه را توجیه می کنند . لذا تعداد عامل های مشترک ممکن است کمتر از تعداد متغیرها باشد.
واریانس خاص کم است.در تحلیل مولفه های اصلی ،مقادیر قطرهای ماتریس همبستگی 1 است. برعکس در تحلیل عاملی مشترک،مقادیر مشترکات در قطر ماتریس قرار میگیرد و عاملها تنها بر اساس واریانس مشترک استنتاج میگردند.

- **معیار تعیین عامل ها:**

برای تعیین عامل ها باید به معیار زیر توجه نمود:

- **معیار مقدار ویژه (ضریب همبستگی را با یک عامل به توان رسانده می شودو باهم جمع می شود . مقدار بیشتر از یک معنی دار بودن مقدار ویژه را مشخص می کند).**
- **معیار طرح سنگ ریزه (عواملی که مقادیر ویژه آنها کمتر یک باشد کنار گذاشته می شود).**
- **معیار درصد واریانس تجمعی تبیین شده (مقادیر کمتر از 50 درصد کنارگذاشته می شود).**

3.استخراج عامل و چرخش عامل ها به منظور به حداکثر رساندن رابطه بین متغیرها و عاملها :

با چرخش عامل توزیع مجدد واریانس بین عامل ها می شود. برای استخراج عامل ها دو روش وجود دارد:

1- عامل های متعامد

2-عامل های متمایل

در روش متعامد ،عاملها به شیوه ای استخراج میشوند که محورهای عاملی در حالت 90 درجه قرار گیرند و این بدین معناست که هر عامل ، مستقل از سایر عامل ها میباشد بنابراین همبستگی بین عاملها بطور قراردادی صفر تعیین میگردد.
مدل عاملی متمایل،پیچیده تر است در این روش همبستگی بین عاملها صفر نیست و عاملها دارای همبستگی میباشند.
مدل عاملی متعامد از نظر ریاضی، ساده میباشد اما مدل متمایل قابل تعدیل بوده و بیشتر واقع گراست.
اگر هدف تحقیق ،تلخیص تعداد متغیرهای اصلی بدون توجه به اینکه نتایج عاملهای استخراج شده تا چه حد معنی دار خواهد بود انجام گیرد در اینصورت روش متعامد روش مناسبی خواهد بود. روش های عمده متعامد برای تحقق عبارتنداز :

1- وریماکس[1] (چرخش متعامد بر روی ضرایب عامل).

[1] - Varimax

2- کوارتیماکس [1] (چرخش متعامد با عامل های کم).

3- اکویماکس (چرخش متعامد با کمینه کردن).

4- اوبلیمین مستقیم .

5- **پروماکس .**

روش کوارتیماکس در ایجاد یک ساختار عاملی ساده تر موفق نمیباشد مشکل این روش در آن است که سعی میکند در فرایند چرخش ، یک عامل عمومی به دست آورد در حالیکه هدف اصلی چرخش ماتریس عاملی ،دست یابی به یک عامل عمومی وسیع نمیباشد. در مقابل روش وریماکس به دنبال مختصر کردن و ساده سازی ستونهای ماتریس عاملی است .هرچند که محاسبات جبری برای روش کوارتیماکس ساده تر از روش وریماکس میباشد اما وریماکس جداسازی عاملها را به طور شفاف تر و واضح تر نشان میدهد. به هر حال روش وریماکس استفاده وسیعتری در بین محققان دارد.

4. محاسبه بار عاملی (نمره عامل ها) برای تعیین عاملهای مورد نظر:

هدف مرحله استخراج عامل ها، به دست آوردن سازه های زیر بنایی است که تغییرات متغیر های مورد مشاهده را موجب شده است. در این مرحله باید ترکیب هایی از متغیر ها را که همبستگی های آن‌ها بالاترین میزان از واریانس کل مشاهده شده را نشان می‌دهد انتخاب می‌کند. در این مرحله باید عامل 1 را مشخص نمود. عامل 2، مجموعه متغیر هایی است که بالاترین سهم را در تبیین واریانس باقیمانده دارد. این شیوه برای عامل سوم، چهارم و عامل های بعدی ادامه پیدا می‌کند تا تعداد عامل های استخراج شده برابر با تعداد متغیر ها گردد. همبستگی هر متغیر با هر عامل بار عاملی [2] نامیده می‌شود و مقدار آن بین1- و 1+ تغییر می‌کند. واریانس تبیین شده توسط هر عامل برابر است با مجذور بار های عاملی آن. این واریانس مقدار ویژه [3] نامیده می‌شود. اولین مقدار ویژه همواره بیشترین بوده و از 1 بزرگتر می‌باشد. مقدار ویژه برای عامل های بعدی کوچکتر می‌باشد. تمام عامل های استخراج شده مورد علاقه محقق نیست. هدف تحلیل عاملی تبیین پدیده های مورد نظر با تعداد کمتری از متغیر های اولیه است. در وهله اول هدف تعیین تعداد عامل هایی است که در تحلیل نگه داشته می‌شود. علی الاصول عامل هایی باید نگه داشته شود که اعتبار صوری یا نظری داشته باشد. منتها قبل از فرایند چرخش نمی‌توان به معنی هر عامل به خوبی پی برد، بنابراین معمولا از ملاک های ریاضی مانند ملاک کایزر یا آزمون اسکری کتل برای نگه داشتن عامل ها استفاده می‌شود. بر اساس ملاک کایزر فقط عامل هایی نگه داشته می‌شوند که مجموع مجذور بارهای عاملی آن‌ها (مقدار ویژه) یک یا بیشتر باشد. این ملاک برای تحلیل عاملی آلفا مناسب است و برای سایر روش های تحلیل عاملی کران پایینی فراهم می آورد. در روش اسکری کتل نمودار مقدار ویژه برای هر عامل ترسیم می‌شود. در نقطه ای که شکل منحنی برای مقادیر ویژه به صورت افقی درآید، آن نقطه اسکری نامیده شده و عامل هایی که سمت چپ آن قرار دارد عامل های واقعی و آن هایی که در سمت راست آن قرار دارند عامل های خطا قلمداد می‌شود. در تفسیر نتایج آزمون اسکری ممکن است میان نظرات پژوهشگران درباره تعداد عامل های واقعی اختلاف نظر پدید آید. همچنین امکان دارد که بیش از یک اسکری موجود باشد. لذا لازم است علاوه بر آزمون اسکری آزمون های دیگری از جمله آزمون کایزر صورت گیرد.

پس از انتخاب عامل ها چرخش آن‌ها ضرورت دارد. هدف از چرخش عامل ها رسیدن به یک ساختار عاملی ساده است. در تحلیل عاملی، ساختار های عاملی متعددی برای یک ماتریس همبستگی وجود دارد. اولین عامل غالبا یک عامل کلی است که تمام یا اکثر متغیر ها بار عاملی بالایی روی این عامل دارد. عامل هایی بعدی معمولا دو قطبی است و بارهای عاملی مثبت و منفی داشته و قابل تفسیر نمی‌باشد با چرخش ساختار عاملی روشنتر می‌شود.

---

[1] - Quartimax

[2] - Factor Loading

[3] - Eigen Value

مشهورترین ملاک برای خوبی یک ساختار عاملی، ملاک مشهور ساختار ساده ثرستون است. طبق این ملاک هر متغیر باید حداقل یک بار عاملی غیر صفر داشته باشد. هر عامل باید فقط با چند متغیر همبستگی بالا داشته باشد. (منظور از همبستگی همان بار عاملی متغیر روی عامل است) و بار عاملی بقیه متغیر ها روی این عامل باید اساسا صفر باشد. هر متغیر باید روی یک عامل بار عاملی بالا داشته باشد. اغلب شیوه های چرخش با توجه به این ملاک ها طرح ریزی شده است.

کاربرد تحلیل عاملی با برنامه SPSS:

ابتدا متغیرهای مورد نظر را در ستونهای مختلف SPSS وارد کرده و عناوین متغیرها را تعریف میکنیم سپس دستور

Analyse → Data Reduction → Factor

را اجرا میکنیم در جدول ظاهر شده ابتدا متغیرها را انتخاب کرده و وارد میکنیم. همانطوریکه اشاره شد در تحلیل عامل دو کار یگی گاهش داده ها و دیگری شناسایی ساختاری را می توان انجام داد.

دکمه Descriptives : در این قسمت روش kmo & Bartletts را برای تشخیص مناسب بودن داده ها انتخاب میکنیم.

دکمه Extraction: با زدن این کلید پنجره ای باز میشود که در قسمت method میتوان روشهای مختلف تحلیل عاملی مانند تحلیل مولفه های اصلی یا تحلیل عاملی مشترک را انتخاب کرد.در پایین این جدول معیار انتخاب تعداد عاملها نوشته شده است. در قسمت Extract بطور پیش گزیده ،معیار مقدار ویژه و همچنین اندازه 1 برای آن نوشته شده است.یعنی تعدادی عامل استخراج میشود که اندازه ویژه آنها بزرگتر از یک باشد.

همچنین در قسمت بعدی میتوان تعداد عاملها را از قبل تعیین کرد.

دکمه Rotation : در این قسمت میتوان نوع چرخش عاملی را مشخص کرد که از قبل واریماکس انتخاب شده است.

دکمه Scores : چنانچه بخواهید مقادیر عاملی را محاسبه و در تحلیل های بعدی نظیر رگرسیون استفاده کنید این دکمه را فشار دهید.نتایج این قسمت در output نمی آید بلکه در قسمت Data Editor ظاهر میشود.

دکمه options: در صورتی که بخواهید بارهای عاملی استخراج کنید که مقدار آن بزرگتر از عدد خاصی مانند0.5 باشد این دکمه را فشار دهید.

در پنجره به طور پیش گزیده عدد 0.1 نوشته شده است.با انتخاب 0.5 به جای آن ،تنها بارهای عاملی بزرگتر از0.5 در نتایج ظاهر میشوند.و این کار تحلیل کردن را راحت تر میکند.

مثال :

فرض کنید داده های بارش ماهانه ایستگاه های فرضی را به صورت جدول زیر داشته باشیم :

داده ها ی دما

ردیف	stations	JAN	FEB	MAR	APR	MAY	JUNE	JULY
1	A	3.00	5.00	12.00	17.00	22.20	25.00	30.00

2	B	3.50	8.00	12.20	20.50	28.40	25.90	29.40
3	C	3.00	5.00	11.80	13.40	19.70	24.50	26.70
4	D	5.80	6.20	13.50	19.70	21.60	27.40	28.50
5	E	4.30	4.80	12.00	19.20	21.60	26.90	27.70
6	F	5.70	10.00	8.80	14.80	21.50	27.80	28.00
7	G	4.20	5.20	12.20	14.30	24.80	30.00	32.10
8	H	4.60	10.10	13.30	17.10	23.20	25.70	27.40
9	I	7.10	10.10	11.90	18.70	21.10	28.70	28.40
10	J	-1.40	6.80	13.60	17.20	24.00	27.40	28.70
11	K	5.60	6.80	12.00	16.50	23.70	26.80	28.10
12	L	7.80	8.60	11.40	16.50	23.50	28.20	27.70
13	M	3.40	6.20	11.10	14.90	22.70	26.60	29.80
14	N	4.30	4.50	13.00	17.20	20.70	28.60	28.00
15	O	3.70	5.30	15.30	17.00	22.00	26.60	28.90
16	P	4.00	7.90	10.40	19.00	24.10	27.80	27.50
17	Q	3.30	6.70	12.90	19.30	26.90	30.20	29.50
18	R	3.80	2.70	11.60	19.40	22.60	29.90	29.50
19	S	1.90	9.80	13.60	20.10	25.60	31.00	30.80
20	T	3.00	4.00	14.30	18.40	24.30	28.60	29.20
21	U	4.30	7.40	11.70	16.40	24.30	28.10	30.50
22	V	7.10	5.00	8.90	16.90	25.00	28.20	31.50
23	W	1.00	7.40	16.20	16.80	23.20	29.10	30.50
24	X	4.80	7.60	12.00	20.20	24.60	28.60	29.00

25	Y	6.20	8.50	10.10	21.50	21.30	29.40	30.90
26	Z	3.50	5.70	11.00	20.40	25.40	28.60	28.30
27	AB	4.80	7.20	12.20	19.10	23.10	26.00	27.40
28	AC	4.50	3.40	10.40	19.10	24.20	28.80	29.30
29	AD	3.60	6.20	8.80	14.20	23.40	27.80	29.20
30	AF	2.40	1.80	12.40	17.50	21.60	25.70	27.30
31	AE	4.90	7.80	10.00	17.30	21.10	27.00	28.50
32	AH	3.40	4.60	7.40	16.80	22.40	26.80	26.60
33	AK	4.40	7.20	11.80	17.00	21.90	25.10	26.60
34	BW	4.20	6.00	10.40	17.20	23.70	27.80	29.10
35	BR	1.40	1.70	9.90	15.70	20.50	27.30	27.70
36	BT	3.00	5.10	10.00	16.70	24.10	26.90	27.30

داده های بارش و دما

ردیف	AUG	SEP	OCT	NOV	DEC	ANNUAL	JAN.P	FEB.P
1	27.50	23.10	14.80	11.80	7.10	16.54	74.13	15.00
2	24.80	22.80	15.30	11.00	5.10	17.24	15.00	24.26
3	24.90	20.00	15.70	8.50	4.70	14.83	166.67	68.31
4	25.90	23.10	16.90	7.00	6.80	16.87	18.98	20.31
5	29.40	25.20	18.70	8.90	3.20	16.83	10.99	39.12
6	27.00	22.90	16.00	9.40	4.80	16.39	73.16	73.16
7	29.20	28.60	20.20	12.90	8.90	18.55	32.51	142.80
8	24.50	20.10	16.60	8.20	5.60	16.37	3.00	8.50

9	26.50	23.40	19.40	10.60	5.20	17.59	74.40	19.30
10	27.90	21.70	14.20	9.00	1.60	15.89	65.00	7.20
11	25.30	21.90	19.60	11.90	5.30	16.96	40.40	24.60
12	25.80	22.70	16.70	8.10	5.80	16.90	25.10	16.80
13	27.20	24.60	17.30	12.60	5.10	16.79	14.10	24.10
14	26.90	23.00	17.60	11.20	6.00	16.75	25.30	.80
15	24.80	21.40	18.50	7.60	7.40	16.54	15.20	6.10
16	29.20	22.10	17.20	11.60	5.90	17.23	8.20	6.40
17	28.70	22.80	20.00	14.00	9.80	18.68	1.20	4.10
18	25.70	24.40	19.20	14.40	4.80	17.33	.00	10.00
19	28.60	24.80	19.90	12.40	6.30	18.73	18.30	56.00
20	28.00	25.10	15.70	12.50	5.40	17.38	9.00	10.90
21	28.20	24.70	18.20	10.80	6.40	17.58	19.70	26.00
22	28.00	26.30	18.70	8.10	5.10	17.40	25.40	15.10
23	28.00	22.40	18.70	10.90	7.50	17.64	23.50	17.90
24	27.10	23.60	18.50	9.70	8.70	17.87	17.20	45.70
25	26.00	23.50	20.40	11.20	6.20	17.93	8.10	9.50
26	26.80	22.50	17.20	11.60	6.40	17.28	16.50	20.60
27	25.50	22.10	18.10	10.80	6.40	16.89	18.30	18.80
28	28.50	23.10	18.90	10.10	3.30	16.97	14.60	.00
29	28.20	23.50	15.10	10.80	4.20	16.25	8.60	10.50
30	27.70	20.70	13.20	11.40	2.90	15.38	14.80	49.10
31	23.20	21.30	15.50	10.70	5.70	16.08	35.60	1.10

32	24.20	20.70	17.90	10.00	3.20	15.33	.70	14.20
33	26.30	23.90	14.60	10.10	6.50	16.28	68.70	1.50
34	25.60	22.60	16.10	10.90	6.80	16.70	23.70	11.90
35	24.70	22.60	16.80	10.30	6.60	15.43	117.50	8.70
36	27.00	22.00	15.60	10.70	4.60	16.08	30.10	2.50

داده های بارش

ردیف	MAR.P	APR.P	MAY.P	JUNE.P	JULY.P	AUG.P	SEP.P	OCT.P
1	24.26	18.98	.00	.00	.00	.00	.00	.00
2	.00	10.99	.00	.00	.00	.00	.00	.00
3	188.94	20.49	6.38	.00	.00	.00	.00	.00
4	68.31	188.94	.00	.00	18.11	.00	.00	.00
5	25.91	24.33	1.99	.00	.00	.00	.00	1.24
6	28.03	38.54	2.39	.00	2.39	.00	.00	.00
7	86.39	55.21	11.09	.00	8.05	.00	.00	5.61
8	53.70	3.00	.00	.00	.60	.00	39.00	21.00
9	21.10	4.10	1.00	.30	.00	.00	.00	.00
10	14.50	4.20	53.80	9.10	.30	1.00	.20	.30
11	35.00	32.60	18.00	.10	.00	.00	.00	.40
12	6.60	16.70	17.40	.00	.00	.00	.00	1.40
13	37.20	5.50	4.30	.00	.00	.00	.00	20.40
14	24.30	9.70	11.10	.00	.00	.00	.00	.00

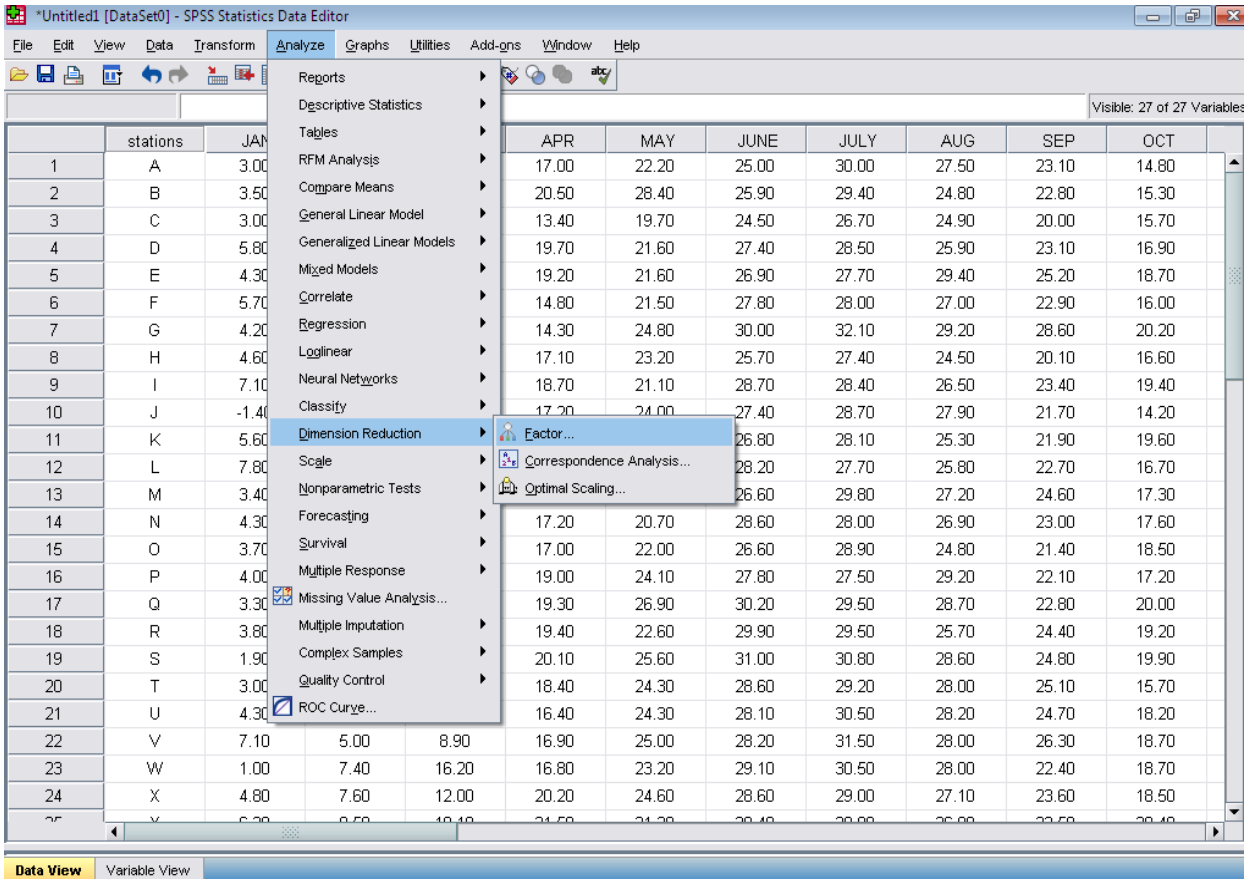
15	49.90	16.20	32.40	3.00	.00	.00	.00	.00
16	21.70	.10	1.50	.00	.10	.00	.00	.00
17	48.50	31.60	31.20	.20	.00	5.80	.00	.00
18	10.00	1.00	.40	.00	5.00	1.10	.00	19.20
19	20.20	9.20	.00	.00	.00	.00	.00	.00
20	24.10	9.70	.70	.00	.00	.00	.00	.00
21	2.40	9.70	.30	.00	.00	.00	.00	.00
22	61.20	2.80	.00	.30	.00	.00	.00	10.30
23	12.10	13.10	2.80	.00	.20	.60	.00	.00
24	45.80	1.50	3.00	.00	.00	.00	.00	1.00
25	38.10	23.50	5.50	.00	.00	.00	.00	26.10
26	26.30	16.80	53.50	1.80	.00	.00	.00	28.00
27	53.90	4.20	11.10	.80	.60	.00	.00	.00
28	31.20	49.00	3.80	6.90	.00	.00	.00	.60
29	31.70	.20	.10	.00	4.00	.00	.00	2.50
30	46.40	15.50	.00	.00	.00	.00	.00	.00
31	2.40	.20	.00	.00	.00	.00	.00	3.00
32	22.20	.00	8.00	.50	.00	.70	.00	.00
33	19.70	14.90	.10	.00	.10	.00	.00	2.90
34	25.40	8.80	.30	.00	.00	3.00	.00	.00
35	29.30	64.40	3.60	.20	6.00	.00	.00	.10
36	1.40	.10	13.40	.00	.00	1.60	.00	.00

داده های بارش

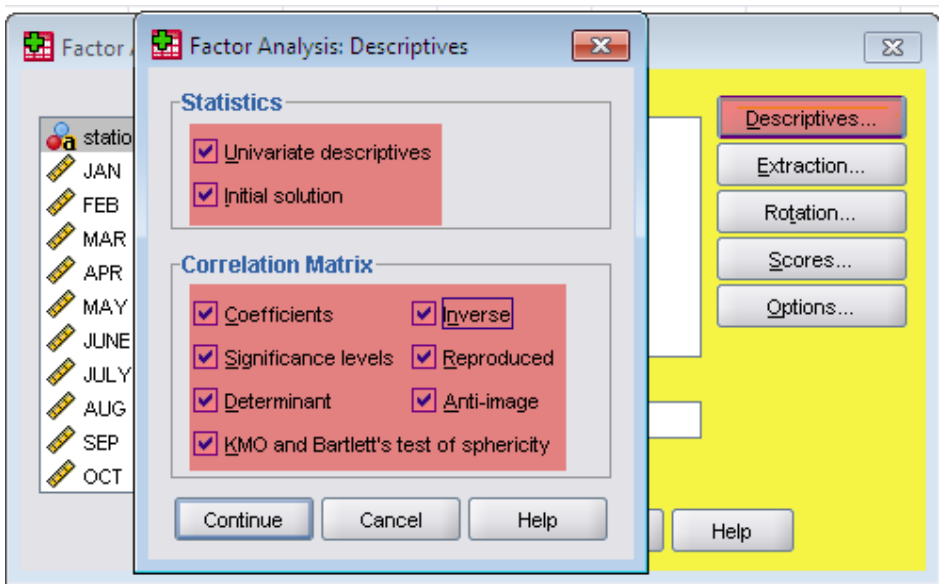
ردیف	NOV.P	DEC.P	ANNUAL.P
1	.00	10.99	143.36
2	6.06	78.35	134.66
3	.00	48.67	499.46
4	1.99	21.37	338.01
5	.00	148.85	252.43
6	6.09	85.72	309.48
7	25.03	77.21	443.90
8	72.50	43.52	244.82
9	3.70	26.40	150.30
10	3.00	78.20	236.80
11	.40	19.20	170.70
12	1.50	1.00	86.50
13	8.50	5.90	120.00
14	.10	7.40	78.70
15	13.10	33.70	169.60
16	8.70	16.00	62.70
17	34.20	125.80	282.60
18	.00	11.40	58.10
19	.00	1.90	105.60
20	16.90	13.80	85.10

21	1.60	10.40	70.10
22	.00	132.70	247.80
23	.30	45.80	116.30
24	5.10	.10	119.40
25	77.50	27.80	216.10
26	.00	27.40	190.90
27	18.20	1.00	126.90
28	5.30	9.90	121.30
29	.00	.10	57.70
30	.10	8.40	134.30
31	25.80	27.00	95.10
32	2.00	92.20	140.50
33	12.90	6.40	127.20
34	.00	86.40	159.50
35	8.10	92.60	330.50
36	4.40	13.60	67.10

مرحله اول : فعال کردن برنامه تحلیل عاملی در SPSS :



مرحله دوم : محاسبه آماره های توصیفی :



با فعال کردن کلید Descriptives امکان محاسبه آماره توصیفی تحلیل عاملی فراهم می شود. در قسمت statistics آماره های univariste descriptive آماره های تک متغیری از قبیل میانگین ، انحراف معیار و تعداد مشاهدات مورد استفاده مشخص می شود . یعنی :

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
JAN	4.0028	1.75897	36
FEB	6.2861	2.16915	36
MAR	11.6694	1.84000	36
APR	17.5833	1.94973	36
MAY	23.1667	1.84809	36
JUNE	27.6333	1.53697	36
JULY	28.7667	1.38378	36
AUG	26.7444	1.59829	36
SEP	23.0333	1.72411	36
OCT	17.3056	1.87509	36
NOV	10.6028	1.71039	36
DEC	5.7028	1.69714	36
ANNUAL	16.8748	.91018	36
JAN.P	31.4900	34.88713	36
FEB.P	23.0794	27.53543	36
MAR.P	34.3928	32.96825	36
APR.P	20.1606	33.06216	36
MAY.P	8.3097	13.81637	36
JUNE.P	.6444	1.91885	36

JULY.P	1.2624	3.45946	36
AUG.P	.3833	1.10802	36
SEP.P	1.0889	6.49913	36
OCT.P	4.0014	8.07279	36
NOV.P	10.0853	18.05867	36
DEC.P	39.9217	41.70867	36
ANNUAL.P	174.8199	107.22661	36

با فعال کردن گزینه Initial Solution برآورد های اولیه از عامل ها محاسبه می شود. در خروجی آن برآورد های اولیه از میزان اشتراک ها ، مقادیر ویژه ماتریس همبستگی متغیرها ، درصد کل واریانس توصیح داده می شود و همچنین با عامل های مشترک و نیر درصد تجمعی واریانس عامل ها مشخص می شود. یعنی :

Correlation Matrix

		JAN	FEB	MAR	APR	MAY	JUNE
Correlation	JAN	1.000	.336	-.330	.116	-.111	.068
	FEB	.336	1.000	.086	.154	.191	.116
	MAR	-.330	.086	1.000	.198	.103	.078
	APR	.116	.154	.198	1.000	.346	.340
	MAY	-.111	.191	.103	.346	1.000	.357
	JUNE	.068	.116	.078	.340	.357	1.000
	JULY	.010	.076	.164	.111	.423	.601
	AUG	-.183	-.067	.144	.059	.348	.448
	SEP	.200	-.060	-.007	.037	.291	.530
	OCT	.328	.161	.093	.313	.172	.664

	NOV	-.247	-.162	.012	.155	.321	.426
	DEC	.174	.210	.312	.119	.221	.317
	ANNUAL	.234	.369	.303	.493	.583	.764
	JAN.P	-.156	-.066	-.072	-.494	-.488	-.399
	FEB.P	.020	.057	.037	-.315	.039	.135
	MAR.P	.051	-.155	.076	-.337	-.273	-.247
	APR.P	.141	-.142	.136	.070	-.170	.045
	MAY.P	-.308	-.030	.203	.077	.238	.108
	JUNE.P	-.388	-.133	.129	.085	.121	.036
	JULY.P	.087	-.167	.048	-.043	-.165	.124
	AUG.P	-.190	-.051	.031	.106	.341	.281
	SEP.P	.056	.302	.153	-.043	.003	-.216
	OCT.P	.163	.064	-.145	.214	.026	.153
	NOV.P	.183	.338	.050	.209	-.023	.061
	DEC.P	-.080	-.110	-.183	-.082	.129	.042
	ANNUAL.P	-.017	-.073	.018	-.296	-.206	-.111

ژ

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.328
Bartlett's Test of Sphericity	Approx. Chi-Square	423.834
	df	276
	Sig.	.000

Correlation Matrix[a,b]

		JULY	AUG	SEP	OCT	NOV	DEC
Correlation	JAN	.010	-.183	.200	.328	-.247	.174
	FEB	.076	-.067	-.060	.161	-.162	.210
	MAR	.164	.144	-.007	.093	.012	.312
	APR	.111	.059	.037	.313	.155	.119
	MAY	.423	.348	.291	.172	.321	.221
	JUNE	.601	.448	.530	.664	.426	.317
	JULY	1.000	.429	.695	.473	.338	.371
	AUG	.429	1.000	.581	.188	.293	.030
	SEP	.695	.581	1.000	.420	.314	.270
	OCT	.473	.188	.420	1.000	.235	.408
	NOV	.338	.293	.314	.235	1.000	.234
	DEC	.371	.030	.270	.408	.234	1.000
	ANNUAL	.711	.489	.644	.704	.440	.581
	JAN.P	-.277	-.230	-.203	-.301	-.245	-.067
	FEB.P	.270	.269	.411	.172	.089	.171
	MAR.P	-.072	-.074	-.091	.063	-.253	.124
	APR.P	.036	-.018	.113	.082	-.285	.199
	MAY.P	-.081	.024	-.209	.029	-.026	.000

	JUNE.P	.011	.138	-.161	-.089	-.221	-.418
	JULY.P	.105	-.041	.234	.042	-.161	.184
	AUG.P	.059	.118	-.092	.131	.348	.297
	SEP.P	-.169	-.240	-.292	-.066	-.242	-.013
	OCT.P	.223	-.120	.121	.215	.179	.017
	NOV.P	.120	-.207	-.082	.217	.031	.250
	DEC.P	.091	.074	.136	.144	-.168	-.045
	ANNUAL.P	.024	-.058	.053	.101	-.293	.149

Correlation Matrix[a,b]

		ANNUAL	JAN.P	FEB.P	MAR.P	APR.P	MAY.P
Correlation	JAN	.234	-.156	.020	.051	.141	-.308
	FEB	.369	-.066	.057	-.155	-.142	-.030
	MAR	.303	-.072	.037	.076	.136	.203
	APR	.493	-.494	-.315	-.337	.070	.077
	MAY	.583	-.488	.039	-.273	-.170	.238
	JUNE	.764	-.399	.135	-.247	.045	.108
	JULY	.711	-.277	.270	-.072	.036	-.081
	AUG	.489	-.230	.269	-.074	-.018	.024
	SEP	.644	-.203	.411	-.091	.113	-.209
	OCT	.704	-.301	.172	.063	.082	.029
	NOV	.440	-.245	.089	-.253	-.285	-.026
	DEC	.581	-.067	.171	.124	.199	.000

	ANNUAL	1.000	-.478	.199	-.194	.028	.009
	JAN.P	-.478	1.000	.229	.474	.116	-.001
	FEB.P	.199	.229	1.000	.479	.176	-.097
	MAR.P	-.194	.474	.479	1.000	.267	-.002
	APR.P	.028	.116	.176	.267	1.000	-.053
	MAY.P	.009	-.001	-.097	-.002	-.053	1.000
	JUNE.P	-.145	.047	-.204	-.070	.007	.580
	JULY.P	.026	.032	.243	.217	.860	-.141
	AUG.P	.214	-.169	-.202	-.052	-.043	.280
	SEP.P	-.097	-.139	-.091	.100	-.089	-.100
	OCT.P	.176	-.266	-.066	.026	-.122	.169
	NOV.P	.200	-.211	-.071	.105	-.027	-.023
	DEC.P	-.017	.104	.165	.136	.082	.127
	ANNUAL.P	-.120	.542	.564	.750	.507	.148

Correlation Matrix[a,b]

		JUNE.P	JULY.P	AUG.P	SEP.P	OCT.P	NOV.P
Correlation	JAN	-.388	.087	-.190	.056	.163	.183
	FEB	-.133	-.167	-.051	.302	.064	.338
	MAR	.129	.048	.031	.153	-.145	.050
	APR	.085	-.043	.106	-.043	.214	.209
	MAY	.121	-.165	.341	.003	.026	-.023
	JUNE	.036	.124	.281	-.216	.153	.061

	JULY	.011	.105	.059	-.169	.223	.120
	AUG	.138	-.041	.118	-.240	-.120	-.207
	SEP	-.161	.234	-.092	-.292	.121	-.082
	OCT	-.089	.042	.131	-.066	.215	.217
	NOV	-.221	-.161	.348	-.242	.179	.031
	DEC	-.418	.184	.297	-.013	.017	.250
	ANNUAL	-.145	.026	.214	-.097	.176	.200
	JAN.P	.047	.032	-.169	-.139	-.266	-.211
	FEB.P	-.204	.243	-.202	-.091	-.066	-.071
	MAR.P	-.070	.217	-.052	.100	.026	.105
	APR.P	.007	.860	-.043	-.089	-.122	-.027
	MAY.P	.580	-.141	.280	-.100	.169	-.023
	JUNE.P	1.000	-.107	.023	-.054	-.060	-.087
	JULY.P	-.107	1.000	-.086	-.033	-.018	-.046
	AUG.P	.023	-.086	1.000	-.059	-.108	.101
	SEP.P	-.054	-.033	-.059	1.000	.361	.592
	OCT.P	-.060	-.018	-.108	.361	1.000	.496
	NOV.P	-.087	-.046	.101	.592	.496	1.000
	DEC.P	.052	.015	.397	.016	-.074	.044
	ANNUAL.P	.030	.410	.068	.112	.038	.191

Correlation Matrix[a,b]

	DEC.P	ANNUAL.P

Correlation	JAN	-.080	-.017
	FEB	-.110	-.073
	MAR	-.183	.018
	APR	-.082	-.296
	MAY	.129	-.206
	JUNE	.042	-.111
	JULY	.091	.024
	AUG	.074	-.058
	SEP	.136	.053
	OCT	.144	.101
	NOV	-.168	-.293
	DEC	-.045	.149
	ANNUAL	-.017	-.120
	JAN.P	.104	.542
	FEB.P	.165	.564
	MAR.P	.136	.750
	APR.P	.082	.507
	MAY.P	.127	.148
	JUNE.P	.052	.030
	JULY.P	.015	.410
	AUG.P	.397	.068
	SEP.P	.016	.112

	OCT.P	-.074	.038
	NOV.P	.044	.191
	DEC.P	1.000	.557
	ANNUAL.P	.557	1.000

Communalities

	Initial	Extraction
JAN	1.000	.830
FEB	1.000	.769
MAR	1.000	.809
APR	1.000	.629
MAY	1.000	.678
JUNE	1.000	.754
JULY	1.000	.712
AUG	1.000	.688
SEP	1.000	.849
OCT	1.000	.737
NOV	1.000	.849
DEC	1.000	.810
ANNUAL	1.000	.991
JAN.P	1.000	.760
FEB.P	1.000	.748
MAR.P	1.000	.728

APR.P	1.000	.928
MAY.P	1.000	.782
JUNE.P	1.000	.852
JULY.P	1.000	.881
AUG.P	1.000	.869
SEP.P	1.000	.857
OCT.P	1.000	.850
NOV.P	1.000	.780
DEC.P	1.000	.843
ANNUAL.P	1.000	.978

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.349	20.573	20.573	5.349	20.573	20.573
2	3.538	13.609	34.183	3.538	13.609	34.183
3	2.745	10.558	44.740	2.745	10.558	44.740
4	2.272	8.738	53.478	2.272	8.738	53.478
5	1.745	6.713	60.191	1.745	6.713	60.191
6	1.491	5.734	65.925	1.491	5.734	65.925
7	1.452	5.584	71.509	1.452	5.584	71.509
8	1.289	4.956	76.465	1.289	4.956	76.465

9	1.081	4.159	80.624	1.081	4.159	80.624
10	.758	2.915	83.539			
11	.683	2.629	86.167			
12	.601	2.311	88.478			
13	.584	2.245	90.723			
14	.495	1.904	92.627			
15	.448	1.723	94.350			
16	.328	1.261	95.612			
17	.283	1.087	96.698			
18	.250	.963	97.661			
19	.199	.766	98.427			
20	.138	.533	98.960			
21	.124	.476	99.436			
22	.068	.263	99.698			
23	.051	.195	99.893			
24	.028	.107	100.000			
25	2.529E-10	9.725E-10	100.000			
26	3.229E-16	1.242E-15	100.000			

Extraction Method: Principal Component Analysis.

Component Matrixa

	Component						
	1	2	3	4	5	6	7
JAN	.144	.178	.604	-.382	.049	.283	-.114

FEB	.221	-.088	.525	.081	-.093	-.258	.144
MAR	.201	.044	-.072	.338	.341	-.687	.230
APR	.477	-.330	.202	.161	.446	.112	-.075
MAY	.617	-.237	-.183	.244	-.077	-.034	-.015
JUNE	.821	.065	-.125	.026	.076	.151	-.013
JULY	.752	.234	-.113	-.041	-.099	.036	.239
AUG	.526	.081	-.498	-.086	-.094	-.008	.277
SEP	.675	.364	-.197	-.353	-.098	.181	.206
OCT	.682	.249	.188	.077	-.013	.152	-.038
NOV	.547	-.206	-.227	-.126	-.305	-.156	-.206
DEC	.508	.338	.232	.082	.023	-.475	-.359
ANNUAL	.967	.091	.091	.018	.036	-.126	.040
JAN.P	-.591	.428	-.190	.007	-.222	-.204	.011
FEB.P	.145	.694	-.141	-.117	-.331	-.147	.275
MAR.P	-.295	.698	.065	.217	-.201	-.129	.086
APR.P	-.034	.649	.007	.034	.681	.092	-.131
MAY.P	.024	-.149	-.302	.705	.090	.072	.112
JUNE.P	-.120	-.219	-.424	.518	.259	.242	.428
JULY.P	.013	.634	.023	-.098	.595	.116	-.091
AUG.P	.282	-.090	-.211	.482	-.147	-.014	-.689
SEP.P	-.143	-.061	.645	.373	-.115	-.091	.214
OCT.P	.233	-.061	.485	.209	-.153	.375	.291
NOV.P	.177	.035	.705	.413	-.146	.035	.051

DEC.P	.016	.334	-.191	.401	-.307	.433	-.310
ANNUAL.P	-.207	.856	-.003	.415	-.137	.099	-.005

Extraction Method: Principal Component Analysis.

a. 9 components extracted.

Component Matrixa

	Component	
	8	9
JAN	.362	.200
FEB	.579	.002
MAR	-.051	-.053
APR	.043	.083
MAY	.206	-.314
JUNE	-.002	.176
JULY	-.077	-.062
AUG	.077	-.240
SEP	-.019	-.115
OCT	.097	.367
NOV	-.520	.101
DEC	-.034	.146
ANNUAL	.135	.050
JAN.P	.094	.302
FEB.P	.054	-.048
MAR.P	-.095	.171

APR.P	-.051	-.111
MAY.P	.051	.377
JUNE.P	.135	.121
JULY.P	-.227	-.205
AUG.P	-.023	-.089
SEP.P	-.105	-.448
OCT.P	-.489	.156
NOV.P	-.215	-.090
DEC.P	.267	-.291
ANNUAL.P	.025	.004

Extraction Method: Principal Component Analysis.

a. 9 components extracted.

Reproduced Correlations

		JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG
Reproduced Correlation	JAN	.830[a]	.418	-.369	.165	-.157	.127	.035	-.237
	FEB	.418	.769[a]	.180	.198	.213	.064	.072	-.065
	MAR	-.369	.180	.809[a]	.173	.210	.096	.159	.162
	APR	.165	.198	.173	.629[a]	.321	.416	.185	.030
	MAY	-.157	.213	.210	.321	.678[a]	.454	.425	.470
	JUNE	.127	.064	.096	.416	.454	.754[a]	.630	.443

	JULY	.035	.072	.159	.185	.425	.630	.712[a]	.558
	AUG	-.237	-.065	.162	.030	.470	.443	.558	.688[a]
	SEP	.171	-.034	-.056	.056	.310	.590	.703	.603
	OCT	.382	.247	.019	.342	.247	.642	.514	.187
	NOV	-.251	-.252	.051	.111	.290	.436	.404	.304
	DEC	.166	.260	.374	.186	.178	.373	.320	.036
	ANNUAL	.224	.368	.297	.462	.575	.781	.725	.476
	JAN.P	-.102	-.137	-.038	-.554	-.481	-.429	-.331	-.222
	FEB.P	.025	.027	.082	-.393	-.025	.120	.385	.337
	MAR.P	-.018	-.065	.075	-.424	-.361	-.205	-.042	-.155
	APR.P	.136	-.194	.180	.089	-.196	.063	.040	-.050
	MAY.P	-.369	-.064	.243	.186	.158	.148	-.018	.026
	JUNE.P	-.405	-.134	.180	.139	.138	.025	-.047	.160
	JULY.P	.115	-.279	.120	.044	-.201	.075	.114	.005
	AUG.P	-.246	-.097	.038	.175	.396	.244	.052	.047
	SEP.P	.033	.345	.150	.023	.027	-.296	-.115	-.263
	OCT.P	.156	.005	-.147	.207	-.034	.202	.224	-.125
	NOV.P	.201	.326	.080	.200	.066	.049	.095	-.258
	DEC.P	-.026	-.068	-.304	-.155	.223	.063	.065	.126
	ANNUAL.P	-.007	-.088	.019	-.364	-.217	-.097	.043	-.062
Residual[b]	JAN		-.083	.039	-.049	.046	-.059	-.025	.054

	FEB	-.083		-.094	-.043	-.021	.053	.004	-.002
	MAR	.039	-.094		.025	-.106	-.019	.005	-.018
	APR	-.049	-.043	.025		.025	-.075	-.074	.029
	MAY	.046	-.021	-.106	.025		-.097	-.002	-.122
	JUNE	-.059	.053	-.019	-.075	-.097		-.029	.005
	JULY	-.025	.004	.005	-.074	-.002	-.029		-.129
	AUG	.054	-.002	-.018	.029	-.122	.005	-.129	
	SEP	.029	-.026	.050	-.019	-.019	-.060	-.008	-.023
	OCT	-.054	-.085	.074	-.029	-.076	.022	-.041	.001
	NOV	.004	.090	-.040	.044	.030	-.010	-.066	-.011
	DEC	.007	-.051	-.061	-.067	.043	-.056	.052	-.006
	ANNUAL	.010	.000	.007	.031	.008	-.016	-.013	.014
	JAN.P	-.054	.071	-.033	.060	-.007	.030	.055	-.008
	FEB.P	-.004	.029	-.044	.078	.064	.015	-.115	-.067
	MAR.P	.070	-.089	.000	.087	.088	-.043	-.030	.081
	APR.P	.006	.052	-.044	-.019	.026	-.018	-.004	.032
	MAY.P	.062	.034	-.040	-.109	.079	-.040	-.063	-.001
	JUNE.P	.017	.001	-.051	-.054	-.017	.011	.058	-.022
	JULY.P	-.028	.112	-.072	-.086	.036	.049	-.009	-.046
	AUG.P	.056	.046	-.007	-.069	-.055	.037	.007	.071
	SEP.P	.023	-.044	.003	-.066	-.023	.080	-.054	.023
	OCT.P	.007	.060	.001	.007	.060	-.049	.000	.005

	NOV.P	-.019	.012	-.029	.009	-.089	.012	.025	.051
	DEC.P	-.055	-.042	.122	.073	-.094	-.022	.026	-.052
	ANNUAL.P	-.010	.015	-.002	.068	.011	-.013	-.018	.004

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 94 (28.0%) nonredundant residuals with absolute values greater than 0.05.

Reproduced Correlations

		SEP	OCT	NOV	DEC	ANNUAL	JAN.P	FEB.P	MAR.P
Reproduced Correlation	JAN	.171	.382	-.251	.166	.224	-.102	.025	-.018
	FEB	-.034	.247	-.252	.260	.368	-.137	.027	-.065
	MAR	-.056	.019	.051	.374	.297	-.038	.082	.075
	APR	.056	.342	.111	.186	.462	-.554	-.393	-.424
	MAY	.310	.247	.290	.178	.575	-.481	-.025	-.361
	JUNE	.590	.642	.436	.373	.781	-.429	.120	-.205
	JULY	.703	.514	.404	.320	.725	-.331	.385	-.042
	AUG	.603	.187	.304	.036	.476	-.222	.337	-.155
	SEP	.849[a]	.463	.341	.213	.634	-.258	.486	-.038
	OCT	.463	.737[a]	.244	.471	.711	-.240	.195	.035
	NOV	.341	.244	.849[a]	.318	.422	-.291	.018	-.217
	DEC	.213	.471	.318	.810[a]	.593	-.070	.221	.173

	ANNUAL	.634	.711	.422	.593	.991[a]	-.504	.211	-.204
	JAN.P	-.258	-.240	-.291	-.070	-.504	.760[a]	.334	.577
	FEB.P	.486	.195	.018	.221	.211	.334	.748[a]	.503
	MAR.P	-.038	.035	-.217	.173	-.204	.577	.503	.728[a]
	APR.P	.137	.107	-.337	.212	.023	.087	.168	.296
	MAY.P	-.245	.126	-.026	-.069	.019	.069	-.165	.064
	JUNE.P	-.170	-.102	-.252	-.440	-.145	.008	-.172	-.071
	JULY.P	.242	.079	-.199	.187	.033	.019	.209	.263
	AUG.P	-.091	.158	.352	.342	.215	-.163	-.184	-.096
	SEP.P	-.285	-.158	-.245	-.011	-.099	-.160	-.072	.109
	OCT.P	.100	.309	.202	-.020	.168	-.293	-.060	.046
	NOV.P	-.106	.245	.009	.255	.205	-.242	-.047	.142
	DEC.P	.101	.090	-.145	-.043	-.021	.087	.195	.207
	ANNUAL.P	.055	.124	-.326	.169	-.129	.506	.547	.761
Residual[b]	JAN	.029	-.054	.004	.007	.010	-.054	-.004	.070
	FEB	-.026	-.085	.090	-.051	.000	.071	.029	-.089
	MAR	.050	.074	-.040	-.061	.007	-.033	-.044	.000
	APR	-.019	-.029	.044	-.067	.031	.060	.078	.087
	MAY	-.019	-.076	.030	.043	.008	-.007	.064	.088
	JUNE	-.060	.022	-.010	-.056	-.016	.030	.015	-.043
	JULY	-.008	-.041	-.066	.052	-.013	.055	-.115	-.030

	AUG	-.023	.001	-.011	-.006	.014	-.008	-.067	.081
	SEP		-.043	-.027	.057	.010	.054	-.075	-.052
	OCT	-.043		-.010	-.064	-.006	-.060	-.023	.029
	NOV	-.027	-.010		-.084	.018	.045	.071	-.036
	DEC	.057	-.064	-.084		-.012	.004	-.049	-.049
	ANNUAL	.010	-.006	.018	-.012		.025	-.012	.010
	JAN.P	.054	-.060	.045	.004	.025		-.105	-.103
	FEB.P	-.075	-.023	.071	-.049	-.012	-.105		-.024
	MAR.P	-.052	.029	-.036	-.049	.010	-.103	-.024	
	APR.P	-.024	-.025	.052	-.012	.005	.029	.008	-.030
	MAY.P	.036	-.097	.000	.069	-.010	-.070	.068	-.066
	JUNE.P	.009	.012	.031	.023	.000	.039	-.032	.001
	JULY.P	-.007	-.037	.038	-.003	-.007	.013	.034	-.047
	AUG.P	-.001	-.028	-.004	-.045	.000	-.006	-.018	.044
	SEP.P	-.007	.092	.003	-.001	.003	.021	-.020	-.009
	OCT.P	.022	-.094	-.023	.037	.008	.027	-.007	-.020
	NOV.P	.024	-.029	.022	-.005	-.005	.031	-.024	-.037
	DEC.P	.035	.054	-.023	-.002	.004	.017	-.031	-.071
	ANNUAL.P	-.002	-.023	.033	-.020	.009	.036	.017	-.011

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 94 (28.0%) nonredundant residuals with absolute values greater than 0.05.

Reproduced Correlations

		APR.P	MAY.P	JUNE.P	JULY.P	AUG.P	SEP.P	OCT.P	NOV.P
Reproduced Correlation	JAN	.136	-.369	-.405	.115	-.246	.033	.156	.201
	FEB	-.194	-.064	-.134	-.279	-.097	.345	.005	.326
	MAR	.180	.243	.180	.120	.038	.150	-.147	.080
	APR	.089	.186	.139	.044	.175	.023	.207	.200
	MAY	-.196	.158	.138	-.201	.396	.027	-.034	.066
	JUNE	.063	.148	.025	.075	.244	-.296	.202	.049
	JULY	.040	-.018	-.047	.114	.052	-.115	.224	.095
	AUG	-.050	.026	.160	.005	.047	-.263	-.125	-.258
	SEP	.137	-.245	-.170	.242	-.091	-.285	.100	-.106
	OCT	.107	.126	-.102	.079	.158	-.158	.309	.245
	NOV	-.337	-.026	-.252	-.199	.352	-.245	.202	.009
	DEC	.212	-.069	-.440	.187	.342	-.011	-.020	.255
	ANNUAL	.023	.019	-.145	.033	.215	-.099	.168	.205
	JAN.P	.087	.069	.008	.019	-.163	-.160	-.293	-.242
	FEB.P	.168	-.165	-.172	.209	-.184	-.072	-.060	-.047
	MAR.P	.296	.064	-.071	.263	-.096	.109	.046	.142
	APR.P	.928[a]	-.067	-.002	.870	-.053	-.078	-.138	-.046

	MAY.P	-.067	.782[a]	.664	-.208	.298	-.093	.096	.028
	JUNE.P	-.002	.664	.852[a]	-.114	-.025	-.078	.016	-.161
	JULY.P	.870	-.208	-.114	.881[a]	-.108	-.045	-.040	-.020
	AUG.P	-.053	.298	-.025	-.108	.869[a]	-.078	-.116	.096
	SEP.P	-.078	-.093	-.078	-.045	-.078	.857[a]	.388	.668
	OCT.P	-.138	.096	.016	-.040	-.116	.388	.850[a]	.609
	NOV.P	-.046	.028	-.161	-.020	.096	.668	.609	.780[a]
	DEC.P	.119	.164	.106	.063	.480	.035	-.082	.058
	ANNUAL.P	.491	.158	.044	.423	.087	.131	.029	.180
Residual[b]	JAN	.006	.062	.017	-.028	.056	.023	.007	-.019
	FEB	.052	.034	.001	.112	.046	-.044	.060	.012
	MAR	-.044	-.040	-.051	-.072	-.007	.003	.001	-.029
	APR	-.019	-.109	-.054	-.086	-.069	-.066	.007	.009
	MAY	.026	.079	-.017	.036	-.055	-.023	.060	-.089
	JUNE	-.018	-.040	.011	.049	.037	.080	-.049	.012
	JULY	-.004	-.063	.058	-.009	.007	-.054	.000	.025
	AUG	.032	-.001	-.022	-.046	.071	.023	.005	.051
	SEP	-.024	.036	.009	-.007	-.001	-.007	.022	.024
	OCT	-.025	-.097	.012	-.037	-.028	.092	-.094	-.029
	NOV	.052	.000	.031	.038	-.004	.003	-.023	.022
	DEC	-.012	.069	.023	-.003	-.045	-.001	.037	-.005
	ANNUAL	.005	-.010	.000	-.007	.000	.003	.008	-.005

	JAN.P	.029	-.070	.039	.013	-.006	.021	.027	.031
	FEB.P	.008	.068	-.032	.034	-.018	-.020	-.007	-.024
	MAR.P	-.030	-.066	.001	-.047	.044	-.009	-.020	-.037
	APR.P		.014	.009	-.010	.010	-.012	.016	.019
	MAY.P	.014		-.084	.067	-.018	-.007	.073	-.051
	JUNE.P	.009	-.084		.007	.048	.024	-.076	.074
	JULY.P	-.010	.067	.007		.023	.012	.023	-.026
	AUG.P	.010	-.018	.048	.023		.019	.008	.005
	SEP.P	-.012	-.007	.024	.012	.019		-.028	-.076
	OCT.P	.016	.073	-.076	.023	.008	-.028		-.113
	NOV.P	.019	-.051	.074	-.026	.005	-.076	-.113	
	DEC.P	-.037	-.036	-.055	-.049	-.083	-.020	.008	-.014
	ANNUAL.P	.016	-.011	-.013	-.013	-.019	-.019	.008	.011

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 94 (28.0%) nonredundant residuals with absolute values greater than 0.05.

Reproduced Correlations

		DEC.P	ANNUAL.P
Reproduced Correlation	JAN	-.026	-.007
	FEB	-.068	-.088

MAR	-.304	.019
APR	-.155	-.364
MAY	.223	-.217
JUNE	.063	-.097
JULY	.065	.043
AUG	.126	-.062
SEP	.101	.055
OCT	.090	.124
NOV	-.145	-.326
DEC	-.043	.169
ANNUAL	-.021	-.129
JAN.P	.087	.506
FEB.P	.195	.547
MAR.P	.207	.761
APR.P	.119	.491
MAY.P	.164	.158
JUNE.P	.106	.044
JULY.P	.063	.423
AUG.P	.480	.087
SEP.P	.035	.131
OCT.P	-.082	.029
NOV.P	.058	.180

	DEC.P	.843[a]	.542
	ANNUAL.P	.542	.978[a]
Residual[b]	JAN	-.055	-.010
	FEB	-.042	.015
	MAR	.122	-.002
	APR	.073	.068
	MAY	-.094	.011
	JUNE	-.022	-.013
	JULY	.026	-.018
	AUG	-.052	.004
	SEP	.035	-.002
	OCT	.054	-.023
	NOV	-.023	.033
	DEC	-.002	-.020
	ANNUAL	.004	.009
	JAN.P	.017	.036
	FEB.P	-.031	.017
	MAR.P	-.071	-.011
	APR.P	-.037	.016
	MAY.P	-.036	-.011
	JUNE.P	-.055	-.013
	JULY.P	-.049	-.013

	AUG.P	*-.083*	*-.019*
	SEP.P	*-.020*	*-.019*
	OCT.P	*.008*	*.008*
	NOV.P	*-.014*	*.011*
	DEC.P		*.014*
	ANNUAL.P	*.014*	

Extraction Method: Principal Component Analysis.

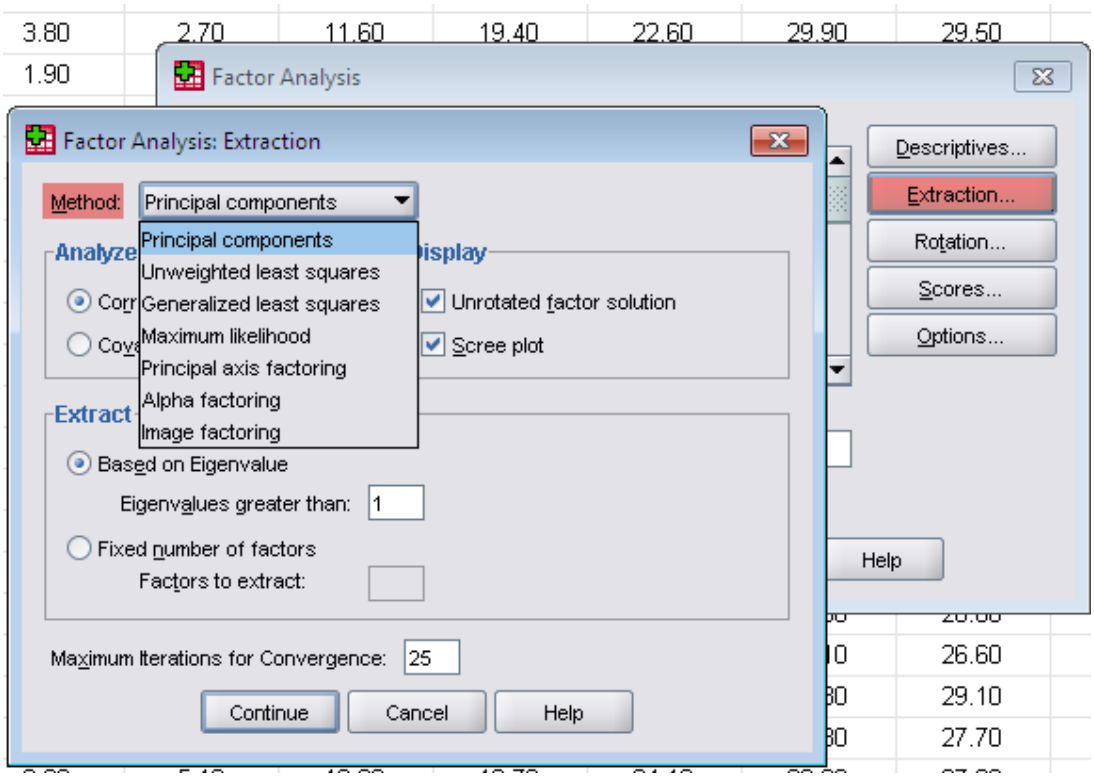
a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 94 (28.0%) nonredundant residuals with absolute values greater than 0.05.

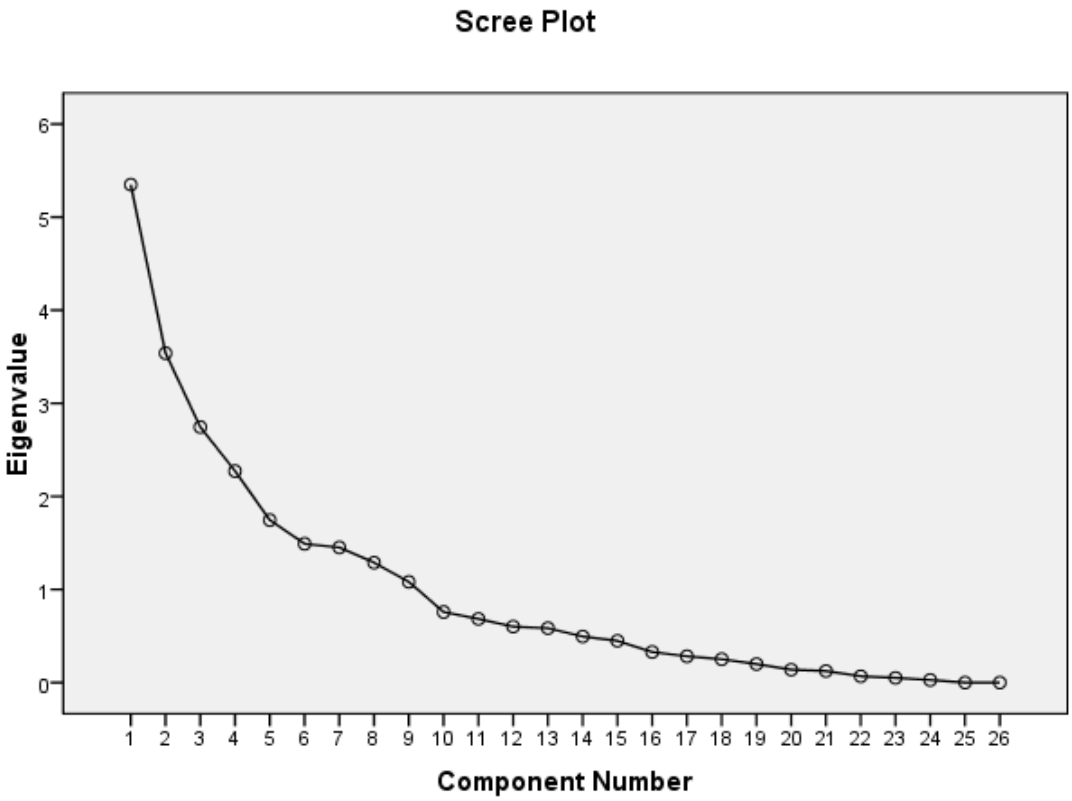
مرحلهٔ سوم : استخراج عامل ها :

برای تعیین استخراج عامل ها باید کلید Extraction را فعال کنید و روش مورد نظر را یعنی یکی از روش مؤلفه های اصلی[1] ، حداکثر درستنمایی[2] ، عامل یابی محور اصلی[3] ، حداقل مربعات غیر وزنی[4] ، حداقل مربعات تعمیم یافته[5] ، عامل یابی آلفا[6] و عامل یابی تصویر[7] انتخاب کنید . یعنی :

[1] - Principle Components
[2] - Maximum Likelihood
[3] - Principle-axis Factoring
[4] - Unweighted Least Squares
[5] - Generalized Least Squares
[6] - Alpha Factoring
[7] - Image Factoring



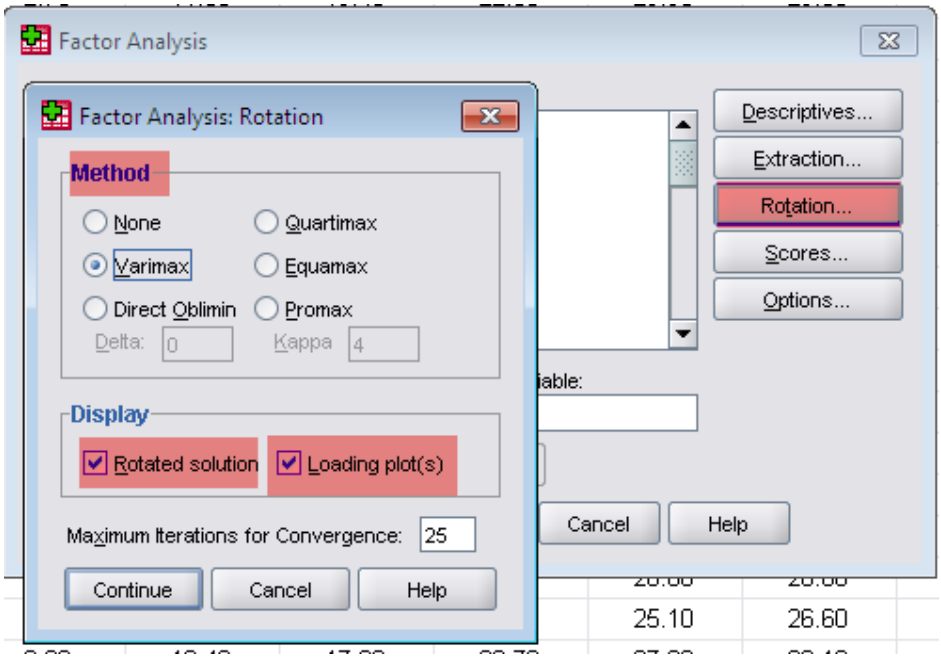
قایل اشاره است اگر هدف محقق خلاصه کردن متغیر ها و دستیابی به تعداد محدودی عامل بارش از روش مؤلفه های اصلی یا مبناء می توان استفاده کند. با انتخاب گزینه Unrotated factor solution ضرایب عامل ها قبل از دوران محاسبه می شوند و با انتخاب Scree Plot نمودار ترسیم شده مقادیر ویژه عوامل انتخابی را از بزرگ تا کوچکترین مقدار نشان می دهد. همچنین گزینه Maximum Iterations for Convergence حداکثر تعداد دفعات تکرار برای همگرایی مدل را نشان می دهد. یعنی :



مرحله چهارم : دوران عامل ها :

در این مرحله با فعال کردن دستور Rotation در کادر اصلی دوران عامل ها فعال می شود . در این قسمت بخش های روش ها (Method) یعنی واریماکس ، کوارتیماکس ، اکویماکس ، دوران مایل [1] ، Rotated Solution که ماتریس ضرایب عامل های دوران یافته ، ماتریس تبدیل عامل ها ، ماتریس الگو و ماتریس همبستگی بین برآوردهای دوران یافتهٔ عامل های مشترک محاسبه می شود و بخش Loading Plot(s) نمودار سه بعدی ضرایب متغیرهای موجود در ماتریس الگو را ترسیم می نماید . یعنی :

[1] - Direct Oblimin



خروجی قسمت دوران عامل ها:

Rotated Component Matrix[a]

	Component						
	1	*2*	*3*	*4*	*5*	*6*	*7*
JAN	*-.036*	*-.069*	*.432*	*.137*	*.091*	*-.364*	*-.184*
FEB	*.047*	*-.073*	*.222*	*-.254*	*.195*	*-.118*	*-.026*
MAR	*.108*	*-.004*	*.042*	*.115*	*.035*	*.178*	*-.082*
APR	*.094*	*-.623*	*.365*	*.145*	*.151*	*.182*	*.030*
MAY	*.530*	*-.356*	*.013*	*-.184*	*.033*	*.112*	*.407*
JUNE	*.625*	*-.234*	*.525*	*.053*	*-.008*	*.106*	*.105*
JULY	*.791*	*-.002*	*.249*	*.031*	*.117*	*-.019*	*-.001*
AUG	*.768*	*-.034*	*-.128*	*-.054*	*-.220*	*.103*	*.064*

SEP	.850	.050	.161	.143	-.081	-.187	-.082
OCT	.420	.000	.717	.056	.144	.043	.054
NOV	.360	-.183	.268	-.335	.022	-.227	.105
DEC	.150	.109	.607	.111	.069	-.357	.219
ANNUAL	.715	-.240	.560	-.014	.099	-.086	.096
JAN.P	-.353	.732	-.063	-.034	-.289	.051	-.087
FEB.P	.488	.682	.001	.075	-.042	-.159	-.066
MAR.P	-.111	.804	.079	.195	.142	.020	-.002
APR.P	-.008	.137	.098	.940	-.078	.009	.019
MAY.P	-.087	.039	.201	-.118	.021	.803	.201
JUNE.P	.034	-.080	-.201	.014	-.067	.893	-.025
JULY.P	.088	.116	.018	.911	.003	-.130	-.050
AUG.P	-.033	-.136	.236	-.081	-.026	.040	.845
SEP.P	-.161	7.110E-5	-.298	-.019	.781	-.101	.060
OCT.P	.106	-.059	.226	-.060	.779	.127	-.214
NOV.P	-.062	.003	.200	-.017	.834	-.073	.107
DEC.P	.154	.285	-.107	.108	.038	.144	.745
ANNUAL.P	.014	.816	.051	.405	.177	.147	.288

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 12 iterations.

Rotated Component Matrixa

	Component	
	8	9
JAN	-.445	.496
FEB	.187	.749
MAR	.859	.061
APR	.101	.103
MAY	.205	.122
JUNE	-.005	-.091
JULY	.073	-.068
AUG	.104	-.068
SEP	-.161	-.063
OCT	-.090	.103
NOV	.147	-.645
DEC	.463	.024
ANNUAL	.241	.157
JAN.P	.004	-.003
FEB.P	.050	.069
MAR.P	.066	-.009
APR.P	.081	.047
MAY.P	.158	-.087
JUNE.P	.026	.026
JULY.P	.036	-.095
AUG.P	.137	-.230

SEP.P	.186	.290
OCT.P	-.250	-.221
NOV.P	.075	.132
DEC.P	-.350	.122
ANNUAL.P	-.048	.083

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 12 iterations.

Total Variance Explained

Component	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	3.929	15.112	15.112
2	3.144	12.092	27.204
3	2.370	9.114	36.318
4	2.266	8.714	45.033
5	2.238	8.609	53.641
6	2.032	7.815	61.456
7	1.769	6.805	68.261
8	1.655	6.365	74.626
9	1.559	5.998	80.624

Extraction Method: Principal Component Analysis.

Component Transformation Matrix

Component	1	2	3	4	5	6	7
1	.761	-.332	.499	-.036	.115	-.071	.146

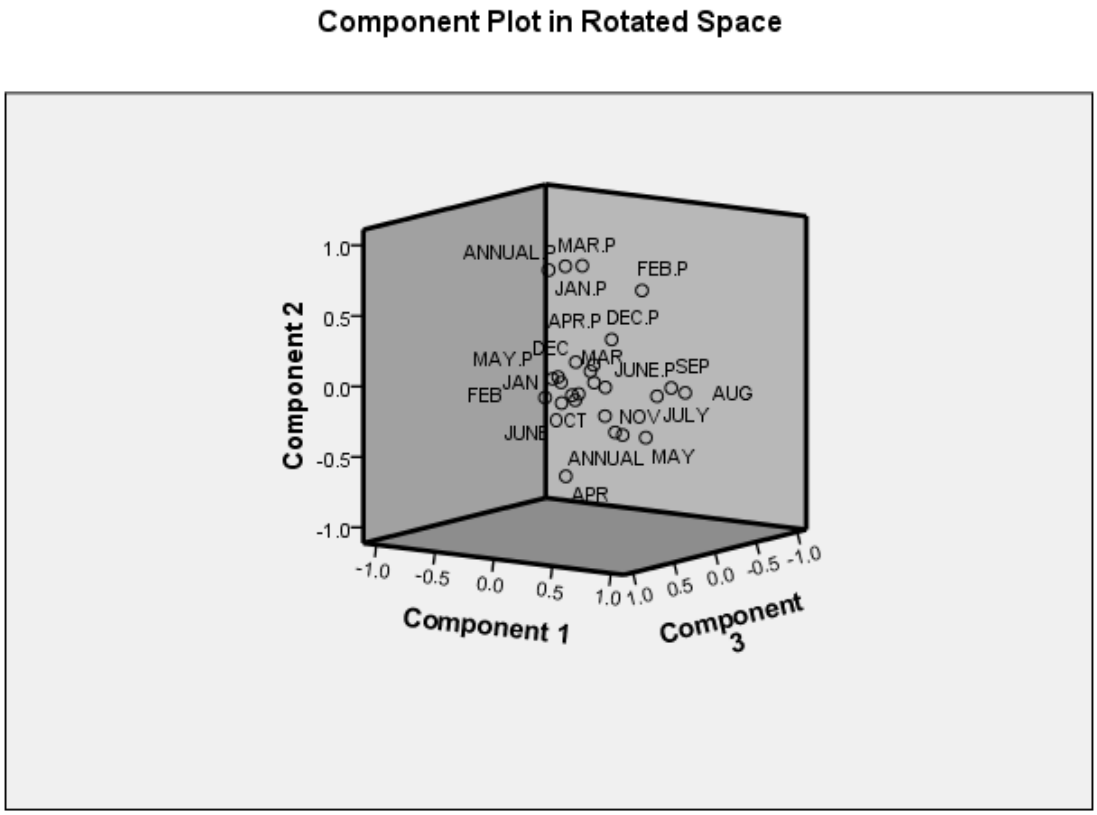
2	.230	.754	.161	.564	-.001	-.159	.062
3	-.298	-.098	.263	.040	.692	-.378	-.184
4	-.140	.130	.059	.000	.426	.647	.515
5	-.164	-.455	.077	.745	-.173	.231	-.236
6	.107	-.157	-.029	.213	.162	.285	.111
7	.387	.159	-.311	-.114	.245	.402	-.671
8	.041	.003	.046	-.133	-.410	.115	.184
9	-.266	.193	.739	-.221	-.204	.315	-.352

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Component Transformation Matrix

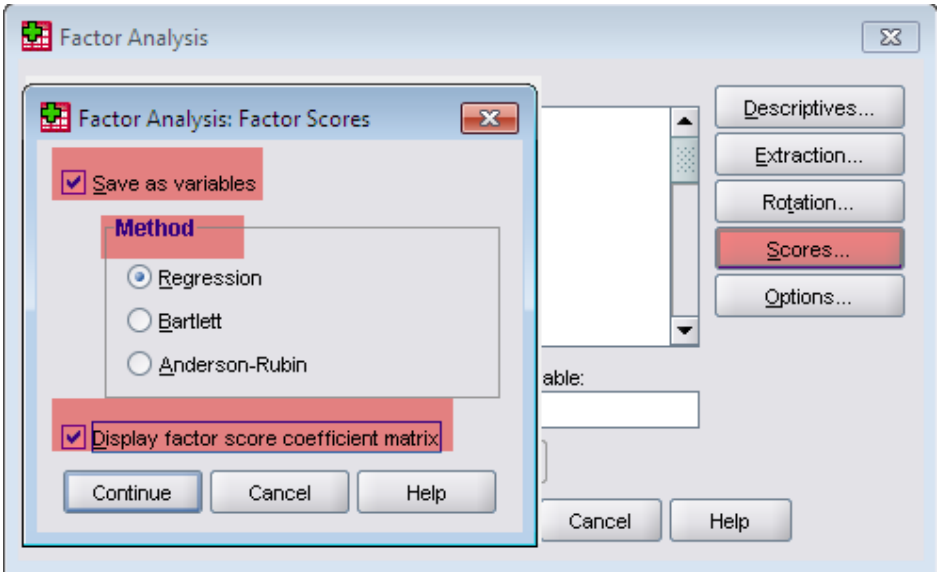
Component	8	9
1	.143	-.017
2	-.023	.069
3	-.062	.415
4	.304	.045
5	.243	.084
6	-.891	-.057
7	.078	.194
8	-.099	.868
9	-.115	-.138

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.



مرحلهٔ پنجم : محاسبهٔ نمره های عامل :

برای محاسبهٔ نمره های عاملی باید کلید Scores را در کار اصلی برنامه فعال کرد در این قسمت علاوه بر ذخیره کردن متغیرها ، قسمت روش ها که شامل رگرسیون ، بارتلت و روبین – اندرسون می باشد. و گزینه Diplay Factor Score Coefficient Matrix وجود دارد . روش های رگرسیون از روش کمترین مربعات بر ای برآورد نمره های عاملی و روش بارتلت از روش کمترین مربعات وزنی و روش روبین – اندرسون روش کمترین مربعات استفاده می کند. یعنی :



خروجی این مرحله :

Component Score Coefficient Matrix

	Component						
	1	*2*	*3*	*4*	*5*	*6*	*7*
JAN	*.027*	*.050*	*.220*	*-.168*	*.028*	*.190*	*-.078*
FEB	*.041*	*-.025*	*.191*	*.036*	*-.053*	*-.173*	*.099*
MAR	*.038*	*.012*	*-.026*	*.149*	*.196*	*-.461*	*.158*
APR	*.089*	*-.093*	*.074*	*.071*	*.255*	*.075*	*-.052*
MAY	*.115*	*-.067*	*-.067*	*.107*	*-.044*	*-.023*	*-.010*
JUNE	*.154*	*.018*	*-.045*	*.011*	*.043*	*.101*	*-.009*
JULY	*.141*	*.066*	*-.041*	*-.018*	*-.057*	*.024*	*.164*
AUG	*.098*	*.023*	*-.181*	*-.038*	*-.054*	*-.005*	*.191*
SEP	*.126*	*.103*	*-.072*	*-.155*	*-.056*	*.121*	*.142*

OCT	.127	.070	.068	.034	-.008	.102	-.026
NOV	.102	-.058	-.083	-.055	-.175	-.104	-.142
DEC	.095	.096	.085	.036	.013	-.319	-.247
ANNUAL	.181	.026	.033	.008	.021	-.085	.028
JAN.P	-.111	.121	-.069	.003	-.127	-.137	.008
FEB.P	.027	.196	-.051	-.051	-.189	-.099	.190
MAR.P	-.055	.197	.024	.096	-.115	-.086	.059
APR.P	-.006	.183	.002	.015	.390	.062	-.090
MAY.P	.004	-.042	-.110	.310	.051	.048	.077
JUNE.P	-.022	-.062	-.155	.228	.148	.162	.295
JULY.P	.002	.179	.008	-.043	.341	.078	-.063
AUG.P	.053	-.025	-.077	.212	-.084	-.010	-.475
SEP.P	-.027	-.017	.235	.164	-.066	-.061	.147
OCT.P	.044	-.017	.177	.092	-.088	.252	.200
NOV.P	.033	.010	.257	.182	-.083	.024	.035
DEC.P	.003	.095	-.069	.176	-.176	.290	-.214
ANNUAL.P	-.039	.242	-.001	.183	-.079	.066	-.003

Extraction Method: Principal Component Analysis.
Component Scores.

Component Score Coefficient Matrix

	Component	
	8	9
JAN	.281	.185

FEB	.449	.001
MAR	-.039	-.049
APR	.034	.077
MAY	.159	-.291
JUNE	-.001	.162
JULY	-.060	-.057
AUG	.060	-.222
SEP	-.015	-.106
OCT	.075	.339
NOV	-.403	.093
DEC	-.026	.135
ANNUAL	.105	.046
JAN.P	.073	.279
FEB.P	.042	-.044
MAR.P	-.074	.158
APR.P	-.040	-.103
MAY.P	.039	.349
JUNE.P	.105	.112
JULY.P	-.176	-.190
AUG.P	-.018	-.082
SEP.P	-.081	-.414
OCT.P	-.379	.144
NOV.P	-.167	-.083

DEC.P	.207	-.269
ANNUAL.P	.020	.004

Extraction Method: Principal Component Analysis.
Component Scores.

Component Score Covariance Matrix

Component	1	2	3	4	5	6	7
1	1.000	.000	.000	.000	.000	.000	.000
2	.000	1.000	.000	.000	.000	.000	.000
3	.000	.000	1.000	.000	.000	.000	.000
4	.000	.000	.000	1.000	.000	.000	.000
5	.000	.000	.000	.000	1.000	.000	.000
6	.000	.000	.000	.000	.000	1.000	.000
7	.000	.000	.000	.000	.000	.000	1.000
8	.000	.000	.000	.000	.000	.000	.000
9	.000	.000	.000	.000	.000	.000	.000

Extraction Method: Principal Component Analysis.
Component Scores.

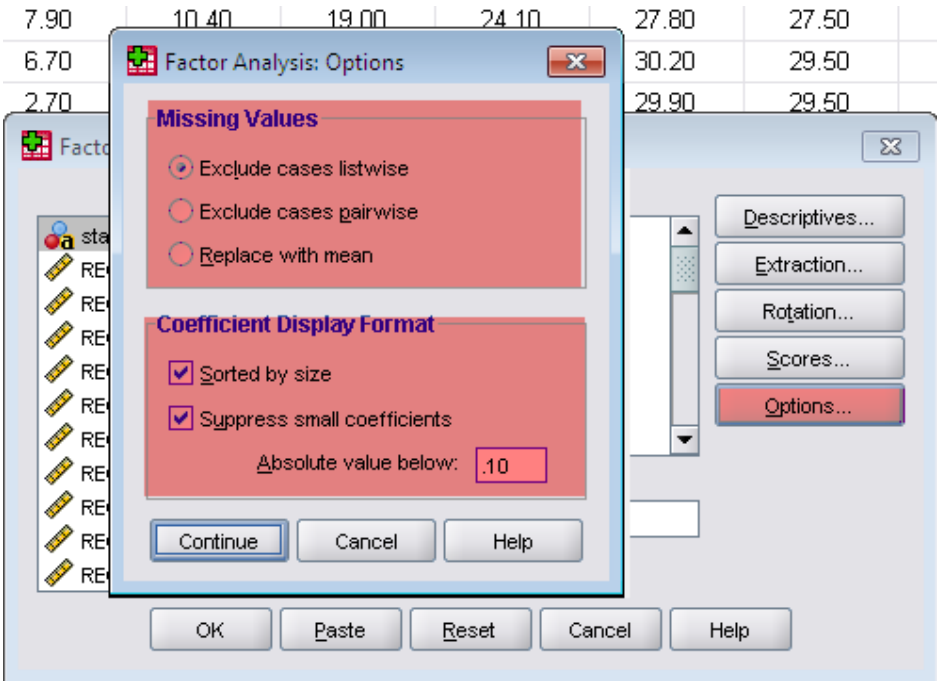
Component Score Covariance Matrix

Component	8	9
1	.000	.000
2	.000	.000

3	.000	.000
4	.000	.000
5	.000	.000
6	.000	.000
7	.000	.000
8	1.000	.000
9	.000	1.000

Extraction Method: Principal Component Analysis.
Component Scores.

در قسمت Options :



تحلیل خوشه ای[1]

گروه بندی تیمارها یا متغیرها در تحلیل های یک متغیره و استنباط آماری از آنها با روش های متداول برای این نوع تحلیل ها انجام می گیرد. مثلا اگر 10 دانشجو را بخواهیم در یک درس به سه طیف خوب متوسط و ضعیف دسته بندی نماییم کافی است که نمرات را بر حسب این سه گروه بندی مشخص می کنیم و در آن صورت بسادگی از روی نمرات می توان این 10 نفر را به گروه های متفاوت نسبت داد . تحلیل خوشه ای به ما کمک می کند که این گروه بندی ها را بنحو مطلوبی انجام دهیم به این ترتیب که m فرد از افراد جامعه لحاظ p صفت به طور توأم به گروه هایی تفکیک می کنیم. اولین قدم در دسته بندی افراد یا اشیاء داشتن معیاری برای شباهت یا عدم شباهت است . در بسیاری موارد تحلیل خوشه ای براساس معیار عدم تشابه صورت می گیرد. هرچه فاصله بین دو فرد کمتر باشد تشابه آنها بیشتر و برعکس هر چه فاصله آنها دورتر باشد عدم تشابه بیشتر است. معیار عدم تشابه با اندازه گیری فاصله بین جفت افراد یا اشیاء صورت می گیرد:

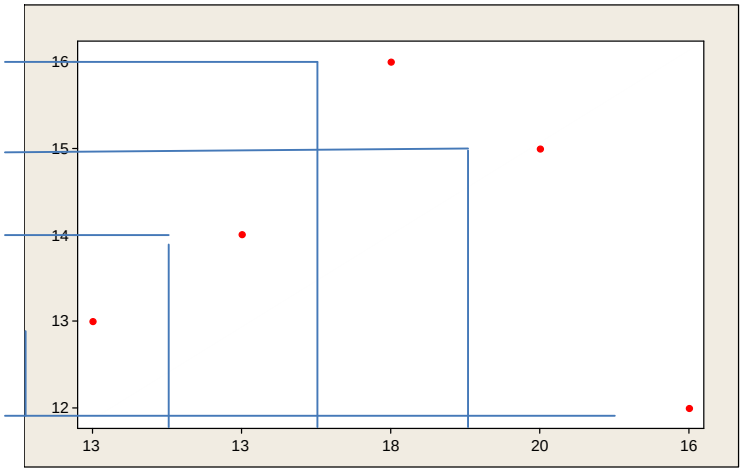
فاصله اقلیدسی :

فرض کنید نمرات 2 درس برای 5 دانشجو به صورت جدول زیر باشد:

دانشجو	زیست شناسی	جانورشناسی
1	13	13
2	13	14
3	18	16
4	20	15
5	16	12

وضعیت هر دو دانشجو در یک فضای دو بعدی بصورت زیر می باشد:

[1] - برای اطلاع بیشتر به کتاب مقدمه ای بر روش های تحقیق در علوم انسانی با تأکید بر جغرافیا ، از مؤلف مراجعه شود.



در شکل بالا فاصله نقاط نسبت به هم با خط مشخص شده است با قضیه اقلیدس می توان فاصله دو نقطه را محاسبه نمود. در یک مثلث قائم الزاویه مربع وتر مساوی مجموع مربع دو ضلع دیگر است .یعنی :

$$c^2 = a^2 + b^2$$

که در آن C وتر ، a و b دو ضلع دیگر هستند . لذا :

$$c = \sqrt{a^2 + b^2} = \left[a^2 + b^2\right]^{\frac{1}{2}}$$

اگر فاصله بین دو فرد را با d_{rs} نشان دهیم فاصله اقلیدسی برای دو صفت به صورت زیر در می آید:

$$d_{rs} = \left[\left(x_{r1} - x_{s1}\right)^2 + \left(x_{r2} - x_{s2}\right)^2\right]^{\frac{1}{2}}$$

فاصله اقلیدسی بین دو دانشجوی سوم و چهارم برابر است با :

$$d_{rs} = \left[\left(18-20\right)^2 + \left(16-15\right)^2\right]^{\frac{1}{2}} = 2.236$$

با استفاده از نرم افزار spss ماتریس عدم تشابه برای دانشجویان را برای دو درس به صورت زیر می توان نوشت:

ماتریس عدم تشابه

Case افراد	Euclidean Distance فاصله اقلیدسی				
	1	2	3	4	5
1	.000	1.000	5.831	7.280	3.162
2	1.000	.000	5.385	7.071	3.606
3	5.831	5.385	.000	2.236	4.472
4	7.280	7.071	2.236	.000	5.000
5	3.162	3.606	4.472	5.000	.000

جدول بالا به صورت مربع و متقارن است و بنابراین به طور معمول در ماتریس عدم تشابه فقط داده های پایین و یا بالای قطر ارائه می شود . یعنی :

ماتریس عدم تشابه

Case افراد	Euclidean Distance فاصله اقلیدسی				
	1	2	3	4	5
1	-	1.000	5.831	7.280	3.162
2		-	5.385	7.071	3.606
3			-	2.236	4.472
4				-	5.000
5					-

در این ماتریس فاصله بین دانشجوی سوم و چهارم $d_{34}=2.236$ می باشد که قبلا محاسبه شد.

فاصله اقلیدسی استاندارد شده :

وقتی متغیرهای اندازه گیری دارای مقیاس متفاوت باشند لازم است داده ها را استاندارد نمود (مقیاس ازبین برود). جهت بدست آوردن فاصله اقلیدسی استاندارد شده[1] می توان از رابطه زیر استفاده کرد:

[1]- Carl Pearson Distance

$$d_{rs} = \left[\left(z_{r1} - z_{s1} \right)^2 + \left(z_{r2} - z_{s2} \right)^2 \right]^{\frac{1}{2}}$$

انواع تحلیل خوشه ای :

روش های متعددی برای تحلیل خوشه ای ارائه می شود که عبارتند از :

- تحلیل خوشه ای سلسله مراتبی [1]
- تحلیل خوشه ای چند میانگین [2]
- تحلیل خوشه ای دو مرحله ای [3]

در تحلیل خوشه ای سلسله مراتبی (شجره ای) ابتدا فاصله تکوینی هر فرد از بقیه افراد محاسبه می شود سپس طبق روش تجمیعی[4] و یا روش تقسیمی یا تجزیه ای[5] گروه ها تشکیل می شود. در روش تجمیعی همه مواد را به طور منفرد در یک دسته قرار داده می شود و سپس دسته های نزدیک به هم را در یکدیگر ادغام کرده تا در نهایت همه گروهها تبدیل به یک دسته شوند. در روش تقسیمی همه مواد در یک گروه قرار می گیرد سپس این گروه به دو زیر گروه و ایندو به زیر گروه های دیگر تقسیم شده تا زمانی که هر ژنوتیپ در گروه خودش قرارگیرد.

تحلیل خوشه ای تجمیعی (ادغام) به روش های پیوند تکی ، پیوند کامل ، پیوند متوسط ، وارد و روش مرکز ثقل تقسیم می شود. روش پیوند تکی بر اساس حداقل فاصله بنا شده است . در پیوند کامل معیار خوشه بندی بر حداکثر فاصله بنا شده است . در پیوند متوسط در خوشه بندی بر میانگین فاصله بین گروهها بنا شده است . در روش وارد براساس مجموع مجذورات بین دو خوشه بدست آمده از کل متغیرها بنا شده است . در روش مرکز ثقل فاصله بین دو خوشه فاصله اقلیدسی بین میانگین های آنها است.

در تحلیل خوشه ای چند میانگینی سعی بر آن دارد تا گروههای همگنی از موارد را براساس خصوصیات انتخاب شده شناسایی کند. این شناسایی بر اساس الگوریتمی انجام می شود که تعداد قابل توجهی از موارد را شامل می شود. داده هایی که در تحلیل خوشه ای چند میانگینی مورد استفاده قرار می گیرند باید از نوع پارامتری و یا کمی باشند. چنانچه داده ها ناپارامتری و یا کیفی هستند باید از تحلیل سلسله مراتبی استفاده کرد.

در تحلیل خوشه ای دو مرحله ای برای داده ها و حجم زیاد داده استفاده می شود . در این نوع انتخاب خودکار مناسب ترین تعداد خوشه ها و همچنین بهترین شاخص برای انتخاب از بین مدل های مختلف خوشه بندی وجود دارد.

---

[1]- Hierarchical Cluster Analysis

[2]- K-means Cluster Analysis

[3]- Two-step Cluster Analysis

[4]- Agglomeration Cluster Analysis

[5]- Partitioning Cluster Analysis

در تحلیل خوشه ای فاصله موضوع بسیار مهمی است که در قالب ماتریس فاصله مطرح می شود. برای سنجش فاصله از شاخص های فاصله اقلیدسی ، نزدیک ترین فاصله ، دورترین فاصله ، روش های جفتی وزن داده شده با استفاده از میانگین ، متوسط پیوند درون گروهی ، وارد ، مرکز ثقل و میانه را می توان اشاره نمود.

پیش فرض های تحلیل خوشه ای :

در تحلیل خوشه ای پیش فرض هایی به شرح زیر باید مدنظر قرار گیرد:

- سطح مقیاس داده باید فاصله ای و نسبی و دووجهی اسمی و برای دو مرحله ای علاوه بر مقیاس فاصله ای و نسبی مقیاس ترتیبی لازم است .
- مشاهدات باید مستقل از همدیگر باشند.
- استاندارد لازم متغیرها
- تمامی پیش فرض های تحلیل های خطی چند گانه لازم است.
- تحلیل خوشه ای چند میانگینی برای اندازه های بزرگتر از 250 داده بکار می رود.
- در تحلیل خوشه ای چند میانگینی نسبت به داده های بیرونی و پرت حساس می باشد که باید حذف بیرونی انجام گیرد.
- در تحلیل خوشه ای چند میانگینی بسته به ترتیب مشاهدات معمولا راه حل های متفاوتی وجود دارد.

کاربرد تحلیل خوشه ای سلسله مراتبی :

روش تحلیل خوشه ای سلسله مراتبی به عنوان پرکاربرد ترین تحلیل خوشه ای برای نمونه های کوچک معمولا کمتر از 250 مناسب است . در این تحلیل برای خوشه بندی هم می توان از داده های خام و هم می توان از داده های استاندارد شده استفاده کرد. لذا در این تحلیل باید به مقیاس داده ها توجه نمود اگر مقیاس داده ها متفاوت باشد یا تقاوت زیادی باهم داشته باشند باید از داده های استاندارد شده استفاده کرد. اولین مرحله در این نوع تحلیل این است که محقق باید نحوه تعریف فاصله ، نحوه ادغام یا تجزیه خوشه ها ، و تعداد خوشه های مورد نیاز را مشخص کند. در این نوع تحلیل تعداد بهینه و مطلوب خوشه ها به هدف تحقیق بستگی دارد. در این تحلیل خوشه ها در همدیگر جای می گیرند. در روش خوشه بندی پیش رو[1] که بعنوان خوشه بندی تجمیعی نیز خوانده می شود خوشه های کوچک با استفاده از یک مقدار تمایز یا برش بالای شاخص تشابه (بیش از 0/9) تشکیل می شوند و در نهایت به یک مقدار شاخص پایین تشابه در یک خوشه قرار گیرند. در روش پس رو[2] که بعنوان خوشه بندی تقسیمی یا تجزیه ای نیز خوانده می شود ماهیت روش قبلی را دارد با این تفاوت که در اینجا نقطه شروع با یک مقدار شاخص تشابه پایین انجام می شود و درنهایت به یک مقدار شاخص تشابه بالا می رسد. این دو روش لزوما نتایج یکسانی ندارند.

پیش فرض های کاربرد تحلیل خوشه ای سلسله مراتبی :

- متغیرهای مستقل باید در سطح سنجش اسمی دو وجهی و فاصله ای باشند.

---

[1]- Forward

[2]- Backward

- اختلاف زیادی نباید بین مقیاس متغیرها وجود داشته باشد(استاندارد سازی برای متغیرها با اختلاف مقیاس زیاد).
- شاخص فاصله یا تشابهی که بکار می رود باید با داده ها همخوانی داشته باشد.

مثال :

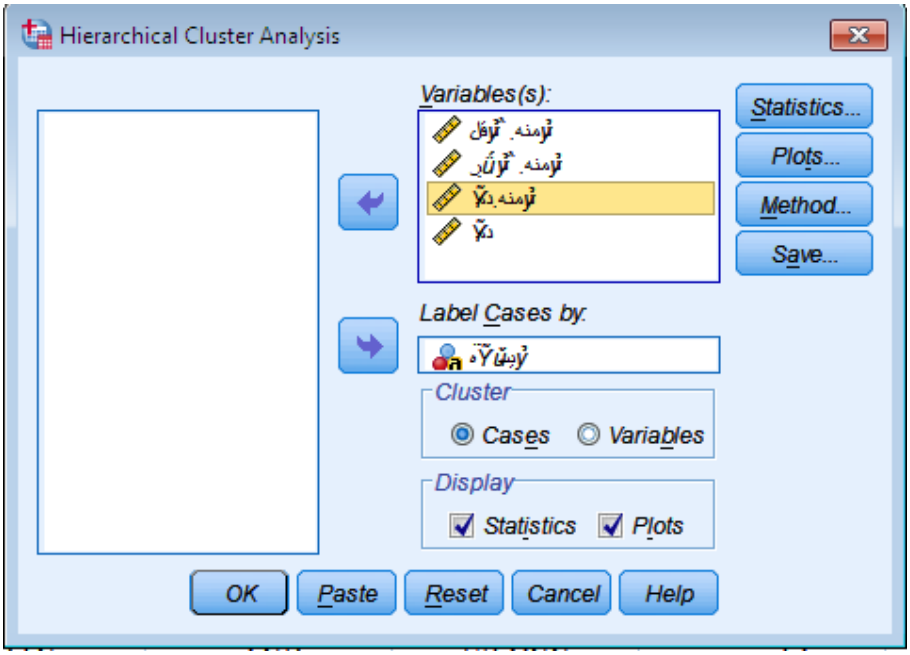
اگر به فرض 26 ایستگاه در دست باشد و هدف توجه بهینه به عناصر اقلیمی (شامل دما ، اختلاف حداکثر دما ، اختلاف حداقل دما و دامنه دما) به صورت جدول زیر باشد . ایستگاه را راستای هدف گروه بندی نمایید:

ردیف	ایستگاه	دما	دامنه حداکثر دما	دامنه حداقل دما	دامنه دما
1	رشت	20.9	0.759	0.646	0.276
2	یزد	16.1	0.778	0.642	0.244
3	اصفهان	15.4	0.789	0.614	0.386
4	تهران	11.3	0.842	0.61	0.441
5	شیراز	20.8	0.785	0.584	0.222
6	سمنان	15	0.751	0.572	0.245
7	ساری	22.1	0.724	0.566	0.245
8	شهرکرد	24.6	0.682	0.564	0.243
9	اراک	21.1	0.750	0.563	0.236
10	کرمان	22.9	0.739	0.562	0.271
11	قم	17.6	0.795	0.555	0.229
12	مشهد	23.1	0.698	0.552	0.338
13	تبریز	25.2	0.719	0.542	0.229
14	کرمانشاه	24.8	0.747	0.535	0.246
15	ایلام	23.7	0.675	0.520	0.226
16	اهواز	23.5	0.746	0.515	0.225

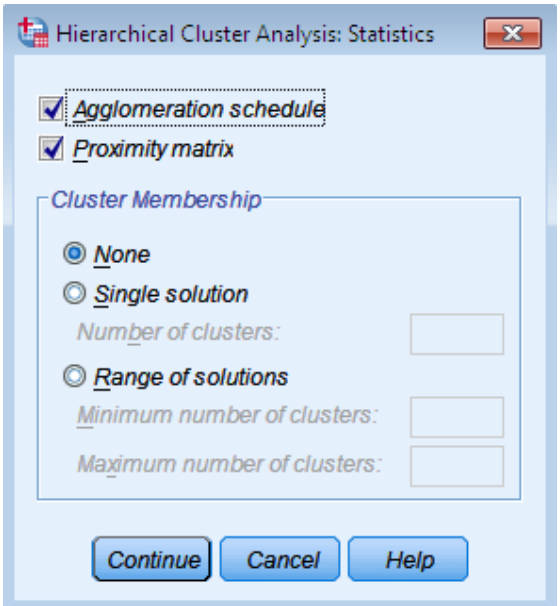
17	اردبیل	29.8	0.705	0.510	0.205
18	بوشهر	21.7	0.706	0.51	0.223
19	خرم آباد	27.6	0.680	0.499	0.207
20	همدان	26.2	0.637	0.492	0.380
21	یاسوج	33.2	0.623	0.491	0.164
22	زنجان	28.5	0.649	0.490	0.251
23	بندرعباس	27.7	0.693	0.486	0.220
24	ارومیه	30.4	0.65	0.474	0.316
25	سنندج	31.7	0.619	0.448	0.225
26	زاهدان	39.5	0.545	0.393	0.220

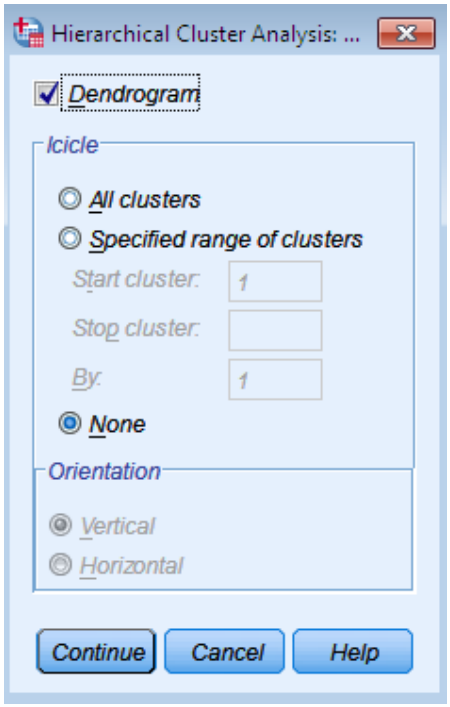
مراحل اجرا با نرم افزار ***SPSS***:

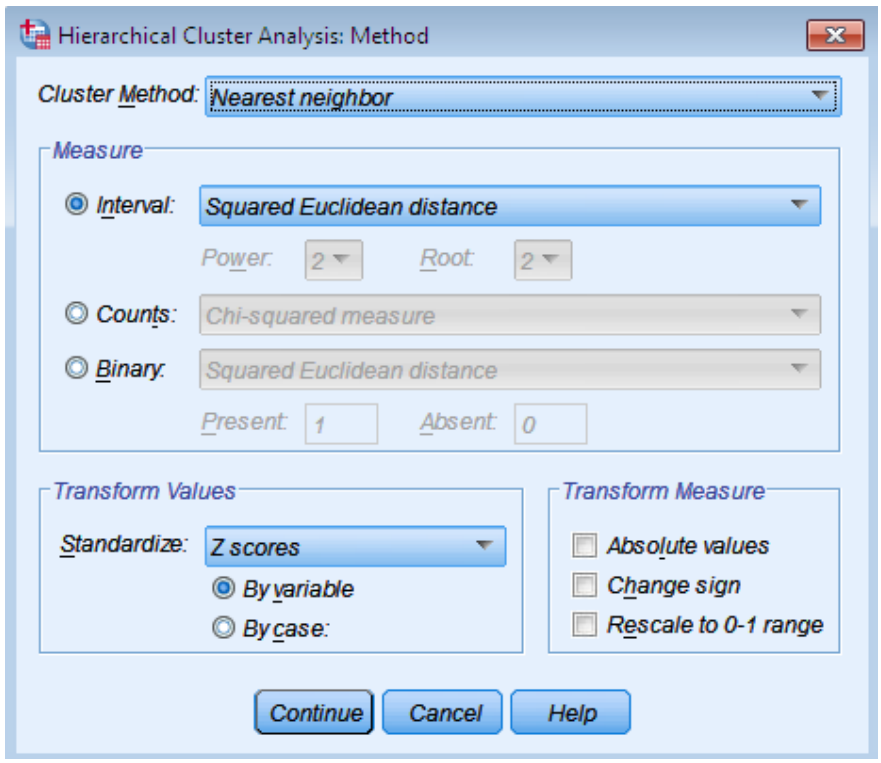
- ابتدا داده ها را در محیط نرم افزار وارد کنید.
- سپس دستور Analyze...> Classify ....> Hierarchical Cluster .
- در پنجره Variablesمتغیرها واردکنید .یعنی :

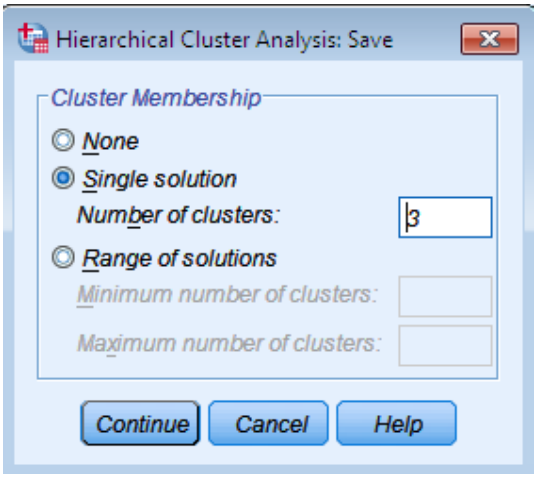


سپس تنظیمات لازم به صورت زیر انجام دهید:




Hierarchical Cluster Analysis: ...
Dendrogram
Icicle
All clusters
Specified range of clusters
Start cluster: 1
Stop cluster:
By: 1
None
Orientation
Vertical
Horizontal
Continue
Cancel
Help


Hierarchical Cluster Analysis: Method
Cluster Method: Nearest neighbor
Measure
Interval: Squared Euclidean distance
Power: 2 Root: 2
Counts: Chi-squared measure
Binary: Squared Euclidean distance
Present: 1 Absent: 0
Transform Values
Standardize: Z scores
By variable
By case:
Transform Measure
Absolute values
Change sign
Rescale to 0-1 range
Continue
Cancel
Help



درجدول زیر مشخص می کند هر ایستگاه در کدام خوشه قرار دارد:

CLU3_1
1
1
2
2
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1

در جدول زیر ماتریس تقریب ها (Proximities) بین متغیرها را نشان می دهد. مقادیر این جدول نشانگر میزان تشابه ها یا تفاوت های هر جفت متغیر می باشد. در این مثال از روش مجذور فاصله اقلیدسی استفاده شد که یک شاخص برای سنجش میزان فاصله بین جفت ها می باشد. دراین روش هر چه مقدار فاصله بیشتر باشد نشانگر آن است که اختلاف بین متغیرها بیشتر است . همانطوری که ملاحظه می شود بین ایستگاه رشت و زاهدان بیش ترین فاصله (33/98) و کمترین فاصله با ایستگاه یزد (با مقدار 0/93) وجود دارد. بنابراین فاصله هر ایستگاه با خودش صفر است.

Proximity Matrix

Case	Squared Euclidean Distance						
	1: 1	2: 2	3: 3	4: 4	5: 5	6: 6	7: 7
1: 1	.000	.933	4.275	11.056	1.990	2.735	2.409
2: 2	.933	.000	5.259	11.430	1.674	1.617	3.265
3: 3	4.275	5.259	.000	1.826	7.673	5.766	7.707
4: 4	11.056	11.430	1.826	.000	15.130	12.168	16.244
5: 5	1.990	1.674	7.673	15.130	.000	1.301	1.116
6: 6	2.735	1.617	5.766	12.168	1.301	.000	1.473
7: 7	2.409	3.265	7.707	16.244	1.116	1.473	.000
8: 8	3.922	5.722	10.573	20.702	3.014	3.475	.565
9: 9	2.412	2.646	7.506	15.443	.458	1.001	.202
10: 10	2.245	3.571	6.074	13.701	1.330	1.834	.240
11: 11	3.522	2.373	7.243	13.527	.542	.762	1.768
12: 12	4.485	7.254	5.097	11.908	5.489	4.585	2.378
13: 13	4.521	5.877	11.191	20.883	2.014	3.233	.483
14: 14	4.217	5.484	9.335	17.787	1.579	2.871	.587
15: 15	7.029	8.297	13.632	24.101	4.165	4.136	1.316
16: 16	5.829	6.405	11.366	20.094	1.916	2.904	1.013
17: 17	9.309	11.467	18.185	29.765	5.200	7.626	2.912

18:	18	6.712	7.146	12.302	21.660	3.030	2.850	1.107
19:	19	10.017	11.855	18.300	29.940	5.851	7.134	2.878
20:	20	13.665	18.254	12.588	20.244	14.375	12.554	8.256
21:	21	18.174	21.188	31.014	46.364	13.275	15.779	8.748
22:	22	11.450	14.455	17.848	29.208	8.509	9.016	4.018
23:	23	10.377	12.302	17.561	28.524	5.938	7.213	3.035
24:	24	13.998	18.456	17.083	27.012	12.218	12.455	6.720
25:	25	19.472	23.016	27.822	41.192	14.698	15.698	9.017
26:	26	38.660	44.565	49.482	66.293	32.694	34.549	23.917

ادامه جدول:

Proximity Matrix

	Squared Euclidean Distance						
Case	8: 8	9: 9	10: 10	11: 11	12: 12	13: 13	14: 14
1: 1	3.922	2.412	2.245	3.522	4.485	4.521	4.217
2: 2	5.722	2.646	3.571	2.373	7.254	5.877	5.484
3: 3	10.573	7.506	6.074	7.243	5.097	11.191	9.335
4: 4	20.702	15.443	13.701	13.527	11.908	20.883	17.787
5: 5	3.014	.458	1.330	.542	5.489	2.014	1.579
6: 6	3.475	1.001	1.834	.762	4.585	3.233	2.871
7: 7	.565	.202	.240	1.768	2.378	.483	.587
8: 8	.000	1.381	1.010	4.243	2.392	.510	1.210
9: 9	1.381	.000	.414	.807	3.330	.791	.606
10: 10	1.010	.414	.000	1.888	1.524	.780	.473
11: 11	4.243	.807	1.888	.000	5.866	2.851	2.046

12:	12	2.392	3.330	1.524	5.866	.000	3.184	2.801
13:	13	.510	.791	.780	2.851	3.184	.000	.269
14:	14	1.210	.606	.473	2.046	2.801	.269	.000
15:	15	.665	2.016	1.963	4.596	3.532	.642	1.377
16:	16	1.741	.849	1.184	1.910	4.087	.456	.269
17:	17	2.018	3.460	3.350	6.403	6.056	1.028	1.642
18:	18	1.292	1.306	1.639	2.833	3.850	.659	.942
19:	19	1.778	3.599	3.526	6.615	5.657	1.150	1.977
20:	20	6.676	10.173	7.013	14.387	2.575	7.927	7.789
21:	21	5.785	10.227	10.090	15.234	12.479	5.546	7.545
22:	22	2.242	5.333	4.254	9.258	4.285	2.300	3.134
23:	23	2.169	3.642	3.393	6.394	5.259	1.244	1.744
24:	24	4.765	8.383	5.997	12.781	3.778	4.994	5.242
25:	25	6.181	10.663	9.564	15.495	9.620	5.930	7.261
26:	26	18.592	26.724	24.585	34.205	23.025	18.613	20.868

ادامه جدول :

Proximity Matrix

	Squared Euclidean Distance					
Case	15: 15	16: 16	17: 17	18: 18	19: 19	20: 20
1: 1	7.029	5.829	9.309	6.712	10.017	13.665
2: 2	8.297	6.405	11.467	7.146	11.855	18.254
3: 3	13.632	11.366	18.185	12.302	18.300	12.588
4: 4	24.101	20.094	29.765	21.660	29.940	20.244

5:	5	4.165	1.916	5.200	3.030	5.851	14.375
6:	6	4.136	2.904	7.626	2.850	7.134	12.554
7:	7	1.316	1.013	2.912	1.107	2.878	8.256
8:	8	.665	1.741	2.018	1.292	1.778	6.676
9:	9	2.016	.849	3.460	1.306	3.599	10.173
10:	10	1.963	1.184	3.350	1.639	3.526	7.013
11:	11	4.596	1.910	6.403	2.833	6.615	14.387
12:	12	3.532	4.087	6.056	3.850	5.657	2.575
13:	13	.642	.456	1.028	.659	1.150	7.927
14:	14	1.377	.269	1.642	.942	1.977	7.789
15:	15	.000	1.157	1.300	.353	.614	6.588
16:	16	1.157	.000	1.510	.456	1.580	8.996
17:	17	1.300	1.510	.000	1.768	.303	9.062
18:	18	.353	.456	1.768	.000	1.148	7.802
19:	19	.614	1.580	.303	1.148	.000	7.895
20:	20	6.588	8.996	9.062	7.802	7.895	.000
21:	21	4.133	6.956	2.351	5.938	2.023	12.855
22:	22	1.162	3.136	1.398	2.240	.743	4.290
23:	23	.829	1.344	.369	1.133	.130	7.120
24:	24	3.915	5.862	4.124	5.178	3.529	1.600
25:	25	3.863	6.706	2.992	5.412	2.114	7.360
26:	26	14.955	20.111	12.277	18.025	11.094	15.656

ادامه جدول :

Proximity Matrix

Case	Squared Euclidean Distance					
	21: 21	22: 22	23: 23	24: 24	25: 25	26: 26
1: 1	18.174	11.450	10.377	13.998	19.472	38.660
2: 2	21.188	14.455	12.302	18.456	23.016	44.565
3: 3	31.014	17.848	17.561	17.083	27.822	49.482
4: 4	46.364	29.208	28.524	27.012	41.192	66.293
5: 5	13.275	8.509	5.938	12.218	14.698	32.694
6: 6	15.779	9.016	7.213	12.455	15.698	34.549
7: 7	8.748	4.018	3.035	6.720	9.017	23.917
8: 8	5.785	2.242	2.169	4.765	6.181	18.592
9: 9	10.227	5.333	3.642	8.383	10.663	26.724
10: 10	10.090	4.254	3.393	5.997	9.564	24.585
11: 11	15.234	9.258	6.394	12.781	15.495	34.205
12: 12	12.479	4.285	5.259	3.778	9.620	23.025
13: 13	5.546	2.300	1.244	4.994	5.930	18.613
14: 14	7.545	3.134	1.744	5.242	7.261	20.868
15: 15	4.133	1.162	.829	3.915	3.863	14.955
16: 16	6.956	3.136	1.344	5.862	6.706	20.111
17: 17	2.351	1.398	.369	4.124	2.992	12.277
18: 18	5.938	2.240	1.133	5.178	5.412	18.025
19: 19	2.023	.743	.130	3.529	2.114	11.094
20: 20	12.855	4.290	7.120	1.600	7.360	15.656
21: 21	.000	2.596	2.678	6.171	1.518	5.967
22: 22	2.596	.000	.700	1.213	1.147	8.542

23:	23	2.678	.700	.000	2.932	2.084	11.080
24:	24	6.171	1.213	2.932	.000	2.508	8.825
25:	25	1.518	1.147	2.084	2.508	.000	3.696
26:	26	5.967	8.542	11.080	8.825	3.696	.000

جدول زیر به عنوان برنامه تجمیع یا مجموعه سازی (***Agglomeration Schedule***)معروف است . در این جدول نشان می دهد چگونه ایستگاه ها در همدیگر ادغام (***Cluster Combined***) و مرحله تشکیل خوشه برای اولین بار (***Stage Cluster First Appears***) انجام می گیرد و همچنین ستون ضرایب (***Coefficients***) نشان داده می شود. در هر مرحله یک ایستگاه یا خوشه با یک ایستگاه و یا خوشه دیگر ادغام می شود. ایستگاههایی که تنها وجود دارند در جدول با نمره صفر مشخص شده اند و به عنوان مثال در مرحله اول که ایستگاه های 19 و 23 در هم ادغام شده اند دوباره در مرحله 6 و ایستگاه های 19 با 17 ترکیب شده و خوشه جدیدی را در مرحله 13 تشکیل داده اند. ستون ضرایب میزان فاصله بین دو ایستگاه یا خوشه را که در همدیگر ادغام شده اند را نشان می دهد . میزان ضرایبی که در اینجا مشاهده می شود ، بستگی به شاخص تقریب و روش پیوندی (Linkage) دارد که در این تحلیل استفاده شد. توجه داشته باشید زمانی تحلیل خوشه ای خوب است که که بتوان تغییرات ناگهانی و قابل توجهی را در ضرایب مشاهده نماییم . لذا قبل از مرحلۀ 17 بهترین مرحله برای توقف خوشه بندی وجود دارد. لذا تعداد 9 خوشه از مرحله 16 تا 26 باقی می ماند که می توانیم ازآنها برای خوشه بندی ایستگاه ها استفاده نمود.

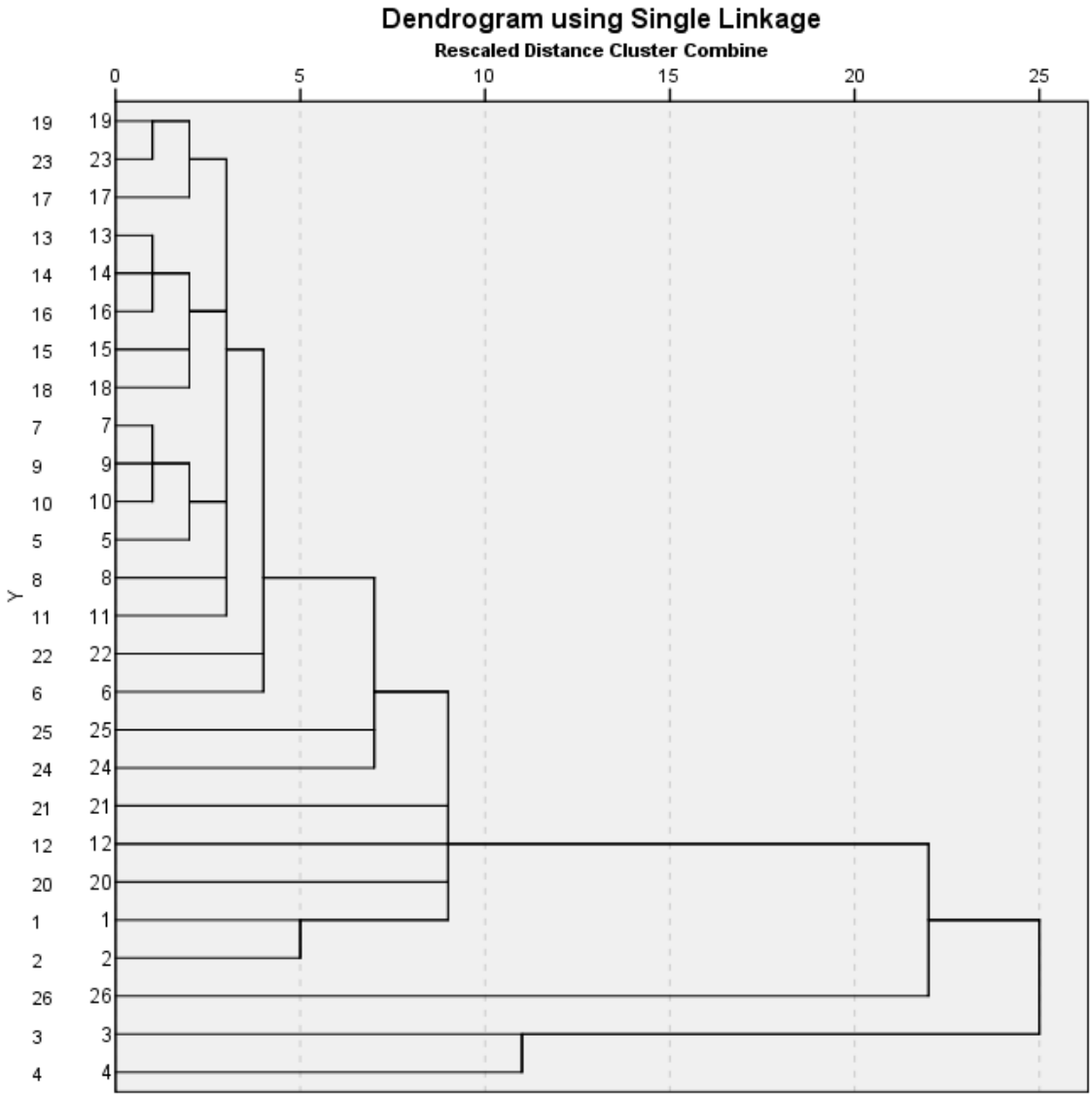
Agglomeration Schedule

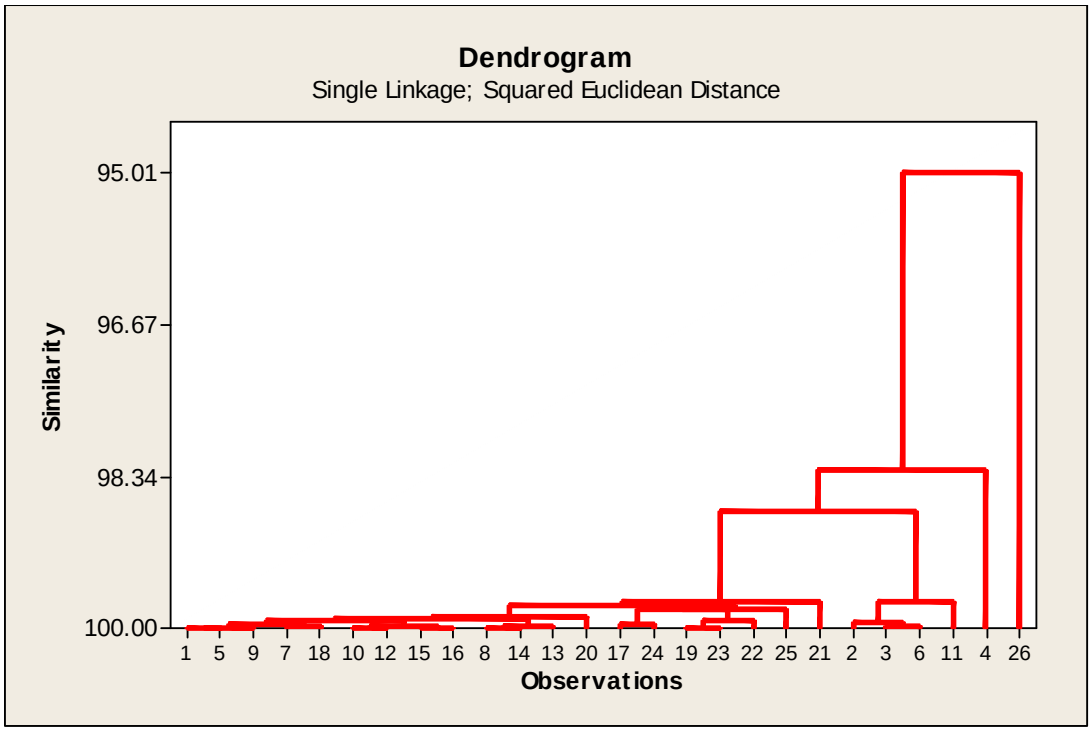
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	19	23	.130	0	0	6
2	7	9	.202	0	0	3
3	7	10	.240	2	0	9
4	13	14	.269	0	0	5
5	13	16	.269	4	0	8
6	17	19	.303	0	1	13
7	15	18	.353	0	0	8
8	13	15	.456	5	7	10
9	5	7	.458	0	3	10

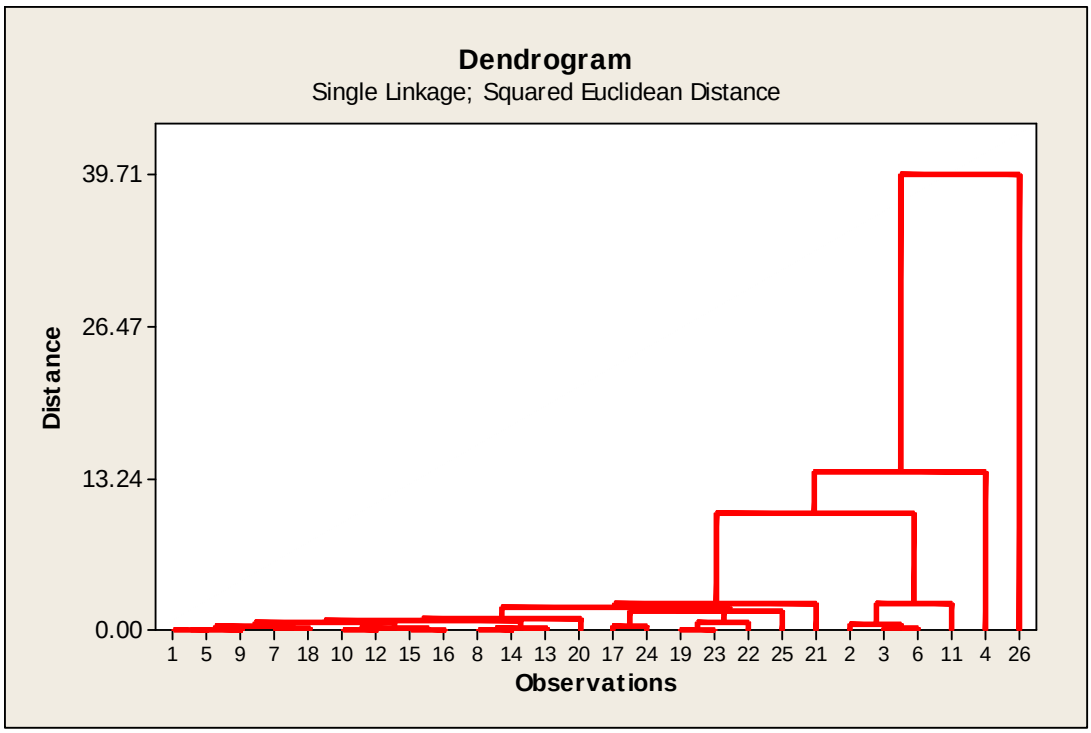
10	5	13	.473	9	8	11
11	5	8	.510	10	0	12
12	5	11	.542	11	0	13
13	5	17	.614	12	6	14
14	5	22	.700	13	0	15
15	5	6	.762	14	0	17
16	1	2	.933	0	0	22
17	5	25	1.147	15	0	18
18	5	24	1.213	17	0	19
19	5	21	1.518	18	0	20
20	5	12	1.524	19	0	21
21	5	20	1.600	20	0	22
22	1	5	1.617	16	21	24
23	3	4	1.826	0	0	25
24	1	26	3.696	22	0	25
25	1	3	4.275	24	23	0

نمودار درختی[1] (دندروگرام) شباهت های نسبی بین متغیرها را نشان می دهد . این نمودار مقدار متغیرهای هر گروه بیش ترین شباهت را باهم دارند و کم ترین شباهت را با گروه های دیگر نشان می دهد. نمودار درختی نموداری از مشاهدات ترکیبی و مقادیر ضرایب خوشه بندی در هرخوشه است. همانطوری که در نمودار مشاهده می شود فواصل بین خوشه ها بین صفر تا 25 مقیاس بندی شده است. در این نمودار متغیرها در محور عمودی سمت چپ و فاصله بین خوشه ها در محور افقی را پس از ادغام نشان می دهد. روش تفسیر نمودار درختی به این صورت است که باید از فاصله بین ادغام خوشه ها در همدیگر برای بیان تعداد خوشه ها استفاده کرد. لذا در نمودار درختی مثال در نهایت دو خوشه بزرگ از مجموع خوشه های کوچک دارا می باشد. در فاصله 22 تا 25 ، دو خوشه تشکیل شده است و در فاصله 12 تا 22 ، سه خوشه ، در فاصله 10 تا 12 ، چهارخوشه ، در فاصله 7 تا 10، پنج خوشه ، در فاصله 5 تا 7 ، هفت خوشه ، در فاصله 3 تا 4 ، هفده خوشه ، در فاصله 1 تا 2 ، بیست و یک خوشه و در فاصله 0 تا 1 ، بیست و پنج خوشه تشکیل شده است :

[1] Dendrogram


Dendrogram using Single Linkage
Rescaled Distance Cluster Combine
0
5
10
15
20
25
Y
19 19
23 23
17 17
13 13
14 14
16 16
15 15
18 18
7 7
9 9
10 10
5 5
8 8
11 11
22 22
6 6
25 25
24 24
21 21
12 12
20 20
1 1
2 2
26 26
3 3
4 4


Dendrogram
Single Linkage; Squared Euclidean Distance
Similarity
95.01
96.67
98.34
100.00
1 5 9 7 18 10 12 15 16 8 14 13 20 17 24 19 23 22 25 21 2 3 6 11 4 26
Observations


Dendrogram
Single Linkage; Squared Euclidean Distance
Distance
39.71
26.47
13.24
0.00
1 5 9 7 18 10 12 15 16 8 14 13 20 17 24 19 23 22 25 21 2 3 6 11 4 26
Observations

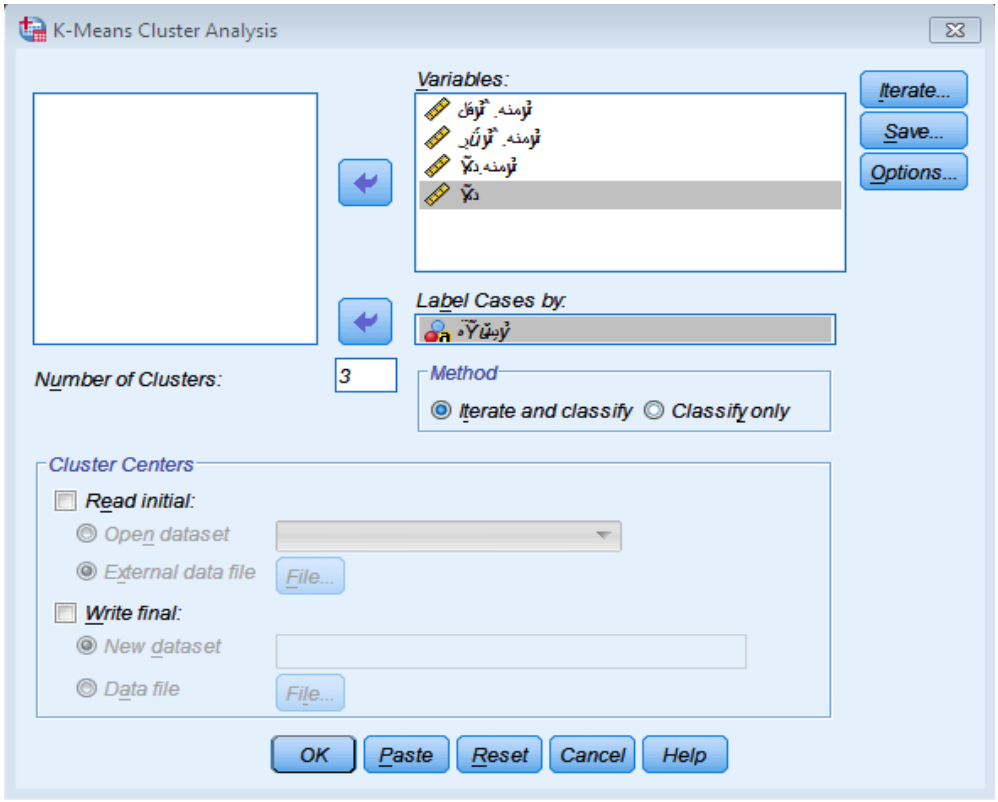
تحلیل خوشه ای چند میانگینی :

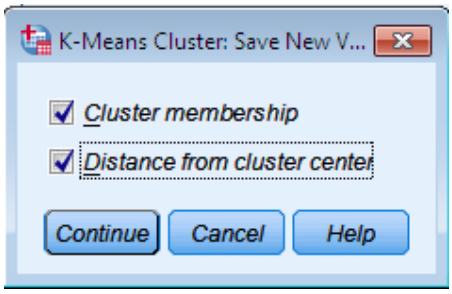
تحلیل خوشه ای چند میانگینی برای داده های بسیار بزرگ به کار می رود . در این روش فرض را بر این می گذارد که داده ها به یک تعداد خوشه مشخص و شناخته شده طبقه بندی می شوند . لذا با چنین فرضی (مشخص بودن خوشه ها) متغیرها را به چندین خوشه طبقه بندی می کند. برای طبقه بندی می توان دو روش اعمال کرد:

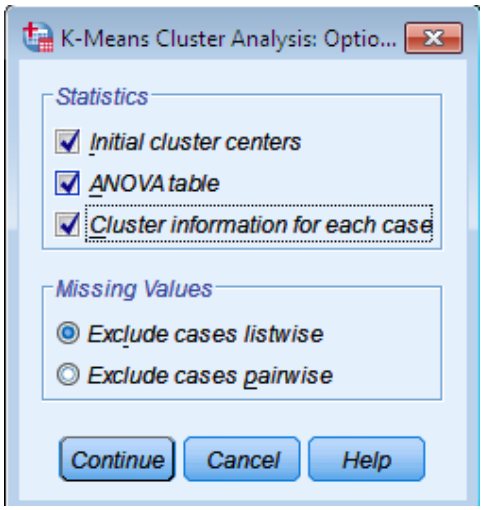
1. به روز کردن مراکز خوشه (Cluster Center).
2. از طریق طبقه بندی مطلق

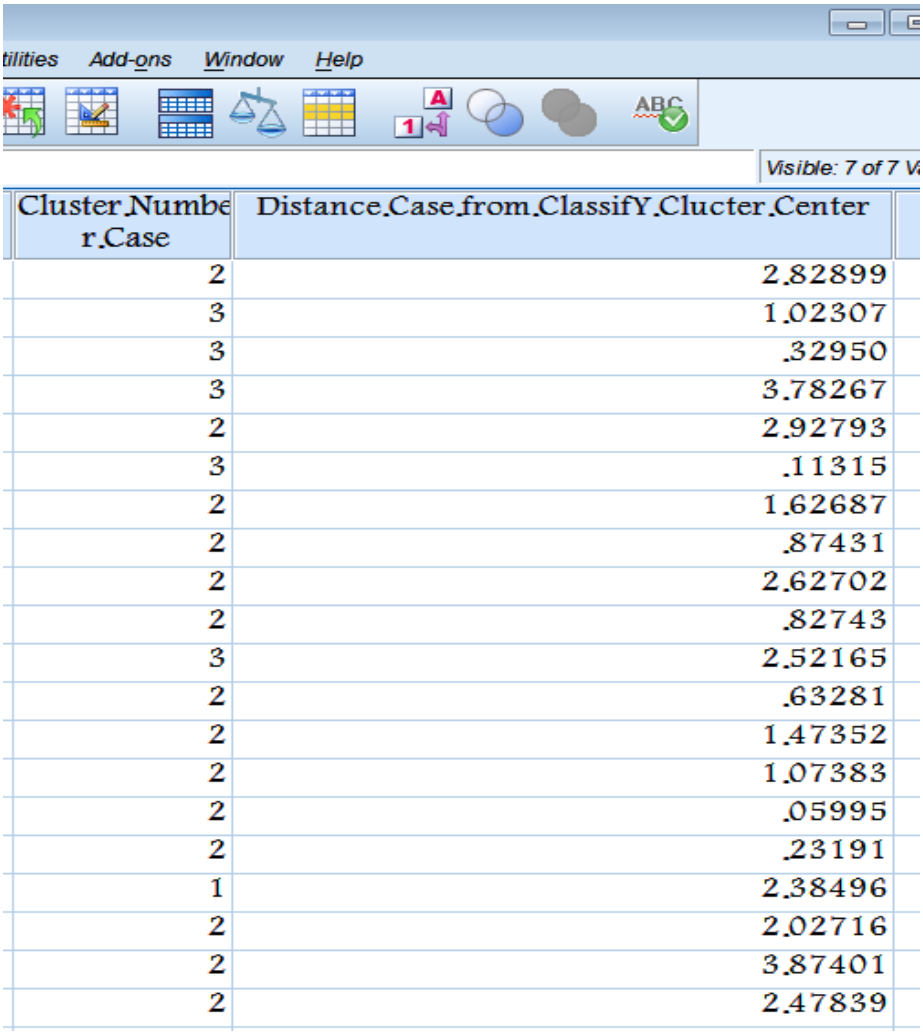
روش تحلیل خوشه ای چند میانگینی برای تعیین فاصله بین دو متغیر از روش فاصله اقلیدسی استفاده می کند. در این روش مراکز اولیه داده در گذر اول بر روی دادها انتخاب می شوند. سپس هر کدام از مشاهدات تکراری گروه ها بر اساس روش نزدیک ترین فاصله اقلیدسی به میانگین خوشه , به آن اضافه می شوند . در این حین محقق باید تعداد مطلوب خوشه خود را مشخص کند. این فرآیند همینطور ادامه می یابد تا جایی که میانگین خوشه ها بیش از یک مقدار مشخص تغییر نکند . با توجه به مثال قبلی مراحل زیر را می توان در نظر گرفت:

1. ابتدا دستور Analyze> Classify> K-Means Cluster فعال کنید.
2. سپس مراحل زیر را انجام دهید:




K-Means Cluster: Save New V...
Cluster membership
Distance from cluster center
Continue
Cancel
Help


K-Means Cluster Analysis: Optio...
Statistics
Initial cluster centers
ANOVA table
Cluster information for each case
Missing Values
Exclude cases listwise
Exclude cases pairwise
Continue
Cancel
Help



Cluster Number Case	Distance Case from ClassifY Clucter Center
2	2.82899
3	1.02307
3	.32950
3	3.78267
2	2.92793
3	.11315
2	1.62687
2	.87431
2	2.62702
2	.82743
3	2.52165
2	.63281
2	1.47352
2	1.07383
2	.05995
2	.23191
1	2.38496
2	2.02716
2	3.87401
2	2.47839

جدول زیر با عنوان مراکز خوشه ها اولیه هر شاخص را در هر خوشه نشان می دهد. همانطوری که مشخص است مقدار میانگین شاخص ها در خوشه اول بیشتر از خوشه دوم و سوم می باشد.

Initial Cluster Centers

	Cluster		
	1	2	3
دامنه.حداقل	.393	.510	.610
دامنه.حداکثر	.545	.706	.842
دامنه.دما	.220	.223	.441
دما	39.500	21.700	11.300

جدول زیر پیشرفت فرآیند خوشه بندی در هر مرحله را نشان می دهد . در تکرار های اولیه مراکز خوشه به میزان نسبتا کمی تغییر می کنند.

Iteration History[a]

Iteration	Change in Cluster Centers		
	1	2	3
1	4.701	2.627	3.152
2	1.881	.301	.630
3	.737	.298	.000
4	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 4. The minimum distance between initial centers is 10.404.

در جدول زیر هر ایستگاه در کدام خوشه قرار دارد. بعنوان مثال ایستگاه رشت درخوشه دوم و ایستگاه یزد در خوشه سوم قرار دارد. ستون چهارم هم فاصله هر ایستگاه را تا مرکز خوشه نشان می دهد. در این جدول مقدار فاصله هر ایستگاه با خودش کم و با خوشه های دیگر زیاد است . از طرفی هر چه مقدار فاصله ایستگاه ها از مرکز خوشه بیشتر باشد نشان می دهد که آن خوشه نماینده خوبی برای خوشه خودش نمی باشد. برای مثال ایستگاه زاهدان از مرکز خوشه اول برابر با 7/318 فاصله دارد.

Cluster Membership

Case Number	ایستگاه	Cluster	Distance
1	1	2	2.829
2	2	3	1.023
3	3	3	.329
4	4	3	3.783
5	5	2	2.928
6	6	3	.113
7	7	2	1.627

8	8	2	.874
9	9	2	2.627
10	10	2	.827
11	11	3	2.522
12	12	2	.633
13	13	2	1.474
14	14	2	1.074
15	15	2	.060
16	16	2	.232
17	17	1	2.385
18	18	2	2.027
19	19	2	3.874
20	20	2	2.478
21	21	1	1.019
22	22	1	3.683
23	23	2	3.974
24	24	1	1.786
25	25	1	.484
26	26	1	7.318

جدول زیر نشان می دهد که مقادیر میانگین متغیرها در خوشه اول بیشتر از خوشه دوم و سوم است به عبارتی در خوشه اول وضعیت بهینه بودن ایستگاه بهتر از خوشه های دیگر است.

Final Cluster Centers

	Cluster		
	1	2	3
دامنه.حداقل	.468	.542	.599
دامنه.حداکثر	.632	.716	.791
دامنه.دما	.230	.252	.309
دما	32.183	23.727	15.080

جدول زیر فواصل اقلیدسی بین مراکز خوشه نهایی را نشان می دهد هر چه میزان فاصله بین این مراکز بیشتر باشد نشانگر این است عدم تشابه بین آنها بیشتر است و برعکس فاصله کم نشان دهنده تشابه بیشتر بین مراکز خوشه ها با همدیگر است. برای مثال فاصله بین خوشه اول و دوم 8/457 است.

Distances between Final Cluster Centers

Cluster	1	2	3
1		8.457	17.105
2	8.457		8.647
3	17.105	8.647	

جدول زیر تحلیل واریانس متغیرها را نشان می دهد. این جدول میزان نقش بیشتر هر کدام از متغیرها را نشان می دهد . هر مقدار F برای هر متغیر بیشتر باشد آن متغیر نقش بیشتری را در جداسازی خوشه ها از همدیگر دارد. برای مثال شاخص دما (53/743) در سطح خطای کوچکتر از 0/001 بیشترین نقش را در جداسازی خوشه ها دارد و برعکس دامنه دما (2/513) در سطح خطای 0/103 کمترین نقش را در جداسازی خوشه ها از همدیگر دارد.

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
دامنه.حداقل	.024	2	.002	23	14.577	.000
دامنه.حداکثر	.035	2	.002	23	20.268	.000

دامنه.دما	.009	2	.004	23	2.513	.103
دما	400.374	2	7.450	23	53.743	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

جدول زیر نشان می دهد چند ایستگاه در هر خوشه قرار گرفته است. همانطوری که مشخص است از مجموع 26 ایستگاه 6 ایستگاه در خوشه اول , 15 ایستگاه در خوشه دو و 5 ایستگاه در خوشه سوم قرار گرفته اند.

Number of Cases in each Cluster

Cluster	1	6.000
	2	15.000
	3	5.000
Valid		26.000
Missing		.000

تحلیل خوشه ای دو مرحله ای :

تحلیل خوشه ای دو مرحله ای با استفاده از الگوریتم تحلیل خوشه ای مقیاس پذیر [1] با مجموعه ای از داده های بسیار بزرگ سرو کار دارد. در تحلیل خوشه ای دو مرحله ای هم برای متغیرهای پیوسته (فاصله ای و نسبی) و هم برای متغیرهای طبقه بندی شده (اسمی و ترتیبی) به کار می رود. در این روش در گام اول که به مرحله پیش خوشه ای [2] معروف است مشاهدات را به چندین خرده خوشه [3] بسیار کوچک تبدیل می شود .سپس این خرده خوشه ها را به تعداد خوشه مطلوب طبقه بندی می نمایند. یکی از مزیت های این روش این است که تعداد خوشه های مطلوب برای محقق ناشناخته است. این روش به طور خودکار تعداد مناسب خوشه ها را برای ما مشخص می کند. با توجه به مثال قبلی مراحل زیر انجام می گیرد:

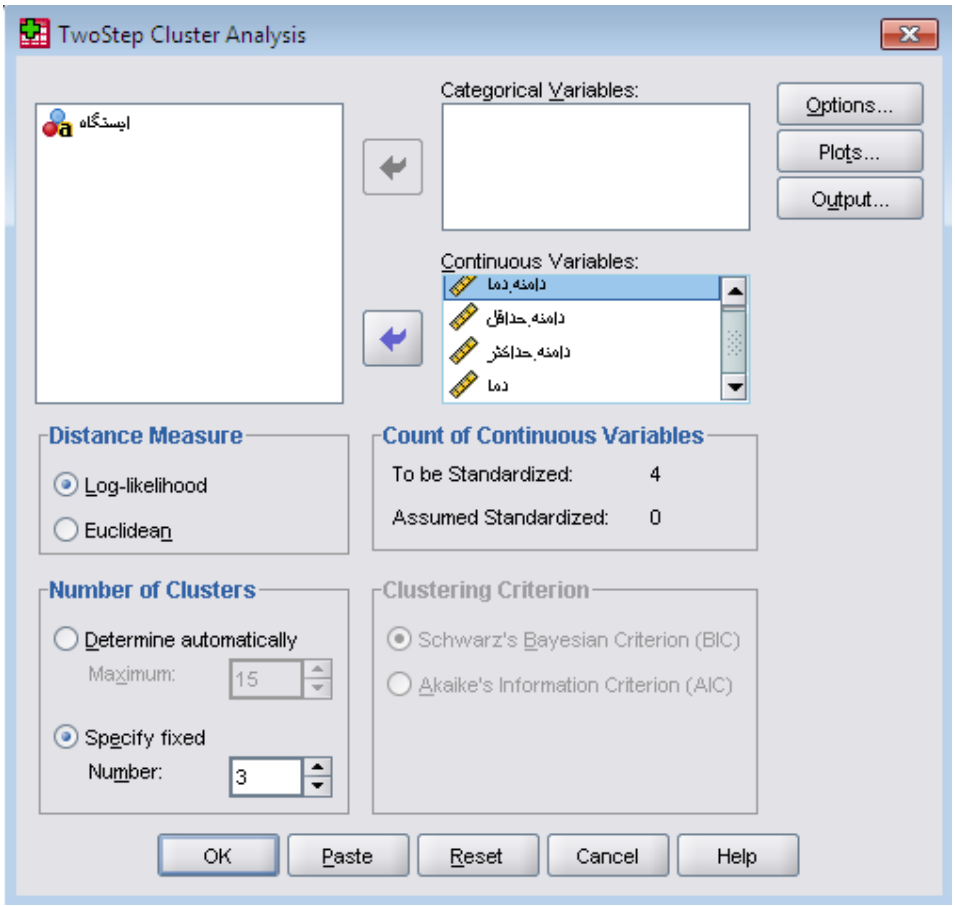
---

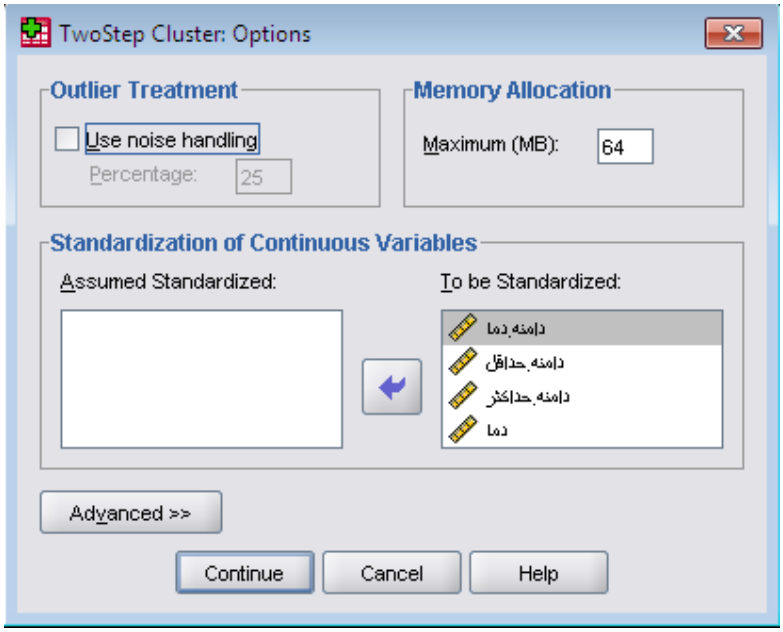
[1]- Scalable

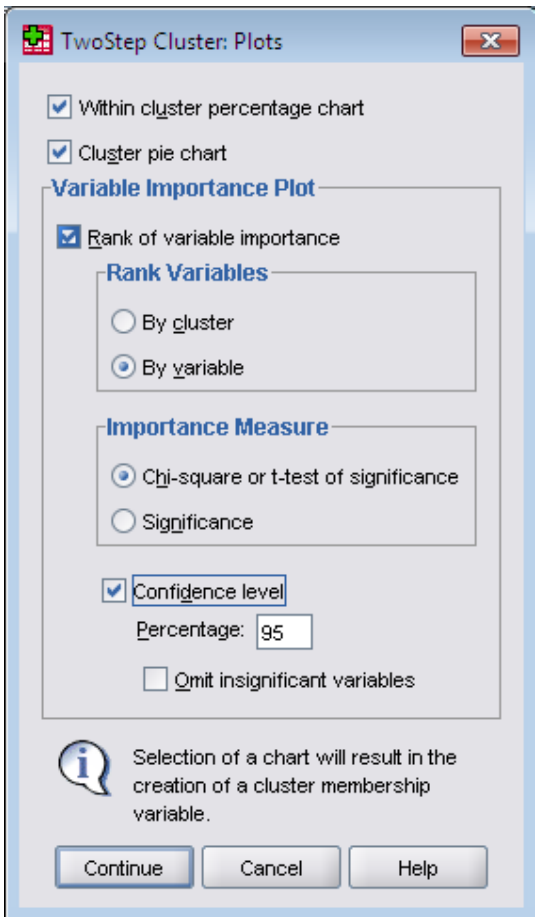
[2]- Pre-Cluster

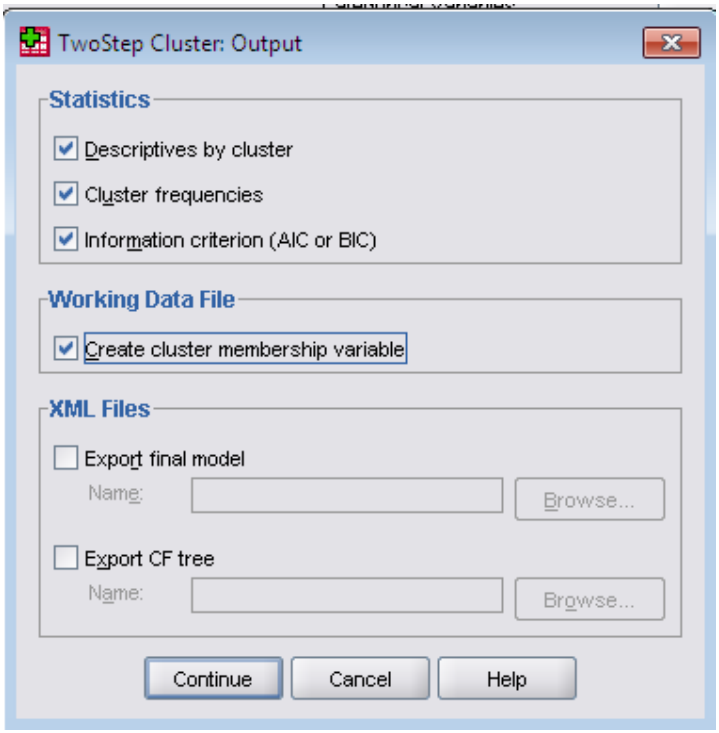
[3]- Sub-Cluster

- ابتدا دستور Analyze> Classify>TwoStep Cluster را فعال کنید.
- مراحل زیر را به ترتیب انجام دهید:




TwoStep Cluster: Options
Outlier Treatment
Use noise handling
Percentage: 25
Memory Allocation
Maximum (MB): 64
Standardization of Continuous Variables
Assumed Standardized:
To be Standardized:
دامنه_دما
دامنه_حداقل
دامنه_حداکثر
دما
Advanced >>
Continue
Cancel
Help


TwoStep Cluster: Plots
Within cluster percentage chart
Cluster pie chart
Variable Importance Plot
Rank of variable importance
Rank Variables
By cluster
By variable
Importance Measure
Chi-square or t-test of significance
Significance
Confidence level
Percentage: 95
Omit insignificant variables
Selection of a chart will result in the creation of a cluster membership variable.
Continue
Cancel
Help



با تکمیل عملیات خروجی جدول زیر در قسمت صفحه گستر نرم افزار با عنوان TSC_... اضافه می شود. که نشان دهنده هر کدام از ایستگاه در هر خوشه می باشد.

TSC_3669
2
2
1
1
2
2
2
2
2
2
2
2
2
2
2
2
3
2
3
3
3
3
3
3

جدول زیر به عنوان جدول توزیع خوشه ها فراوانی هر خوشه را نشان می دهد. براساس جدول تعداد ایستگاه ها در خوشه اول 2 عدد ، 15 ایستگاه در خوشه دوم و 9 ایستگاه در خوشه سوم قرار گرفته است . تمام ایستگاه ها در خوشه ها ادغام شده است . بنابراین هیچ ایستگاهی از

تحلیل خارج نمانده است . از طرفی حدود 7/7 درصد از ایستگاه ها در خوشه اول ، 57/7 درصد در خوشه دوم و 34/6 درصد در خوشه سوم قرار گرفته اند.

TwoStep Cluster

Cluster Distribution

		N	*% of Combined*	*% of Total*
Cluster	*1*	*2*	*7.7%*	*7.7%*
	2	*15*	*57.7%*	*57.7%*
	3	*9*	*34.6%*	*34.6%*
	Combined	*26*	*100.0%*	*100.0%*
Total		*26*		*100.0%*

جدول زیر جدول مرکز ثقل تحلیل را نشان می دهد. جدول زیر میزان میانگین و انحراف معیار خوشه ها را نشان می دهد . برای مثال هر کدام شاخص ها میزان اهمیت آنها را نشان می دهد . میزان نقش خود دما با خوشه های مختلف را نشان می دهد . میزان بالای میانگین هر شاخص اهمیت زیادی در تحلیل را نشان می دهد. برای مثال میانگین شاخص ها درخوشه سوم اهمیت بیشتر آن خوشه را نسبت به خوشه های دیگر بوده است . برای مثال میانگین شاخص دما در خوشه سوم (30/5) را نشان می دهد.

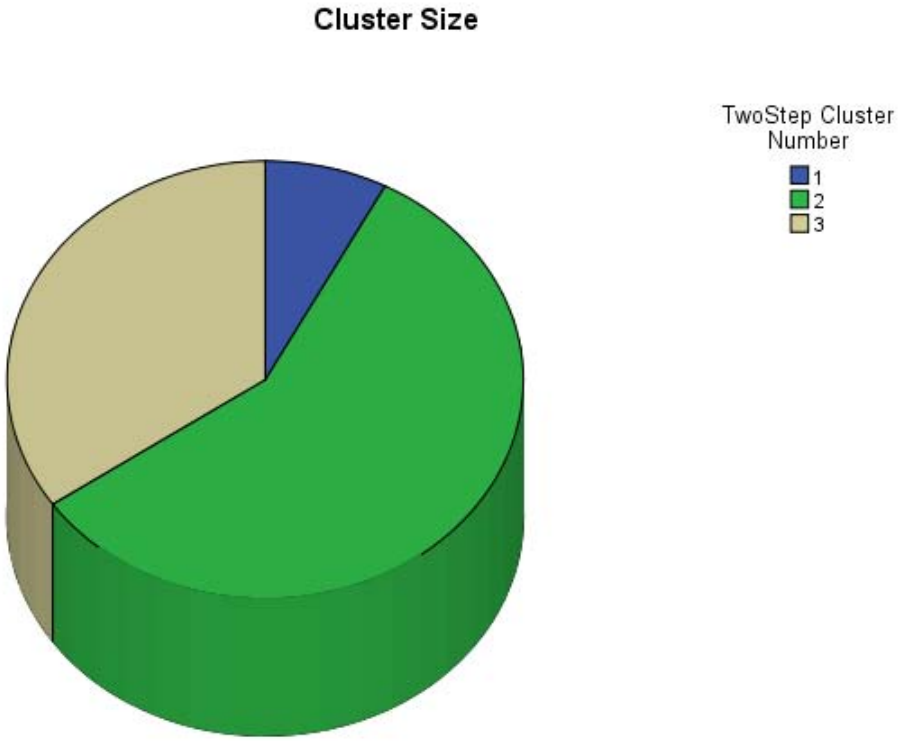
Centroids

		دامنه.دما		دامنه.حداقل
		Mean	*Std. Deviation*	*Mean*
Cluster	*1*	*.41350*	*.038891*	*.61200*
	2	*.24653*	*.030028*	*.56187*
	3	*.24311*	*.065693*	*.47589*
	Combined	*.25819*	*.063557*	*.53596*

Centroids

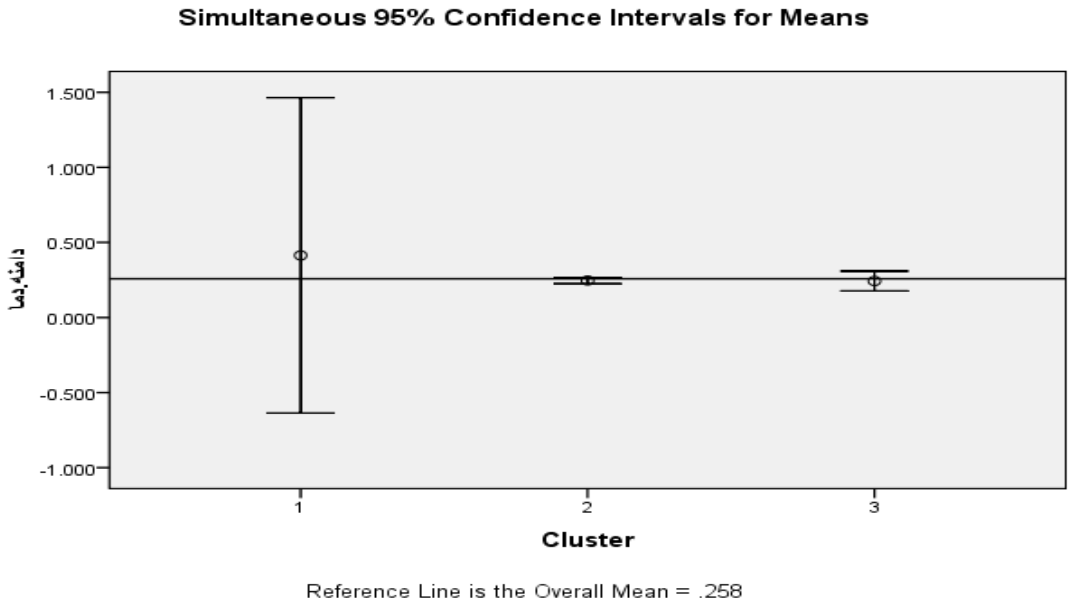
		دامنه.حداقل	دامنه.حداکثر		دما	
		Std. Deviation	*Mean*	*Std. Deviation*	*Mean*	*Std. Deviation*
Cluster	*1*	*.002828*	*.81550*	*.037477*	*13.35000*	*2.899138*
	2	*.039655*	*.73693*	*.036082*	*21.54000*	*3.111683*
	3	*.035600*	*.64456*	*.047948*	*30.51111*	*4.015110*
	Combined	*.058744*	*.71100*	*.066238*	*24.01538*	*6.235684*

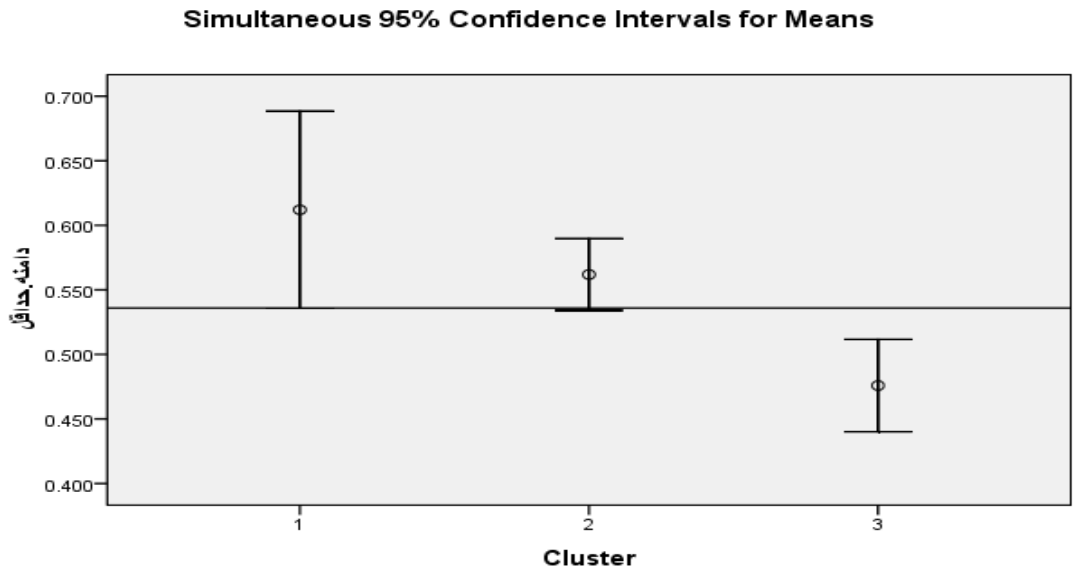
نمودار زیر تعداد و سهم هر خوشه را نشان می دهد . در نمودار زیر سهم خوشه سوم در تحلیل مذکور را نشان می دهد.

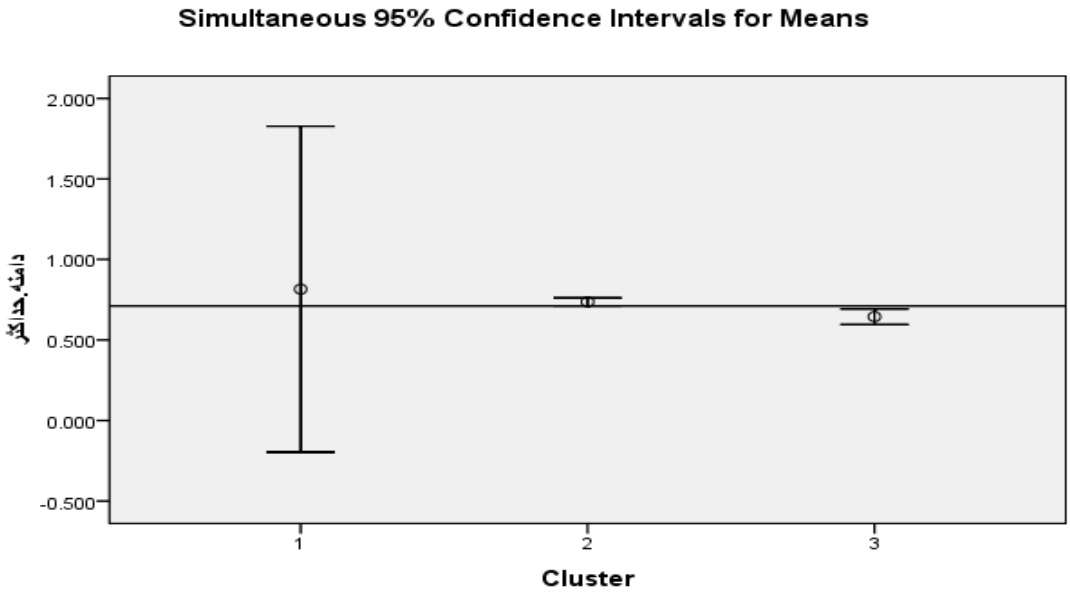


نمودارهای زیر میانگین مرکز ثقل دامنه دما برابر با ***0/258*** است در حالی که مقدار آن برای خوشه اول ***0/4135*** برای خوشه دوم ***0/2465*** و برای خوشه سوم ***0/2431*** می باشد. مرکز ثقل دامنه حداقل برابر با*0/536* است و مرکز ثقل دما برابر با *24/015* است .

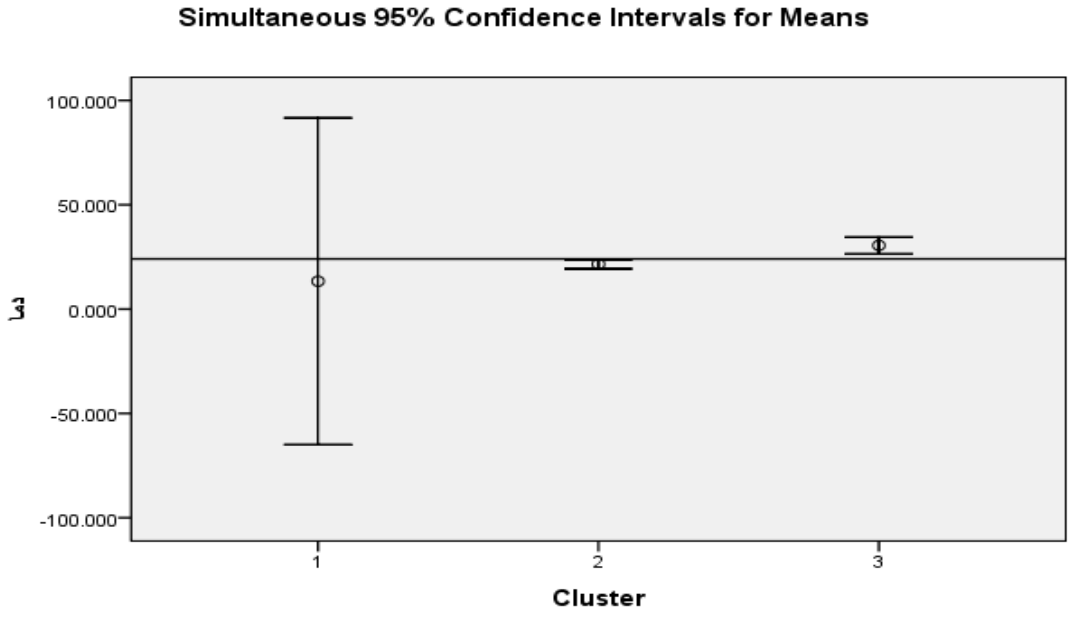
Within Cluster Variation





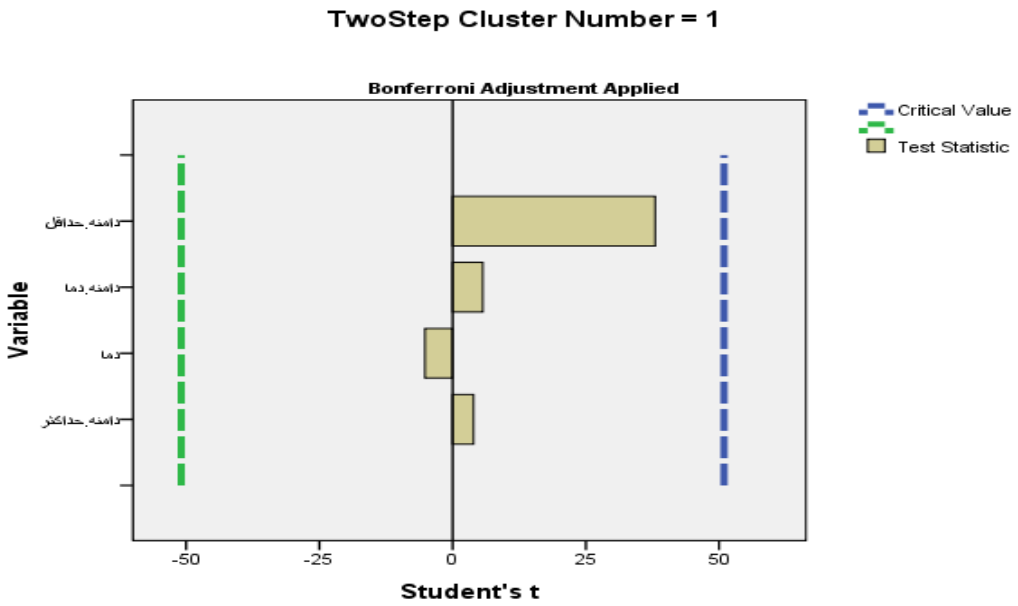


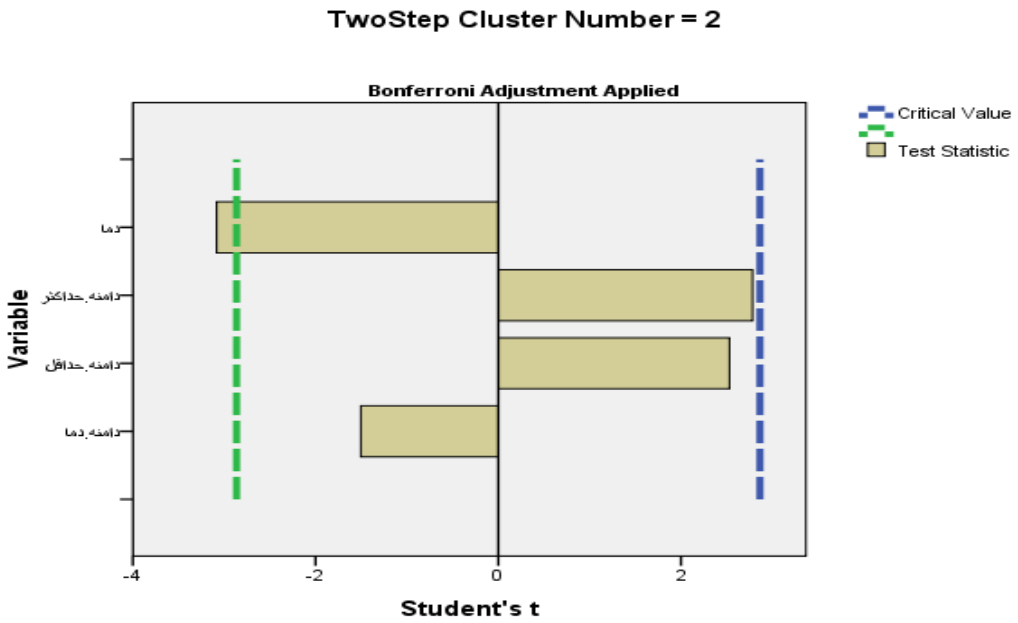
Reference Line is the Overall Mean = .711

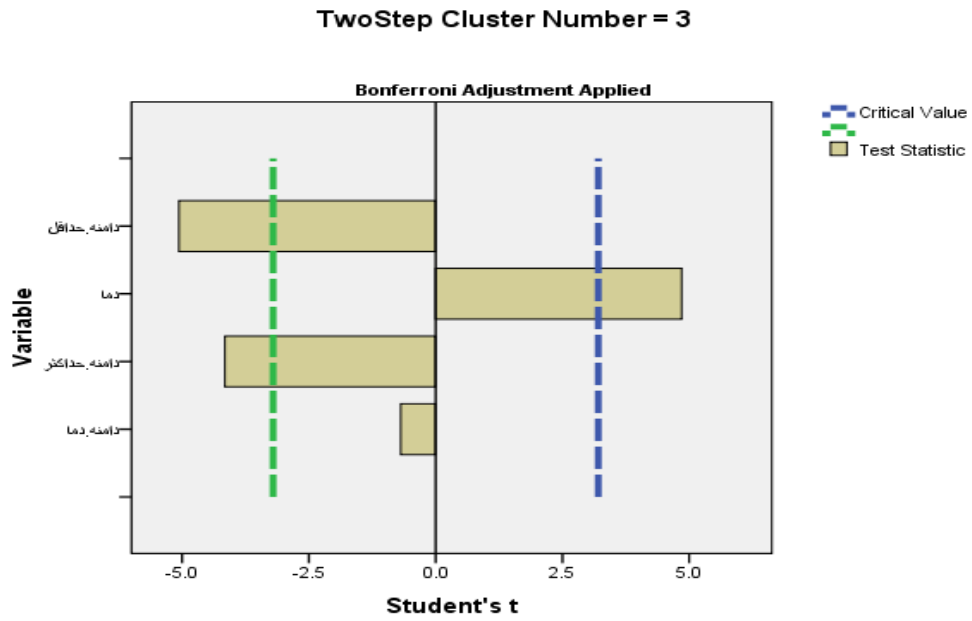


Reference Line is the Overall Mean = 24.015

نمودارهای زیر به نمودار های اهمیت بر حسب متغیر معروف اند. در هر نمودار متغیرها به صورت نزولی بر اساس ترتیب اهمیت شان روی محور Y نشان داده شده است . خطوط عمودی هر نمودار مقادیر بحرانی برای تعیین اهمیت هر متغیر را علامت گذاری می کنند. به این معنی که زمانی مقدار آماره t برای یک متغیر (چه در جهت مثبت و چه درجهت منفی) بزرگتر از عمودی باشد آن متغیر مهم است . مثلا در نمودار زیر مقدار t برای متغیرها کوچک تر از خطوط عمود هستند لذا می توان گفت این متغیرها اهمیت زیادی در تشکیل خوشه ها ندارند. وقتی که مقدار t یک متغیر مثبت باشد نشان می دهد میانگین آن متغیر از میانگین خوشه بزرگ تر است و برعکس مقدار منفی t نشان می دهد که میانگین متغیر از میانگین خوشه کوچک تر است. برمثال برای دو خوشه مقدار t برای متغیر دما بزرگ تر از خطوط عمودی می باشد لذا می توان گفت این متغیر اهمیت زیادی در تحلیل خوشه ای این نمونه دارد و مقادیر t برای دامنه دماحداکثر و حداقل مثبت است و نشان می دهد که میانگین متغیرهای دامنه دمای حداقل و حداکثر از میانگین خوشه ها بزرگ تر است و چون مقادیر t متغیرهای دما و دامنه دما منفی است لذا مقدار میانگین دما و دامنه دما کوچک تر از میانگین خوشه ها هستند و ...




TwoStep Cluster Number = 2
Bonferroni Adjustment Applied
Critical Value
Test Statistic
Variable
دما
دامنه_حداکثر
دامنه_حداقل
دامنه_دما
-4
-2
0
2
Student's t


TwoStep Cluster Number = 3
Bonferroni Adjustment Applied
Critical Value
Test Statistic
Variable
دامنه_حداقل
دما
دامنه_حداکثر
دامنه_دما
-5.0
-2.5
0.0
2.5
5.0
Student's t

منابع و مأخذ:

منابع فارسی :

1- ابراهیم پور , ملیحه و نوروززاده رحیمی , ماندانا: تحلیل رگرسیون همراه با $Minitab_{14}$, انتشارات اقلیدس , 1386.

2- ابونوری اسمعیل : آمار توصیفی و کاربرد آن , انتشارات دانشگاه مازندران , 1386.

3- آذر , عادل و مومنی , منصور : آمار و کاربرد آن در مدیریت , جلد اول , انتشارات سمت , 1380.

4- _____ : آمار و کاربرد آن د ر مدیریت ، جلد دوم ، سمت ، 1380

5- افشین نیا ، منوچهر : روشهای آماری و کاربرد آن در علوم ، انتشارات اتا ، 1372 .

6- بایزیدی , ابراهیم و خسرو آفریدون , عباس : تحلیل آماری با $Minitab_{15}$, انتشارات عابد, 1389.

7- بخشی , بهنام : کاربرد SPSS در تجزیه و تحلیل آماری کشاورزی ..., انتشارات سپهر , 1388.

8- بی همتا , محمد رضا و زارع چاهوکی , محمد علی : اصول آمار در علوم منابع طبیعی , انتشارات دانشگاه تهران , 1387.

9- تهرانیان ، محمدصادق و بزرگ نیا ، ابوالقاسم (ترجمه) : آمار کاربردی ، انتشارات جهاد دانشگاهی مشهد ، 1374 .

10- تیموریان یانسری , مهدی (ترجمه) : آمار و احتمال برای مهندسی مدرن جلد اول , انتشارات دانشگاه امیرکبیر, 1386.

11- _____ : آمار و احتمال برای مهندسی مدرن جلد دوم , انتشارات دانشگاه امیرکبیر, 1386.

12- جاوری ، مجید : شیوه های تجزیه و تحلیل کمی در اقلیم شناسی (با تأکید برمدل های روند) , انتشارات پیام رسان , 1388.

13- _____ : شیوه های تجزیه و تحلیل کمی در اقلیم شناسی (با تأکید برمدل های فصلی) , انتشارات پیام رسان , 1389

14- _____ : مقدمه ای بر روش های تحقیق در علوم انسانی با تأکید بر جغرافیا ، انتشارات طلور ، 1384.

15- _____و صابری فر ، رستم : روش تحقیق در جغرافیا : انتشارات دانشگاه پیام نور ، 1389.

16- _____ و بابازاده ، سید احمد (زیرچاپ): تجزیه و تحلیل داده های اقلیمی با استفاده از تحلیل رگرسیونی ، جغرافیا و برنامه ریزی منطقه ای ، پیام نور ،1388 .

17- جمشیدیان , احمد رضا و نوری زاد, مهدی : طرح و تجزیه و تحلیل آزمایشها با نرم افزار$Minitab_{14}$, انتشارات ارکان دانش , 1385.

18- حسینی , سید یعقوب : آمار ناپارامتریک ..., انتشارات دانشگاه علامه طباطبایی , 1382.

19- حسینی , سید یعقوب : آمار ناپارامتریک ..., انتشارات دانشگاه علامه طباطبایی , 1382.

20- رضایی ، عبدالمجید و سلطانی افشین : مقدمه ای بر رگرسیون کاربردی ، انتشارات دانشگاه صنتعتی اصفهان ، 1377.

21- صادقپور , بهرام و مرادی وهاب : کنترل کیفت آماری در $STATISTCA_{R7}$, انتشارات مبعث, 1386.

22- _____تحلیل پیشرفته آماری با نرم افزار $STATISTCA_{R7}$, انتشارات دانشگاه مازندران , 1386.

23- صالحی صدقیان , جمشید و ابراهیمی , ایرج : تحلیل آماری پیشرفته , انتشارات هوای تازه , 1381.

24- طالبی ، هوشنگ و ... (ترجمه) : طرح و آزمایش ها و روش های آماری ، انتشارات دانشگاه اصفهان ، 1379 .

25- علیجانی ، بهلول و کاویانی ، محمد رضا ، مبانی آب و هواشناسی ، انتشارات سمت ، 1374.

26- غیاثوند , احمد : کاربرد آمار و نرم افزار SPSS در تحلیل داده ها , انتشارات لویه 1387.

27- فرشادفر , عزت اله : اصول و روشهای آماری چند متغیره , انتشارات دانشگاه رازی , 1380.

28- کاظم نژاد, انوشیروان و... (ترجمه) 100 آزمون آماری به همراه راهنمای نرم افزار SPSS , انتشارات دیباگران تهران , 1380.

29- کیامنش , علیرضا و کبیری , مسعود(ترجمه) : رگرسیون و همبستگی کاربردی , انتشارات جهاد دانشگاهی علامه طباطبایی , 1388.

30- گودرزی , سعید : کاربرد آمار در علوم اجتماعی همراه با دستورات ... , انتشارات جامعه شناسان , 1388.

31- ملکمیان , لینا و برومند زاده , تقی (ترجمه) : در آمدی بر روش های آماری ... , انتشارات دفتر پژوهشهای فرهنگی , 1379.

32- مهدوی ، مسعود : آمار و روش های تجزیه و تحلیل داده های جغرافیا ، انتشارات ققنوس ، 1375 .

33- منصورفر , کریم : روش های آماری پیشرفته , انتشارات دانشگاه تهران , 1387.

34- نصیری , رسول : آموزش گام به گام نرم افزار SAS , انتشارات گستر , 1384.

35- نیکوکار , مسعود و تربیتی قره باغ , حبیب و... : آموزش کاربردی و آموزش نرم افزاری S-PLUS ، انتشارات گسترش علوم پایه , 1384.

36- نیرومند ، حسینعلی (ترجمه) : رگرسیون خطی کاربردی ،انتشارات دانشگاه فردوسی مشهد ، 1374 .

37- نیرومند ، حسینعلی (ترجمه) : تحلیل سری های زمانی ، روش های یک متغیرو چند متغیری ، انتشارات دانشگاه فردوسی مشهد ، 1376.

38- هژبرکیانی , کامبیز(ترجمه) : تئوری و مسائل آمار و اقتصادسنجی , انتشارات نی , 1379.

39- هومن , حیدرعلی : تحلیل داده های چند متغیری در پژوهش رفتاری , انتشارات پارسا, 1380.

منابع خارجی :

38- Abraham,B.&Ledolter ,J.(1983) : Statistical Methods for Forecasting , Newyork , John Wiley & sons , USA .

39- Armstrong .J.S (1978): Long-Range Forecasting , Newyork , John Wiley , USA .

40- Barlow ,R.J (1989) : Statistics - A Guide ... John Wiley Englond .

41- Brown , R. G (1959) : Statistical Forecasting for Inventory Control , Newyork , McGraw- Hill , USA .

42- Cilchrist . W (1976): Statistical Forecasting , Nowyork , John Wiley,USA.

43- Draper . N & Smith , H (1981) : Applied Regression Analysis , Englewood cliffs - Prentice -Hall .

44- Folland , C. K & Parker D. E(1999) : Worldwide Surface Temperature Trends to the Mid-19th Century . Meteorological office Landan .

45- Grotch .S. L (1999) : A statistical - Intercomparison Predictied by Four General Circulation Models in the Historical Data , Elsevier ,Newyork .

46- Howell D. C. (1989) : Fundamental Statistics for the Behavioral Sciences . Pwskent . USA.

47- King . L . J (1969) : Statistical Analysis in Geography , Prentice - Hall ,USA .

48- Thomopoalos ,N.T(1980) : Applied Forecasting Methods , Englewood cliffs- prentice - Hall .

This document was created with Win2PDF available at http://www.daneprairie.com.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.