

کاربرد آمار و احتمالات در اقلیم شناسی

جلد اول

تألیف:

دکتر مجید جاوری

مقدمه:

آمار و مفاهیم آن در ادبیات علمی و دینی ما زیاد به آن اشاره شده است در قرآن کریم در آیات بی شماری به مفاهیم آماری اشاره شده است. به طوری که در سوره ابراهیم در آیه 34 خداوند متعال می فرماید:

**وَأَتَاكُمْ مِنْ كُلِّ مَا سَأَلْتُمُوهُ وَإِنْ تَعَدُّوا نِعْمَتَ اللَّهِ لَا تَحْصُوهَا إِنَّ الْإِنْسَانَ لَظَلُومٌ كَفَّارٌ**

> و از هر چه از او خواستید به شما عطا کرد و اگر نعمت خدا را شمارش کنید نمی توانید آن را به شمار در آورید. قطعا انسان ستم پیشه ناسپاس است <

از دیدگاه اقلیم شناسی آمار ابزاری است که می توان به کمک آن راز و رمز داده ها و متغیرهای جمع آوری شده را در یک تحقیق آشکار نمود. وقتی مسئله و طرح تحقیق فراهم گردید و مطالعه انجام شد. اقلیم شناس به جمع آوری می پردازد. اقلیم شناس برای روش کردن پاسخ سؤال های تحقیق خود به ابزارهای مختلفی رو می آورد که مهمترین آن آمار است. این ابزار به اقلیم شناس کمک می کند تا آنچه را در مطالعه انجام داده است بسنجد و سپس تجزیه و تحلیل کند و بعد از تحلیل دست به پیش بینی بزند و در نهایت مطالعه انجام گرفته را در قالب طرح تحقیقی خود کنترل نماید. بر این اساس آمار بخشی از فرآیند تحقیقی هر اقلیم شناس می باشد. نقش این ابزار را می توان بیشتر هنگام تصمیم گیری و پیش بینی مشاهده نمود. روش های آماری در تحقیقات اقلیمی برای خلاصه کردن و توصیف کردن داده ها، سنجش داده ها، تحلیل داده ها، پیش بینی داده ها و کنترل داده ها مؤثر هستند. آمار در اقلیم شناسی جنبه توصیفی، استنباطی و پیش بینی دارد. بر این اساس این موضوع در سه جلد در حال تدوین است که در این جلد سعی شده است به جنبه های توصیفی آن بیشتر توجه شود. این کتاب در پنج فصل تنظیم شده است. فصل اول به کلیات اشاره داد و فصل دوم به کاربرد نمودارها در اقلیم شناسی پرداخته است، فصل سوم به بررسی آماره های توصیفی و تغییر پذیری آنها پرداخته است

، فصل چهار به شاخص های پراکندگی متغیرهای اقلیمی اشاره نموده است و فصل پنجم به کاربرد همبستگی و رگرسیون در تحلیل های اقلیمی اشاره می کند.

جاوری

1389

## فصل اول : کلیات :

## آمار و حدود قلمرو آن در اقلیم شناسی :

اقلیم شناسی سطحی وسیعی از متغیرهایی را شامل می شود که برای تحلیل آنها نیاز به شاخص های مختلف آماری دارد . آماره های مختلف آماری اهمیت زیادی در مطالعات و تحقیقات اقلیمی دارند. کاربرد شاخص های آماری در اقلیم شناسی دامنه وسیعی از شیوه های ساده تا پیچیده را در بر می گیرد . گستره وسیع کاربرد اقلیم و شاخه های مختلف آن زمینه های فنون مختلف آماری را ضرورت می بخشد . کاربرد فنون آماری در اقلیم شناسی با مفاهیم متعددی همراه است که باید اقلیم شناس به آن توجه نماید. این مفاهیم عبارتند از :

- مفهوم نمونه و جامعه
- مفاهیم آمار توصیفی و استنباطی
- مفهوم مدل آماری
- مفهوم شیوه های عاملی و ناعاملی
- مفهوم آزمون ها و سنجش آزمون ها
- مفهوم چند متغیره بودن سری ها
- مفهوم مدل های سری زمانی

## 1-2- توصیف آماری و فهم اقلیم :

اقلیم شناسی اساساً شاخه ای از جغرافیای طبیعی است و عمدتاً جنبه های مختلف جو و آب و هوا را توصیف می کند . اقلیم برآورد اولیه حالت متوسط و برآورد قابلیت تغییر پذیری در خصوص آن حالت مانند انحرافات معیار و دیگر شاخص های قابلیت تغییر پذیری جو می باشد. اقلیم شناسی بیشتر توجه خود را به حال معطوف می دارد . هدف اصلی آن توصیف

تعیین شرایط معیار و انحرافات معیار می باشد که باید آن شرایط را با نقشه نمایش داد. این نقشه ها برای پهنه بندی، منطقه بندی و برنامه ریزی اهمیت زیادی دارند. پارادایم تحقیقات اقلیمی از شیوه های توصیفی محض به طرف فهم دینامیسم اقلیم با استفاده از کامپیوتر و شبیه سازی حالت اقلیم و قابلیت تغییرپذیری هدایت می شود. کاربرد شیوه های آماری نقش مهمی در این پارادایم جدید دارد. اقلیم یک نظام دینامیکی است که نه تنها تحت تأثیر عوامل بیرونی قرار دارد بلکه تحت تأثیر عوامل درونی مانند تابش خورشید و ناهمواری سطح زمین قرار می گیرد. روند تکاملی آن کم و بیش با اصول فیزیکی شناخته شده مانند ثابت تکانه زاویه ای کنترل می شود. در نهایت باید حالت کامل نظام اقلیمی (شامل نظام های اتمسفری، هیدروسفری، لیتوسفری، بیوسفریو پدوسفری) را بررسی کرد و شناخت. البته در این راستا شناخت مدل عمومی گردش جوی به عنوان یک نظام واقعی اهمیت دارد. قابل اشاره است که نمی توان روند تکامل یک نظام اقلیمی را در سطح وسیع فهم نمود. براین اساس نمی توان به راحتی نقشه حالت جو را تبیین نمود. در نظام اقلیمی متغیر های حالت شامل مجموعه ای از تمام متغیرها در هر قلمرو جغرافیایی (طول، عرض جغرافیایی و ارتفاع و عمق) می باشد. دلایل متعددی وجود دارد که یک نظام اقلیمی نیاز به فهم آماری دارد. مهمترین آنها عبارتند از:

- اقلیم بوسیله عوامل متعددی کنترل می شود و تنها تعداد کمی از این عوامل را می توان بررسی نمود. برای فهم بهتر اقلیم لازم است بیشتر به منبع یا منابع درونی تغییرات یک نظام اقلیمی توجه نمود.
- دینامیسم یک نظام اقلیمی غیر خطی است. مؤلفه های غیر خطی هیدرودینامیک هستند. فرآیندهای هیدرودینامیک شامل فرآیندهای غیرخطی متعددی مانند تراکم می باشد.
- حالت دینامیسم نظام اقلیمی اساساً شامل فرآیند ناپایداری مانند ناپایداری باروکلینیک در تروپوسفر عرض های میانی است.
- دینامیسم های اقلیم متعدد و متنوع هستند. فرآیندهای هیدرودینامیک انرژی انتقالی از مکانها با مقیاس وسیع تا مکان ها با مقیاس کوچک می باشند.

ناپایداری ها و غیر خطی بودن نظام اقلیمی در زمان معین نظام اقلیمی غیر قابل پیش بینی را مشخص می کند. این مقیاس های زمانی متفاوت در سیستم های فرعی مانند تروپوسفر ، اقیانوس ها در عرض های میانی متفاوت هستند. برای فهم نظام اقلیمی لازم است آن را در قالب مدل های مختلف اقلیمی مورد بررسی قرار داد. مدل های اقلیمی با اقلیم واقعی تفاوت عمده ای در مقیاس مکانی و زمانی در حالت طبیعی دارد. پارادایم مدل های اقلیمی قالب تصادفی و... دارند. برای اقلیم شناس مهم ترین اصل آماری حالت میانگین می باشد. میانگین مذکور بعنوان برآورد حالت میانگین بلندمدت واقعی نظام اقلیمی نیست. میانگین برآورد شده دارای خطاهایی است که در طول مشاهدات هنگام اندازه گیری اتفاق افتاده است. از طرفی حالت میانگین حالت نمونه نمی باشد براین اساس به نقشه های فشار در سطوح 500 هکتوپاسکالی و سطح زمین توجه نمایید. شاخص های دیگر آماری برای بررسی حالت اقلیم مورد استفاده قرار می گیرند که بعضی از آنها خطی و بعضی غیر خطی می باشند . همبستگی در قالب آمار شاخص دیگری است که رابطه خطی بین دو متغیر تصادفی را نشان می دهد. این شاخص آماری خطی میانگین یک متغیر تصادفی را با متغیر دیگر مورد بررسی قرار می دهد. قوی ترین رابطه خطی قوی ترین همبستگی را نشان می دهد. همبستگی جزئی بین متغیر های متوالی به طریقی خودهمبستگی سری را نشان می دهد و همبستگی بین سری متغیرها همبستگی کلی یا همبستگی متقابل را نشان می دهد. نمونه بسیار مهم همبستگی سری شاخص نوسان جنوبی ماهانه می باشد که ناهنجاری های ماهانه اختلاف فشار بین داروین و تاهیتی را مشخص می کند.

یکی از مفاهیم مهم آماری دیگر که در تحقیقات اقلیمی مورد توجه قرار می گیرد مفهوم پایایی یا ایستایی و پایایی دوره ای می باشد. یک متغیر تصادفی یا یک فرآیند تصادفی وقتی که عامل های آماری آن مستقل از زمان باشد ، آن پایا یا ایستا می باشد. بیشتر عامل های اقلیمی که بیش از یک سال جمع آوری می شوند ایستا نمی باشند بلکه ایستای دوره ای<sup>1</sup> می باشند چونکه عامل های اقلیمی قالب های مختلفی از تغییرات دارند. میانگین های بلند مدت ماهانه میانگین متوسط فشار در سطح

دریا یک چرخه سالانه را نشان می دهد که تقریباً سینوسی می باشد. از مفاهیم مهم آماری دیگر پیش بینی و کیفیت پیش بینی می باشد. پیش بینی ها در قالب مدل ها مورد توجه قرار می گیرد و معمولاً پیش بینی طبقه ای متغیرها دامنه تغییرات بالاتر و پایین تر از حد نرمال را مشخص می سازد. نکته ای که در پیش بینی اهمیت دارد بیش بینی کمی نسبت به پیش بینی طبقه ای می باشد. از مفاهیم مهم دیگری که در تحقیقات آماری در اقلیم باید به آن توجه شود مشخصه های مدل های زمانی و مکانی متغیرها می باشد. متغیرهای اقلیمی که در تحلیل های آماری مورد توجه قرار می گیرند قالب زمانی و مکانی دارند.

### تعریف آمار:

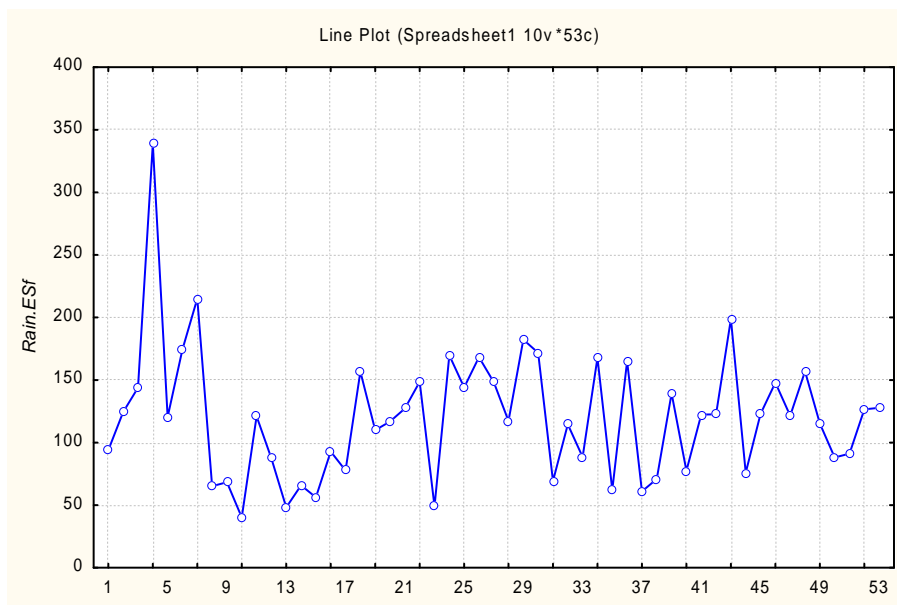
لغت آمار در زبان فارسی معادل واژه انگلیسی **Statistics** است که از ریشه لاتین آن یعنی **State** به معنی دولت گرفته شده است. آمار در لغت به معنی حساب کردن و شمارش کردن است. از نظر روش و محتوا آمار را به جمع آوری، نمایش، طبقه بندی، پردازش داده ها و سپس تحلیل و استخراج برآوردی اطلاعات مورد نیاز تعریف می کند. بر این اساس آمار به روش های جمع آوری، خلاصه کردن، طبقه بندی، تحلیل و پیش بینی و استخراج نتایج از داده ها اطلاق می شود. از نظر اقلیمی کاربرد آمار را می توان به جمع آوری، طبقه بندی، سنجش، تحلیل و پیش بینی داده ها به منظور تصمیم گیری اقلیمی در فرآیند برنامه ریزی به منظور مهیا نمودن مکان بهینه برای زندگی در قالب اقلیم تعریف نمود. در کاربرد آمار از مفاهیم متعددی مانند آماره، پارامتر، متغیر و... استفاده می شود. آماره اندازه عددی است که مشخصه های نمونه آماری را توصیف می کند. پارامتر یا مقدار واقعی کمی است که ویژگی های جامعه آماری را توصیف می کند. متغیر کمی است که مقادیر مختلفی را اختیار می کند. از این نظر اقلیم شناس در کاربرد آمار سه فرآیند زیر را مد نظر قرار می دهد:

1. اقلیم شناسی آماری
2. کاربرد آماری اقلیم
3. اقلیم سنجی آماری

اقلیم شناسی آماری بر فرآیند های جمع آوری ، طبقه بندی ، سنجش و تحلیل داده ها و متغیر های اقلیمی شامل دما ، بارش و... تأکید دارد و کاربرد آماره های اقلیمی شامل جمع آوری ، طبقه بندی ، سنجش ، تحلیل و پیش بینی به منظور تصمیم گیری اطلاق می شود و اقلیم سنجی آماری شامل جمع آوری ، طبقه بندی ، سنجش ، تحلیل و پیش بینی به منظور شناسایی مکان بهینه اقلیمی برای زندگی می باشد. به موقعیت های زیر توجه نمایید:

### موقعیت شماره 1:

دانشجوی اقلیم شناسی داده های 53 ساله بارش ایستگاه اصفهان را در قالب آماره های مختلف جمع آوری ، طبقه بندی ، سنجش و تحلیل می کند و در نهایت داده های بارش 30 ساله ایستگاه اصفهان را در قالب آماره های توصیفی و استنباطی تجزیه و تحلیل می کند . در این مرحله دانشجو در قالب فرآیند اول یعنی اقلیم شناسی آماری اقدام نموده است .

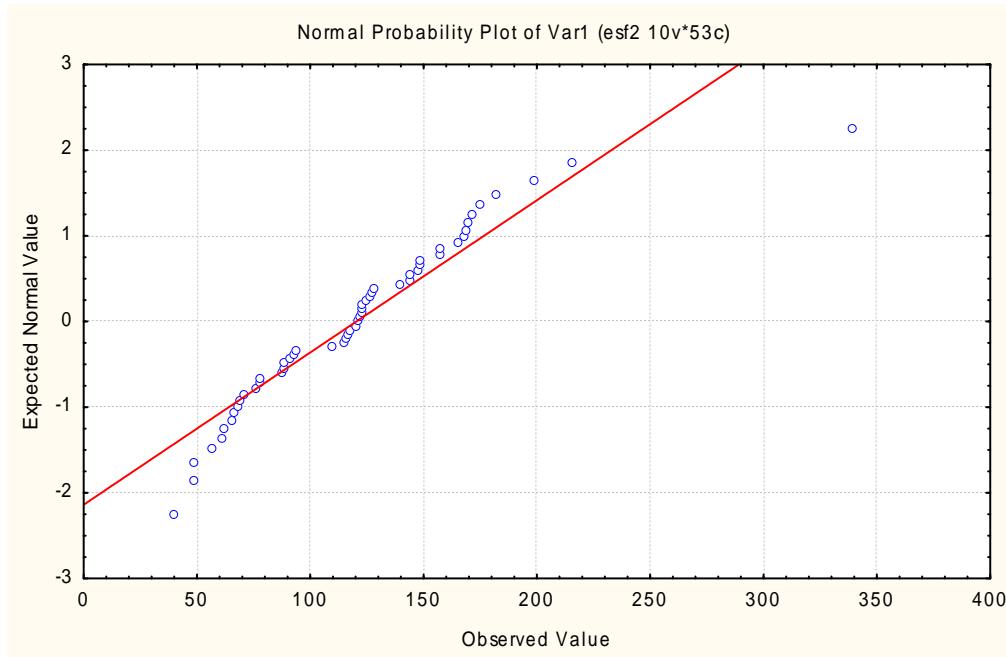


### موقعیت شماره 2:

دانشجوی اقلیم شناسی دیگری داده های 53 ساله بارش ایستگاه اصفهان را جمع آوری ، طبقه بندی ، سنجش و تحلیل می کند و با آماره های استخراج شده به پیش بینی تغییرات بارش در سال های آینده و تحلیل تغییرات زمانی بارش در سال

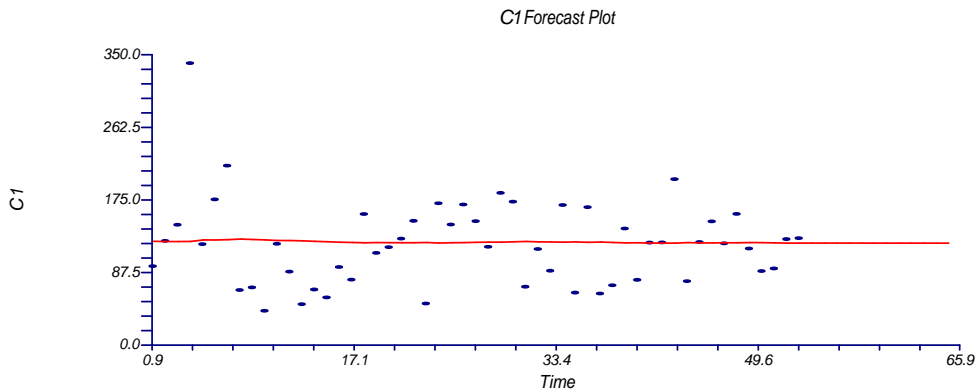


های بعدی را ارائه می دهد . در این مرحله دانشجویان با استفاده از آماره های مختلف به پیش بینی تغییرات اقلیمی پرداخته است و فرآیند دوم را یعنی کاربرد آماره ها در اقلیم شناسی انجام داده است .



### موقعیت شماره 3:

دانشجوی اقلیم شناسی دیگری داده های 53 ساله بارش ایستگاه اصفهان را جمع آوری ، طبقه بندی ، سنجش ، تحلیل ، پیش بینی می کند و با پیش بینی ، اصفهان را از نظر شرایط حاکم بارش پهنه بندی و مناطق بهینه و مناسب از نظر بارش را طبقه بندی می کند . بر این اساس در این مرحله دانشجویان فرآیند سوم یعنی اقلیم سنجی آماری را انجام داده است .



## جامعه و نمونه آماری:

یک جامعه آماری به مجموعه ای از افراد یا اشیاء یا واحدهایی اطلاق می شود که حداقل دارای یک صفت مشترک باشند. جامعه آماری همان محدوده یا قلمروهایی است که تأثیر متغیرها را در آن می سنجد. معمولاً<sup>1</sup> در هر تحقیق، جامعه مورد بررسی یک جامعه آماری است. جامعه آماری با توجه به هدف بررسی و نوع واحد مورد نظر به گونه های مختلف تعریف می شود. تعریف جامعه در جغرافیا باید از نظر زمانی و مکانی شامل همه موضوعات مورد مطالعه باشد. جامعه در تحقیقات جغرافیایی به طور نسبی به جامعه محدود و جامعه گسترده تقسیم می شود. در بیشتر تحقیقات چون مطالعه کل جامعه متناسب با مقتضیات زمانی و مکانی عملی نیست، باید دست به نمونه گیری زد<sup>1</sup>.

نمونه گیری و روش های آن:

محقق به دلایلی مانند صرفه جویی در وقت و هزینه، برنامه ریزی آسان، عملی تر بودن تحقیق و... مجبور است دست به نمونه گیری بزند. نمونه بخشی از یک جامعه است که معرف کل جامعه می باشد. بهترین نمونه آن است که ویژگی های بیشتری از جامعه را در برداشته باشد. در نمونه گیری مهمترین سؤال این است که برای انتخاب نمونه ای که معرف جامعه باشد، نمونه گیری چگونه انجام می گیرد؟ به عبارتی روش نمونه گیری و حجم نمونه چه باید باشد؟ پاسخ به این سؤال بستگی به نوع تحقیق مورد نظر دارد. بنابراین در تحقیقات جغرافیایی سنجش دقت داده ها تحت تأثیر عوامل متعددی است و حجم نمونه ها در این تحقیقات متفاوت خواهد بود. بهتر است برای کاربرد و بررسی نمونه ابتدا واحد نمونه مورد نظر مشخص شود. بعد از مشخص کردن نمونه، اعتبار ویژگی واحد مورد نظر با استفاده از آزمون های دقت سنجی سنجیده می شود. برای انجام این کار می توان از شاخص های مختلف دقت سنجی مانند شاخص میانگین مجذورات خطا، ریشه میانگین مجذورات خطا و... استفاده کرد. با توجه به موضوع و هدف تحقیق باید روش نمونه گیری مناسبی را انتخاب کرد. برای نمونه گیری باید فهرست افراد جامعه در دسترس باشد. این فهرست قالب نمونه گیری را مشخص می کند. روش های نمونه گیری متعددی وجود دارد که به طور کلی می توان روش های نمونه گیری را به دودسته نمونه گیری غیر احتمالی و احتمالی تقسیم نمود. نمونه

<sup>1</sup> برای اطلاع بیشتر به کتاب مقدمه ای بر روش تحقیق در علوم انسانی با تأکید بر جغرافیا، مؤلف مراجعه شود.

گیری غیر احتمالی روش هایی می باشد که براساس قوانین احتمالات صورت نمی گیرد و نمونه ها بر اساس قضاوت شخصی محقق انتخاب می شوند. از روش های نمونه گیری غیر احتمالی در مواردی که نخواهند نتایج نمونه آماری را به جامعه آماری تعمیم دهند ، زمانی که روش نمونه گیری قالب معینی نداشته باشد و زمانی که پیش آزمون وجود دارد می توان استفاده کرد. روش های نمونه گیری غیر احتمالی شامل روش های اتفاقی ، سهمیه ای ، زنجیری و گروهی می باشد . در نمونه گیری احتمالی قوانین احتمال بر نمونه گیری حاکم است و هر یک از اعضای جامعه شانس معینی برای حضور در نمونه را دارند . روش های نمونه گیری احتمالی شامل روش های تصادفی ساده ، سیستماتیک ، طبقه ای ، چند مرحله ای و خوشه ای می باشند . برای اطلاع بیشتر به کتاب روش تحقیق مؤلف مراجعه شود.

### متغیر های اقلیمی :

متغیر از نظر لغوی ، به معنی چیزی است که تغییر می کند ، اما در فعالیتهای تحقیقی ، متغیرها دارای ویژگی هایی می باشند که باید متناسب با انواع آنها مورد مطالعه قرار گیرند . با این وجود هر ویژگی یا خصیصه ای را که کمیت پذیر باشد ، متغیر گویند . تقریباً در هر تحقیق شناخت متغیرها ، اندازه گیری آنها و بررسی نحوه آثار و روابط آنها بایکدیگر مورد نظر است . متغیر یک سبب است که می توان عدد یا ارزش را جایگزین آن کرد. برای تبیین بهتر این موضوع لازم است به انواع متغیر اشاره شود:

#### 1- انواع متغیر :

از نظر ماهیت متغیرها را می توان به دو دسته کلی ذیل تقسیم کرد :

1- متغیرهای کیفی

2- متغیرهای کمی

1- متغیرهای کیفی که متغیرهای مقوله ای نیز نام دارند ، متغیرهایی هستند که مقادیر عددی به خود نمی گیرند یا کمیت پذیر نیستند . متغیرهای کیفی شامل ویژگی هایی از قبیل جنسیت ، مذهب ، نژاد ، زبان ، سواد ، دین و ... می باشند.

2- متغیرهای کمی متغیرهایی هستند که برای اندازه گیری آنها می توان اعداد را به وضعیت های آزمودنی و قابل سنجش برطبق قاعده ای، معین نمود . این متغیرها را متغیرهای کمیت پذیر هم می گویند . مانند : بارش ، دما ، فشار ، سرعت باد ،

رطوبت و ... بر این اساس متغیرهای مورد تحلیل های آماری را در اقلیم می توان به عنوان متغیرهای هواشناسی ، اقلیم شناسی و هیدرواقلیم در نظر گرفت.

لازم است متغیرها را از حالت انتزاعی و مفهومی به صورت عملی تبدیل کرد. یعنی متغیرها به صورت مشاهده پذیر و اندازه پذیر بیان شوند . متغیرهای کمی را می توان به دو نوع ذیل تقسیم کرد :

الف : متغیرهای کمی پیوسته

ب : متغیرهای کمی گسسته

متغیر کمی پیوسته متغیری است که بین دو حد متوالی آن مقادیر بی شماری وجود دارد مانند وزن ، (50/5 کیلو، 50/5 کیلو ، 50/75 کیلو و...) ، قد و ... . در حالیکه برای مقدار متغیر کمی گسسته نمی توان مقدار دیگری را پیدا کرد. مانند: تعداد جمعیت یک شهر یا جنسیت و یا تعداد فرزندان . اگر مرد بود ، زن نیست یا بالعکس و یا اگر تعداد فرزند 2 یا 4 و... بود ، نمی شود گفت که تعداد فرزندان 2/5 یا 4/5 نفر باشند . متغیرها بر اساس نقش خود در تحقیق به انواع ذیل تقسیم می شوند :

1- متغیر مستقل

2- متغیر وابسته

3- متغیر ناخواسته

4- متغیر تعدیل کننده

1- متغیر مستقل :

متغیر مستقل ، متغیری است که مقادیر آن توسط دستکاری محقق تغییر پیدا می کند . به عبارت دیگر محقق آن را ثابت و بی تأثیر از عوامل دیگر فرض کرده ، خود به دستکاری و یا کنترل می پردازد تا تأثیر آن را بر روی عوامل دیگر بصورت کنترل کننده ، مورد بررسی قرار دهد . متغیر مستقل را متغیر درونداد یا محرک هم می گویند . متغیر مستقل ، متغیری است که نقش علت را ایفا می کند . وجه تسمیه این متغیر اینست که مستقل از نتیجه است زیرا نه به چیزی وابسته است و نه چیزی آن را تعیین می کند. متغیر مستقلی که محقق با دخل و تصرف آن را بوجود می آورد متغیر فعال می نامند و متغیر مستقلی که محقق بدون دخالت

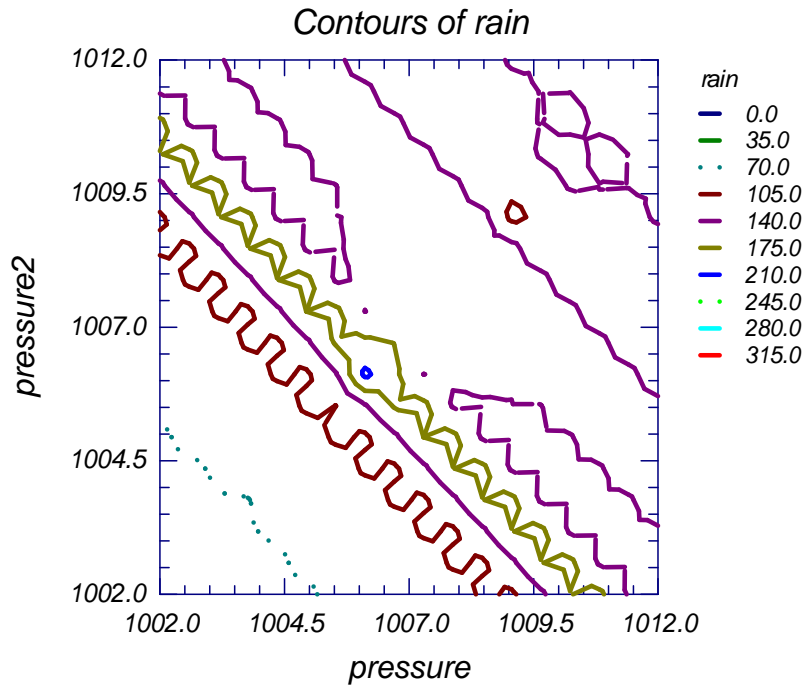
آن را می‌سنجد، متغیر منسوب گفته می‌شود. برای فهم بهتر متغیر مستقل لازم است به چند نمونه اشاره شود. برای مثال اگر محقق اثر ارتفاع را در میزان بارش سواحل خزر بررسی کند، در این تحقیق ارتفاع بعنوان متغیر مستقل می‌باشد. یا هدف محقق بررسی تغییرات زمانی دما و بارش ایران است. در این طرح طول دوره، متغیر مستقل است. در مطالعات از نوع همبستگی به جای متغیر مستقل از اصطلاح متغیر پیش‌بینی و به جای متغیر وابسته، متغیر ملاک استفاده می‌شود. مثلاً "پژوهشگری درصدد پیش‌بینی نوسانات فصلی بارش 35 ساله آینده ایران است، متغیر مستقل مورد استفاده، طول دوره 35 سال آینده است. در بعضی از تحقیقات جغرافیایی نمی‌توان متغیر مستقل را از متغیر وابسته بطور کامل تمایز داد. باید به این نکته توجه داشت که ماهیت متغیر مستقل در هر تحقیق بستگی به مسأله تحقیق دارد.

## 2- متغیر وابسته:

متغیر وابسته، متغیری است که مقادیر آن در اثر تغییر مقادیر متغیر مستقل یا متغیر پیش‌بینی، تغییر پیدا می‌کند. به عبارت دیگر در آنها، تغییر، مستلزم تغییر در متغیرهای مستقل یا پیش‌بینی است. لذا متغیر وابسته، متغیری است که تحت تأثیر متغیر پیش‌بینی کننده است. متغیر وابسته را متغیر برونداد یا پاسخ هم می‌گویند. برای مثال پژوهشگری خواسته است تغییرات زمانی بارش ایران را بررسی کند. در این تحقیق متغیر بارش، بعنوان متغیر وابسته در نظر گرفته شده است و متغیر مستقل یعنی طول دوره بارش را با توجه به اثر آن در متغیر وابسته تحلیل و پیش‌بینی می‌کند. اگر محقق بخواهد اثر بادهای غربی را در گسترش پرفشار سیبری بررسی کند. در این تحقیق پرفشار سیبری بعنوان متغیر وابسته قلمداد می‌شود. فرق متغیر مستقل را با متغیر وابسته می‌توان در موارد زیر جستجو نمود:

- متغیر مستقل نقش علت و متغیر وابسته نقش معلول را بازی می‌کند.
- متغیر مستقل در اختیار محقق است ولی متغیر وابسته در اختیار محقق نمی‌باشد.
- متغیر مستقل برای محقق پیش‌بینی کننده است ولی متغیر وابسته مورد پیش‌بینی قرار می‌گیرد.

برای فهم بهتر بررسی اثر فشار (متغیر مستقل) را بر روی بارش (متغیر وابسته) اصفهان در موقعیت شماره 5 می‌توان دید.  
موقعیت شماره 5:



نمودار بالا رابطه بارش و فشار هوا را در ایستگاه مذکور نشان می دهد به طوری که هر چه فشار افزایش یافته میزان بارش کاهش یافته است .

### 3- متغیر ناخواسته :

متغیر های ناخواسته ، متغیر هایی هستند که در جریان تحقیق از طریق عوامل تحقیق ، اختلال و مزاحمت ایجاد می کنند . به همین جهت به آنها متغیرهای مزاحم یا مانع نیز گفته می شود. تأثیر این متغیرها می تواند تخریبی و حتی اصلاحی باشد . یعنی در نتیجه پژوهش تأثیر بگذارد . بر این اساس تأثیر ناخواسته ای بدون نظر محقق در متغیر مستقل و نتایج تحقیق ایجاد می کند . از این متغیرها گاهی بعنوان متغیر های کنترل هم یاد می کنند. در مطالعات اقلیمی عواملی مانند عوامل طبیعی ، قلمرو جغرافیایی ، طول دوره های آماری و ... بعنوان عوامل ناخواسته در نظر گرفته می شوند . در بعضی موارد این متغیرها را می توان کنترل نمود . کنترل این متغیر ها به دو صورت ذیل انجام می گیرد :

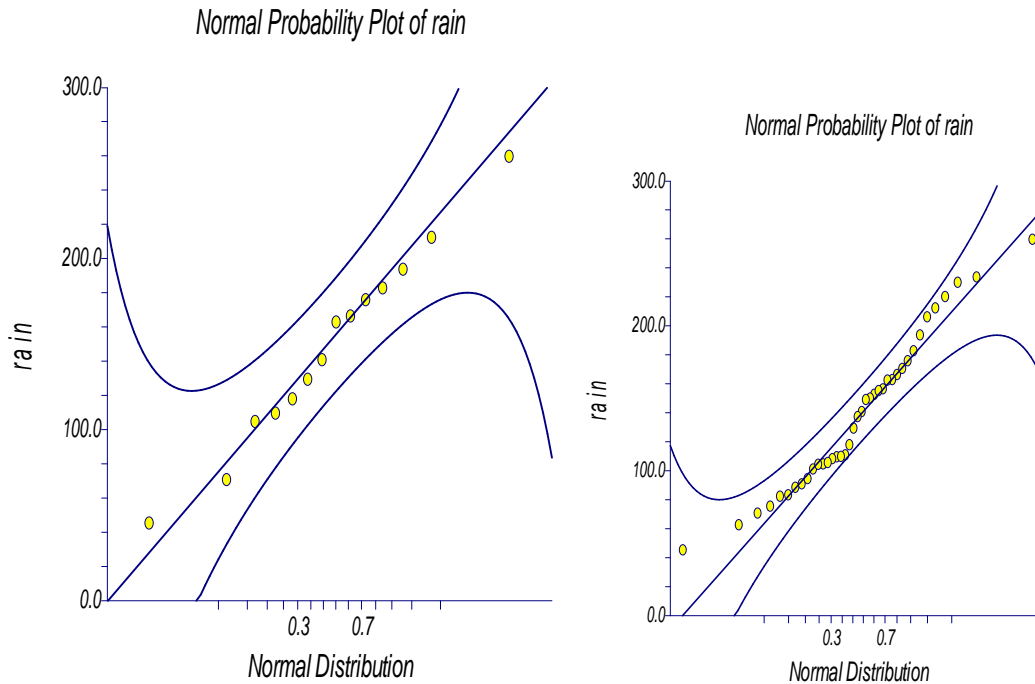
- کنترل تحقیقی .
- کنترل آماری .

کنترل تحقیقی در قالب طرح تحقیق مطرح است . برای مثال محققى به منظور بررسی تغییرات زمانی دما و بارش ایران مجبور به کنترل تحقیق خود می شود . لذا سعی می کند در طرح تحقیق خود ایستگاه هایی را انتخاب کند که طول دوره آماری آنها کمتر از 35 سال نباشد . براین اساس عامل طول دوره با کنترل طرح تحقیق خنثی شده است . در کنترل آماری با روش های مختلفی مانند روش های رگرسیونی ، کوواریانس و ... می توان عملیات کنترل را انجام داد .

#### 4- متغیر تعدیل کننده :

مطالعه این نوع متغیر به طور مستقیم مورد نظر نیست ، بلکه به عنوان یک عمل ثانوی در تحقیق مورد توجه محقق می باشد . به طریقی این متغیر، متغیر مستقل ثانوی است که می تواند متغیر وابسته را تحت تأثیر خود قرار دهد . برای مثال پژوهشگری می خواهد اثر تغییرات زمانی میزان بارش ایران را بررسی کند . در این طرح طول دوره سری به عنوان متغیر مستقل در نظر گرفته شده ، میزان بارش ، متغیر وابسته و طولانی بودن دوره بعنوان یک متغیر تعدیل کننده می باشد . زیرا چگونگی تأثیر متغیر مستقل بر متغیر وابسته ، دوره زمانی طولانی می باشد . لذا نقش دوره طولانی ، به عنوان یک عامل ثانوی است . برای اطلاع بیشتر به کتاب روش تحقیق مؤلف مراجعه شود . به هر حال اقلیم بعنوان یک نظام جوی تحت تأثیر متغیر های بی شماری است که در تحلیل های آماری باید به آنها توجه شود . نظام اقلیم از تمام فرآیندهایی که به طور مستقیم و غیرمستقیم محیط جوی اکوسیستم ها و جوامع انسانی تحت تأثیر قرار می دهند ، تشکیل شده است . مؤلفه های اصلی نظام حالت های هیدرودینامیک جو و اقیانوس است . مقادیر و تعداد بی شمار متغیرهای اقلیمی توزیع چولکی دارند . به عبارت دیگر میانگین ها یا فراوانی ها در دوره های طولانی تر میل به توزیع نرمال پیدا می کنند . برای تبیین موضوع به موقعیت شماره 6 که توزیع بهنجاری بارش ایستگاه کاشان (1967 تا 2005) را در دو دوره زمانی نشان می دهد ، توجه نمایید .

دوره 39 ساله بارش ایستگاه کاشان :



دوره 15 ساله بارش ایستگاه کاشان :

از نکات مهمی که در تحلیل آماری داده های اقلیمی باید به آن توجه نمود این است که اقلیم شناس باید مقیاس داده ها را مدنظر قرار دهد. دو مفهومی که معمولاً در تحلیل های اقلیمی بکار می رود مقیاس زمانی و مقیاس فاصله (مکانی) است. مقیاس فاصله مانند فرآیند یک طوفان جنب حاره ای است که مقیاس فاصله آن ممکن است بعنوان قطر آن یا فاصله آن بین یک حداقل فشار و حداکثر فشار بسته باشد. مقیاس زمانی مدتی است که یک پدیده در محیط اتفاق می افتد. اندازه گیری مدت ظهور هر پدیده اقلیمی از مهمترین شاخص آماری است. از مهمترین موضوعاتی که در کاربرد آمار در اقلیم شناسی مورد توجه قرار می گیرد ، متغیرهای جوی و اقلیمی می باشد . متغیرهای جوی و اقلیمی را می توان برای ردیابی حالت فیزیکی جو مورد استفاده قرار داد.

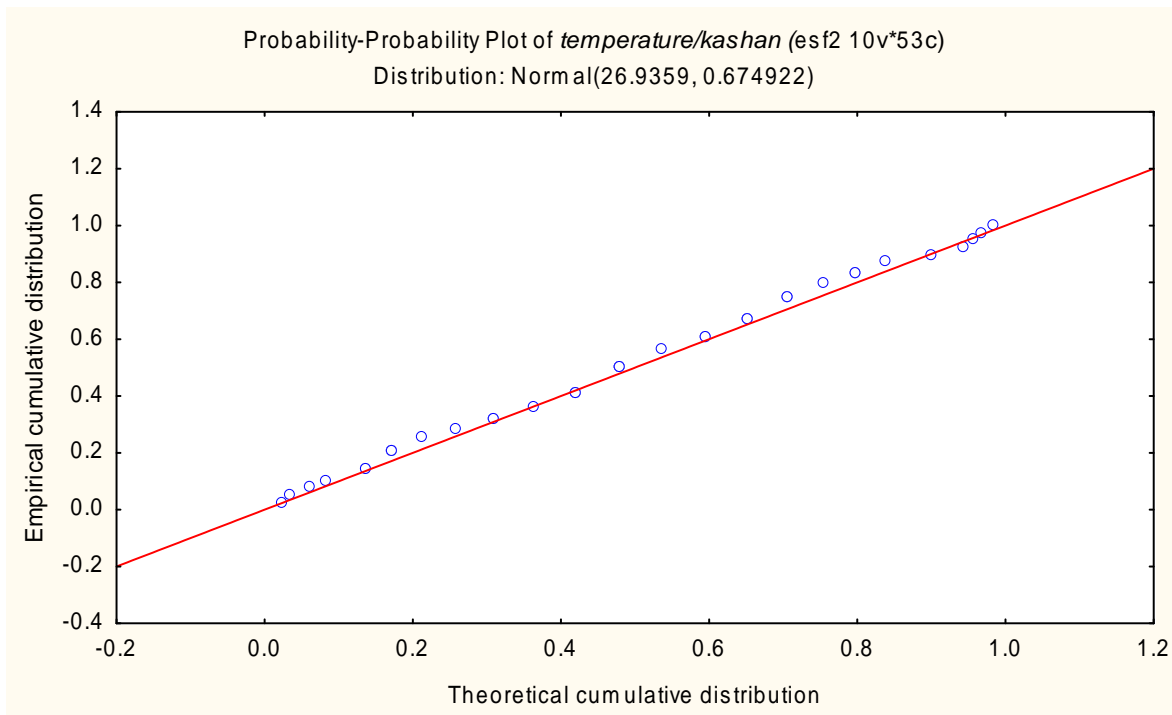


اقلیم محلی معمولا اغلب با ردیابی داده های ایستگاهی مانند دما ، بارش ، فشار ، رطوبت ، ابرناکی ، تابش و باد قابل بررسی می باشد. اقلیم ماکرو معمولا آب و هوا را با داده های شبکه ای<sup>1</sup> مانند فشار سطح دریا ارتفاع ژئوپتانسیل ، دمای ژئوپتانسیل ، باد برداری ، واگرایی ، رطوبت نسبی و تابش موج بلند خروجی می تواند بررسی نماید. داده های شبکه ای اهمیت زیادی در توزیع کامل مکانی دارند. بارش به شکل باران یا برف از متغیرهای بسیار مهم اقلیمی است که برای جو بعنوان آزاد کننده گرمای نهان در ستون هوا قلمداد می شود. دو فرآیند دینامیکی متفاوت یکی همرفت که امکان شرایط عمودی ناپایداری هوا را مشخص می کند و دیگری فرآیند تولید بارش می باشد. کاربردهای متعددی در این سطح از لحاظ آماری وجود دارد . آمار مقدار بارش نمونه ای از آن است که بستگی به زمان تراکم دارد و تعداد روزهای بارانی در ماه اغلب بستگی به مقدار بارش دارد و زمان بین دو وقوع بارش از نکات دیگر می باشد. برای نمایش آماری متغیرهای بارش استفاده از نمودارهای احتمال<sup>2</sup> است . این نمودار نمایش تابع تجربی مربوط به خروجی های ممکن  $y$  در محور قائم و احتمال های تراکمی برآورد شده در محور افقی است . موقعیت شماره 6 نمودار احتمال برای دمای ایستگاه کاشان را نشان می دهد.

موقعیت شماره 6:

---

<sup>1</sup> - Gridded data  
<sup>2</sup> - probability plots



تقارن و عدم تقارن متغیرها از نکات مهمی است که باید آنها باید توجه نمود. در حالت تقارن تفاوتی بین تمایز متغیرهای وابسته و مستقل در نظر گرفته نمی شود در صورتی که در حالت عدم تقارن به وسیله متغیر مستقل، متغیر وابسته پیش بینی می شود. شاخص های تقارن و عدم تقارن در قالب سطوح مختلف سنجش متفاوت هستند. در اقلیم شناسی در حالت تقارن متغیرهای فاصله ای را می توان با شاخص های آماری همبستگی پیرسون، ضریب تعیین و در حالت ناتقارن متغیرهای فاصله ای را می توان با شاخص های آماری رگرسیون و... اندازه گیری نمود. این شاخص ها در جلد دوم کتاب به آن اشاره شده است.

### سنجش اقلیمی<sup>1</sup>:

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 نسبی 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 نسبی

<sup>1</sup> - برای اطلاع بیشتر به کتاب شیوه های تجزیه و تحلیل کمی در اقلیم شناسی (با تأکید بر مدل های روند) مؤلف مراجعه شود.

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 ، متغیرهایی هستند که مقادیر عددی به خود نمی گیرند یا کمیت پذیر نیستند . متغیرهای 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، معین نمود. این متغیرها 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 متغیرهایی 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 جنبه ریاضی \_ آماری ندارد . وقتی متغیر مستقل از چند سطح تشکیل شده باشد ، متغیر 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 دهنده 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 کننده 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 فاصله آنها ، مشخص می کند . مقیاس فاصله ای نه تنها به ترتیب اشیاء توجه دارد بلکه فاصله 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 فاصله آنها با یکدیگر مشخص می کند . در این مقیاس نه تنها مقدار صفت مورد نظر 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 . اکثر داده 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 متغیر مورد مطالعه در تحلیل های کمی دارای توزیع بهنجار باشد و آزمودنی ها بوسیله نمونه گیری تصادفی بدست آمده باشند ، برای تحلیل می توان آزمون های عاملی یا پارامتریک را بکار گرفت. چنانچه متغیر مورد نظر توزیع بهنجاری نداشته باشد برای تحلیل باید از آزمون های ناعاملی یا ناپارامتریک استفاده نمود. آزمونهای ناپارامتری معمولاً "برای متغیرهایی که در آن مقیاس های اسمی و رتبه ای اندازه گیری شده اند ، بکار می رود . در مواردی که تعداد نمونه کوچک باشد برای متغیرهای اندازه گیری شده در مقیاس فاصله ای نیز از آزمونهای ناپارامتریک می توان استفاده نمود. در حالیکه از آزمونهای پارامتریک را می توان برای متغیرهایی که حداقل در مقیاس های فاصله ای اندازه گیری شده اند، استفاده کرد. در تحلیل های پارامتریک برای توزیع متغیر مورد اندازه گیری در جامعه ای که نمونه از آن بدست آمده است ، پیش فرض هایی صورت می گیرد . اعتبار نتایج آزمون پارامتری به صادق بودن این پیش فرض ها بستگی دارد . از جمله این پیش فرض ها بهنجار بودن ، استقلال مشاهدات و برابری واریانس می باشد . آزمونهایی که هیچ گونه فرضی در باره توزیع آنها وجود ندارد ، بصورت تحلیل های آزاد توزیع مطرح می شوند که اغلب ناپارامتریک هستند. در این راستا آزمون های فرضیه ای که استفاده می شوند می توانند یک دامنه و یا دو دامنه باشند .

## آزمون های آماری مربوط به انواع مقیاس :

با گسترش روزافزون تحقیقات اقلیمی و روش های آماری جدید اولین تکنیک های استنباطی بوجود آمده فئونی بودند که برای داده هایی بکار گرفته شده اند که مفروضات فراوانی داشته اند. با توجه به مفروضات کاربرد آزمون ها در این خصوص لازم 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 و رابطه ای می باشد. فرضیه های توصیفی قالب یک متغیره و فرضیه های تفاوتی و رابطه ای دو یا چند متغیره دارند. برای مثال اگر فرضیه تحقیق اقلیم شناس این باشد که با افزایش ارتفاع کوههای البرز بارش کم می شود، این فرضیه توصیفی است و چنانچه فرضیه اقلیم شناس این باشد که بارش از نوع برف نسبت به باران با افزایش ارتفاع البرز کاهش می یابد، فرضیه تفاوتی و چنانچه فرضیه اقلیم شناس میزان بارش در ارتفاعات البرز با سرعت توده هوا رابطه دارد، باشد، فرضیه مورد استفاده فرضیه رابطه ای است. از طرفی در کاربرد آزمون ها باید به مستقل و وابسته بودن داده ها توجه نمود. باید به ادعای محقق 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. فرضیه های تحقیق در مقام آزمون بصورت مخالف یا صفر نوشته می شوند و زمانی که استنباط آماری موجب رد فرضیه صفر بشود، فرض تحقیق پذیرفته می شود. اگر فرضیه تحقیق مبین تفاوت بین دو متغیر باشد، در اینصورت رابطه دو میانگین فوق چنین است  $(\mu_1 \neq \mu_2)$  ولی اگر فرضیه تحقیق، مبین جهت تفاوت فوق نیز باشد، یعنی اگر فرض، مبین بیشتر یا کمتر بودن یکی

نسبت به دیگری باشد، در این حالت رابطه بصورت ( $\mu_1 < \mu_2$  یا  $\mu_1 > \mu_2$ ) خواهد بود. اگر احتمال رابطه فوق که به صورت یک مقدار عددی از آزمون بدست آمده است، تحت فرضیه صفر برابر یا کمتر از  $\alpha$  یعنی سطح معنی داری که غالباً به شکل درصد ( $\alpha = 1\%$ ,  $\alpha = 5\%$ ) مورد استفاده است، باشد فرض صفر رد شده و فرض تحقیق قبول می شود و برعکس. سطح معنی داری در تحلیل های آزمون ها، مبین پذیرش یا عدم پذیرش فرض صفر در سطح معین (99% یا 95% یا 90%) یا (0/01 یا 0/05 یا 0/1) می باشد. همچنین درجه آزادی در این قالب مورد استفاده قرار می گیرد. درجه آزادی به میزان مستقل بودن مشاهدات از یکدیگر اشاره می کند. با توجه به مطالب جدول شماره 1 سطح سنجش و آماره های مربوط را نشان می دهد.

جدول شماره 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 و...)	ناپارامتریک
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 زمانی اقلیم :

اقلیم سنجی از نظر لغوی اصطلاحی است که محقق در این تحقیق مطرح ساخته است و به معنای سنجش و اندازه گیری اقلیم است. اقلیم سنجی تحلیل و مقایسه کمی عناصر اقلیمی در مکان های جغرافیایی به منظور پیش بینی و تصمیم گیری علمی در قالب برنامه ریزی های فضایی است. در راستای اقلیم سنجی بررسی تغییرات تصادفی، گرایش دار، فصلی و دوره ای اهمیت زیادی دارد. 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 مؤلفه مهمی از فرآیند های تصادفی را تشکیل می دهند. مشخصه های مدل ساده 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 هستند. آزمون های ناعاملی سنجش سری های تصادفی، آزمون های توالی، نقاط گردش، دنیلز، تاوکنندال و... هستند و آزمون های عاملی سنجش سری های تصادفی، آزمون های وان نیومن، تابع خودهمبستگی، تومپ، پرایش -بوکس Q و... هستند. کاربرد آزمون ها در قسمت های بعدی کتاب به آن اشاره شده است.

مطالعه 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 دنیلز و برای عاملی آزمون  $t$  بیشتر استفاده می شود. برای آزمون دنیلز قاعده تصمیم گیری به این صورت است که اگر نمره آزمون از نمره جدول کمتر باشد، فرض صفر رد نمی شود و نتیجه گرفته می شود که سری تصادفی هست<sup>1</sup>.

سری زمانی که معمولاً "کوتاهتر از سال می باشد اغلب الگویی از نوساناتی سال به سال را نشان می دهند، این الگوی دوره ای، حرکات فصلی خوانده می شود. این الگوهای فصلی در قالب مدل های ضربی و جمعی مطرح می شوند. هر دوره 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 بعنوان 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<sup>2</sup>. سنجش سری ها در قالب سری های دوره ای هم از موضوعاتی که باید در سنجش داده ها به آن توجه نمود. نکته مهم این است که در متغیری که در سطح سنجش بالاتری قرار دارد، می توان به سطح پایین تر تبدیل کرد و سنجش نمود.

<sup>1</sup> - برای اطلاع بیشتر به کتاب شیوه های تجزیه و تحلیل کمی در اقلیم شناسی (با تأکید بر مدل های روند) از مؤلف مراجعه شود.

<sup>2</sup> - برای اطلاع بیشتر به کتاب شیوه های تجزیه و تحلیل کمی در اقلیم شناسی (با تأکید بر مدل های فصلی) از مؤلف مراجعه شود.



## فصل دوم : کاربرد نمودار ها در اقلیم شناسی:

### انواع نمودار ها :

داده های اقلیمی را که به صورت توزیع فراوانی ها در تحقیقات آماری اقلیم استفاده می شوند ، می توان به صورت نمودار نشان داد. نمایش تصویری داده های اقلیمی برای نتیجه گیری و مدل سازی آماری تغییرات عناصر اهمیت زیادی در تحقیقات اقلیمی دارد. بر این اساس نمودار ابزاری بسیار مناسب برای نمایش و توصیف داده های اقلیمی است. نمودار های مورد استفاده در تحقیقات اقلیمی معمولاً " دو بعدی و سه بعدی هستند. دو بعدی بودن نمودار ها به صورت محور عمودی یا محور  $y$  ها و محور افقی یا محور  $x$  ها نمایش داده می شوند. محل تقاطع دو محور ، مبدأ مختصات گفته می شود و با صفر نشان داده می شود . طول نقاط روی محور افقی در سمت راست مبدأ مختصات با مقدار مثبت و عرض نقاط روی محور عمودی در بالای مبدأ مختصات با مقدار مثبت و طول نقاط روی محور افقی در سمت چپ مبدأ مختصات با مقدار منفی و عرض نقاط روی محور عمودی در پایین مبدأ مختصات با مقدار منفی نشان داده می شوند. نمودار های سه بعدی علاوه بر محور های طولی و عرضی ، محور ارتفاعی یا محور  $Z$  هم برای نمایش داده ها استفاده می شود. نمایش داده ها با نمودار ها دارای مزایایی مانند امکان مقایسه بهتر سری ها اقلیمی ، فهم و تبیین بهتر توزیع و تغییرات داده ها ، نمود مناسب توزیع داد ها و توصیف ساده تر داده ها می باشند.

### انواع نمودار ها مورد استفاده در تحقیقات اقلیمی:

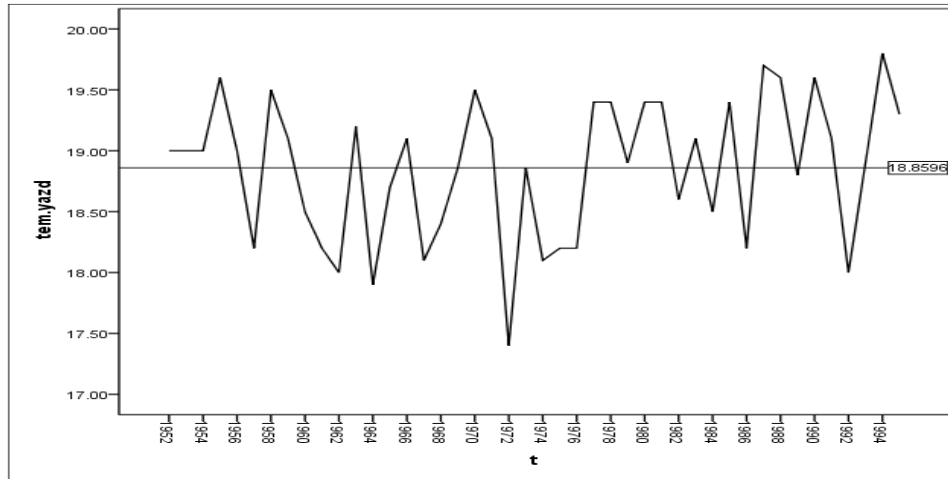
با توجه ماهیت نمودار ها و تحقیقات آماری اقلیم می توان نمودارهای مورد استفاده در تحقیقات اقلیمی به انواع زیر تقسیم نمود :

- 1- نمودار خطی
- 2- نمودار میله ای
- 3- نمودار هیستوگرام
- 4- نمودار جعبه ای
- 5- نمودار چند ضلعی
- 6- نمودار تراکمی
- 7- نمودار شاخه و برگ
- 8- نمودارهای توزیع بهنجاری
- 9- نمودار های کنترل

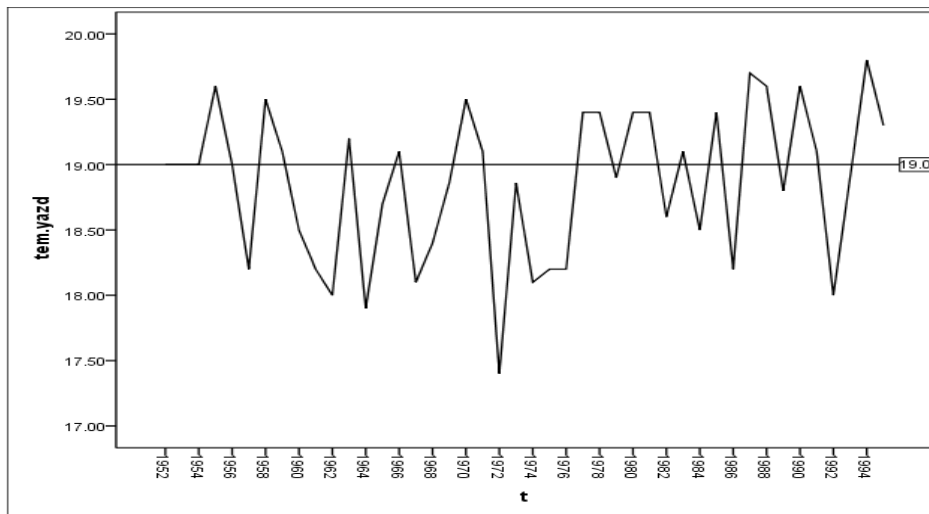
- 10 نمودار علت و معلولی
- 11 نمودار پارتو
- 12 نمودار پراکنش
- 13 نمودار تابع چگالی احتمال
- 14 نمودار خطوط هم ارزش
- 15 نمودار دایره ای
- 16 نمودار سری زمانی
- 17 نمودار برداری
- 18 کلیماتوگرام ها
- 19 نمودار آمپروترمیک
- 20 نمودار شماتیک و ...

#### 1- نمودار خطی

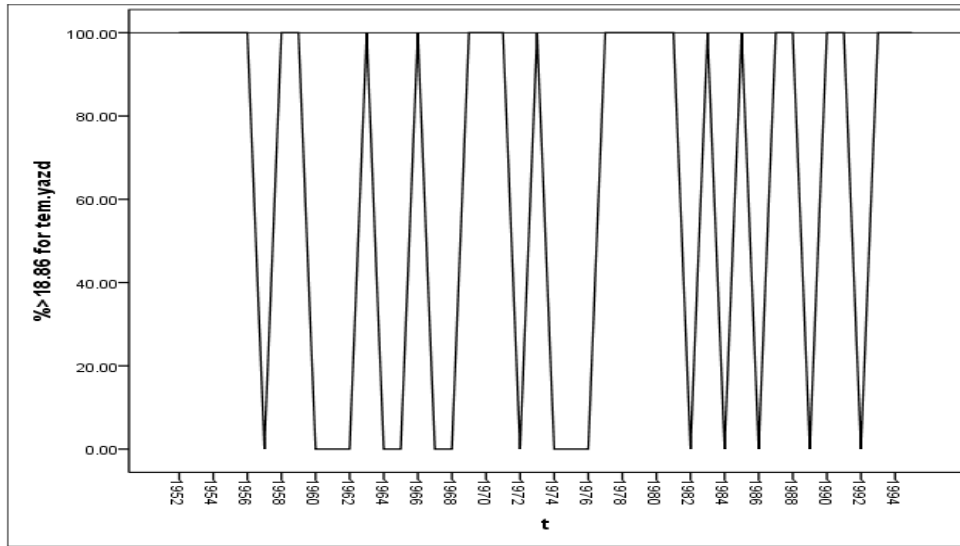
نمودار خطی از مهمترین نمودار هایی است که در تحقیقات اقلیمی از آن استفاده می شود. چون سری های اقلیمی به صورت کمیت های پیوسته ای در نظر گرفته می شود نمایش آنها به صورت خطی مخصوصاً " به صورت فراوانی ها ضروری است . برای ترسیم این نمودار دو سری داده لازم است که در محورهای افقی و عمودی آن مرتب نمود. با مرتب نمودن داده ها در محورهای افقی و عمودی نقاط اتصال به همدیگر را به صورت خطی نشان داده می شود. معمولاً بسته به مطالعات و تحقیقات اقلیمی محور افقی زمان و محور عمودی داده های اقلیمی مانند عناصر دما ، بارش و ... تنظیم می شود. نمایش خطی سری های اقلیمی به صورت تک نمونه و چند نمونه انجام می گیرد. بر این اساس استفاده از نمودار های خطی برای نمایش فراوانی های اقلیمی از اهمیت زیادی برخوردار است . نمودار خطی در مطالعات اقلیمی تنوع کاربردی فراوانی دارد . از این نمودار می توان جهت نشان دادن نمود خطی داده ها در برابر شاخص های مختلف آماری مانند میانگین ، نما ، مد ، انحراف معیار و ... استفاده کرد . همچنین می توان برای درصد فراوانی داده ها نسبت به یک حد معیار استفاده کرد. برای مثال اگر هدف نمایش سری دمای یزد (1995-1952) در برابر میانگین آن باشد ، نمودار خطی آن به صورت زیر نمایش داده می شود:



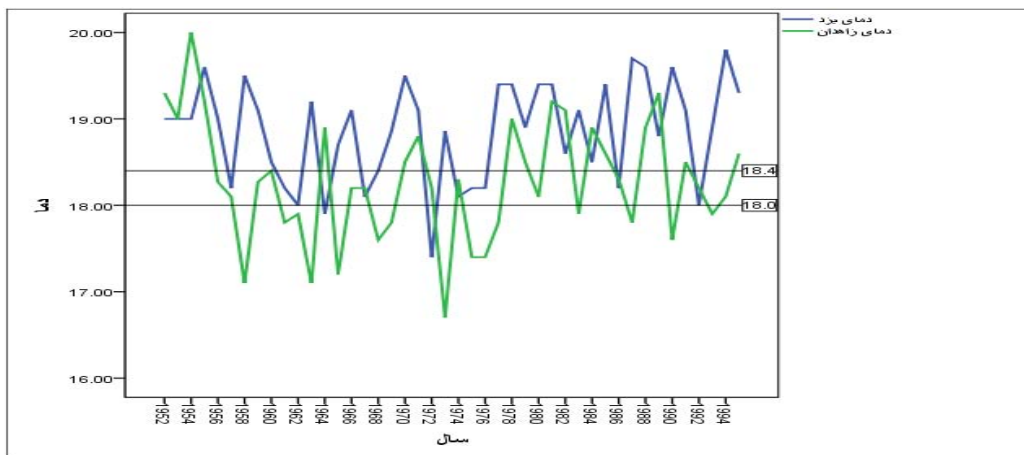
نمودار شماره ۱ توزیع دمای ایستگاه یزد را نسبت به میانگین سری در طول دوره مورد نظر نشان می دهد. اگر هدف نمایش سری دمای یزد در برای میانه سری باشد. نمودار شماره ۲ را می توان نشان داد:



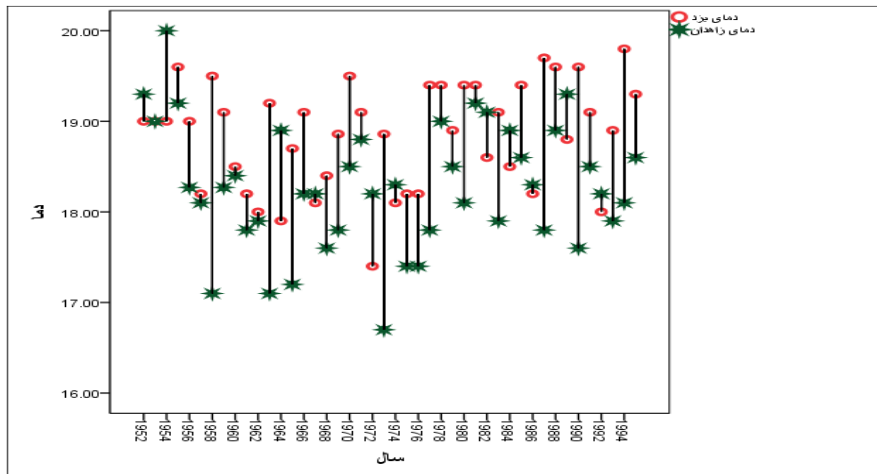
اگر هدف بررسی درصد بالای میانگین سری باشد می توان نمودار مذکور را به صورت زیر نشان داد:



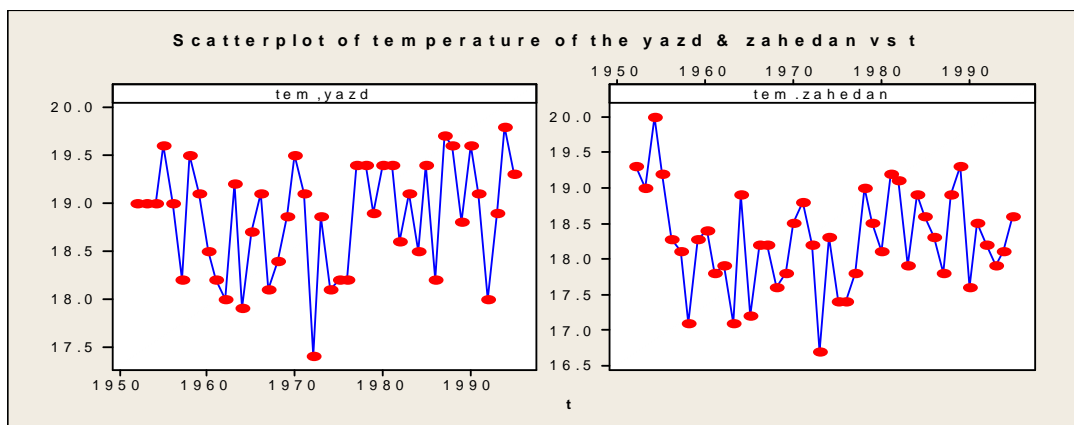
اگر هدف مقایسه دو سری دما برای مثال دمای ایستگاه های یزد و زاهدان باشد ، نمود دو نمونه نمودار خطی به صورت نمودار شماره نشان داده می شود :



قابل اشاره است نمایش نمودار های خطی از تنوع زیادی برخوردار است که می توان آنها ویرایش نمود و سری های اقلیمی را بهتر نمایش داد. برای مثال :



یا:



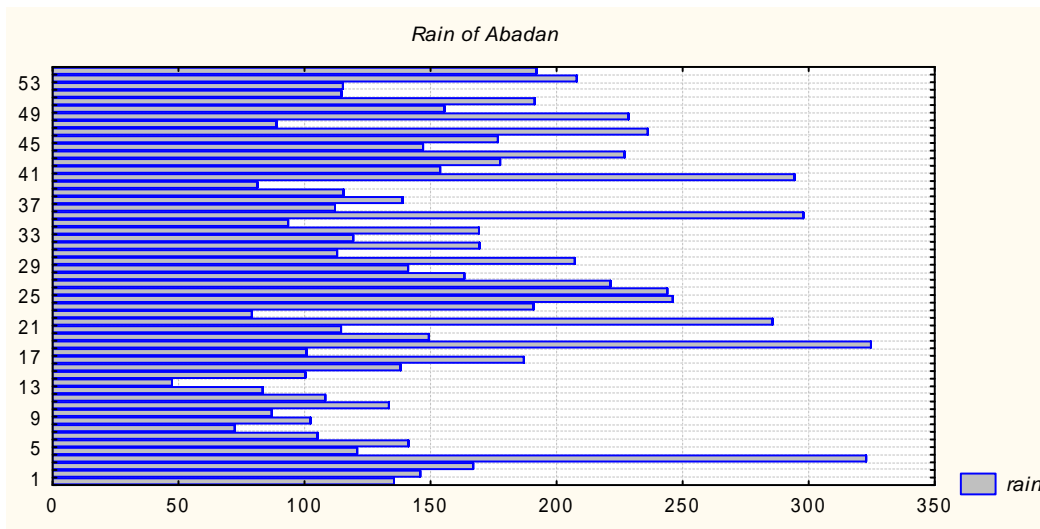
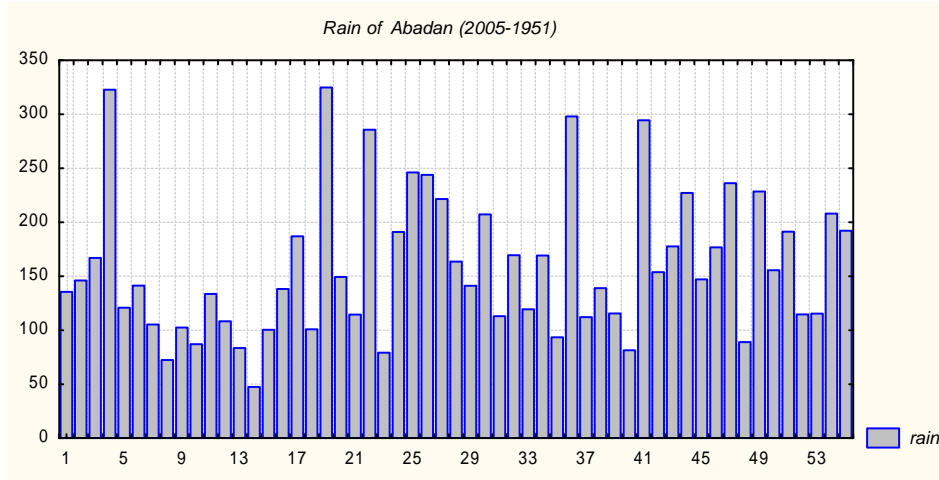
به هر حال نمودار خطی را می توان برای مدل های خطی و منحنی خطی و برازش آنها استفاده کرد. نمایش پراکندگی سری های اقلیمی با نمودار خطی از اهمیت زیادی برخوردار است. نمودار های خطی بسته به ماهیت خطی و منحنی خطی آن به اشکال و انواع مختلفی قابل نمایش می باشند و این قابلیت اهمیت زیادی جهت نمایش داده های اقلیمی در تحقیقات دارند.

## 2- نمودار میله ای :

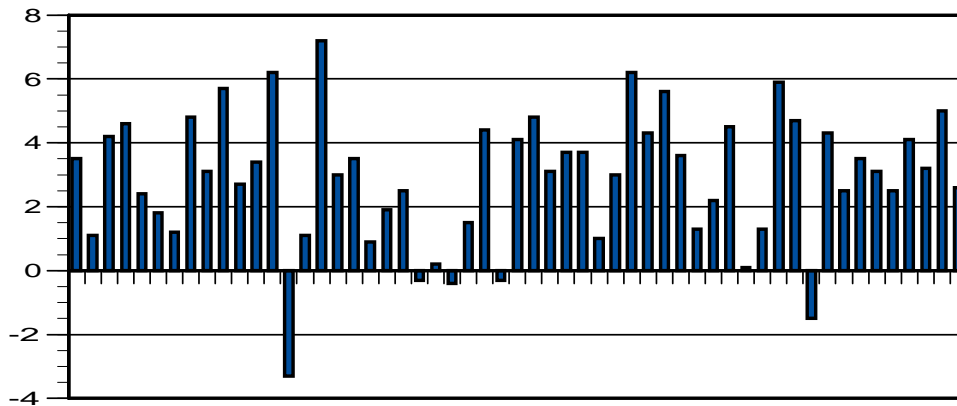
معمولاً برای نمایش توزیع عناصر اقلیمی از نمودار میله ای یا ستونی<sup>1</sup> استفاده می شود. در این نمودار هر میله متناظر با یک رده از توزیع می باشد. میله ها فقط در طول نمایش دامنه تغییرات توزیع را نشان می دهند. نمود کاهشی و افزایشی توزیع با نمودار میله ای قابل توجه می باشد. در نمودار میله ای مقادیر رسته ها را بر روی محور عمودی (Y) و عناوین رسته ها بر روی محور افقی (X) نشان داده می شود. اگر ترتیب انتخاب محورها عوض شود، نمودار میله ای افقی خواهد بود. بر این اساس نمودار های میله ای را می توان به صورت عمودی و افقی نشان داد. ارتفاع هر میله به اندازه فراوانی هر مقوله از متغیر می باشد. اگر در تحلیل با دو متغیر سر و کار داشته باشیم متغیر مستقل را در محور افقی و متغیر وابسته

<sup>1</sup> Bar Chart

را در محور عمودی تنظیم می شود. ماهیت داده های اقلیمی نقش مهمی در نمود نمودار های میله ای دارد. نمود سری دما با نمود سری های اقلیمی دیگر مانند رطوبت، بارش متفاوت است. سری دما به علت داشتن مقدار منفی نمودار میله ای قالب دو بعدی به خود می گیرد. یعنی:



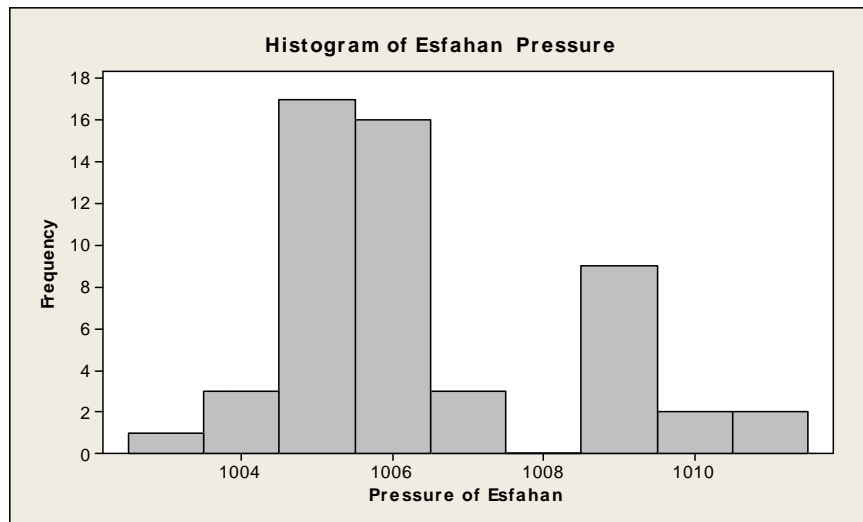
برای مثال دمای ماه ژانویه ایستگاه اصفهان در دوره 1951 – 2005 با نمودار میله ای زیر نشان داده شده است:



نمودار بالا دامنه نوسان دما را به خوبی نشان می دهد. بر این اساس نمود عنصر دما با عناصر دیگر اقلیمی با نمودار میله ای متفاوت می باشد.

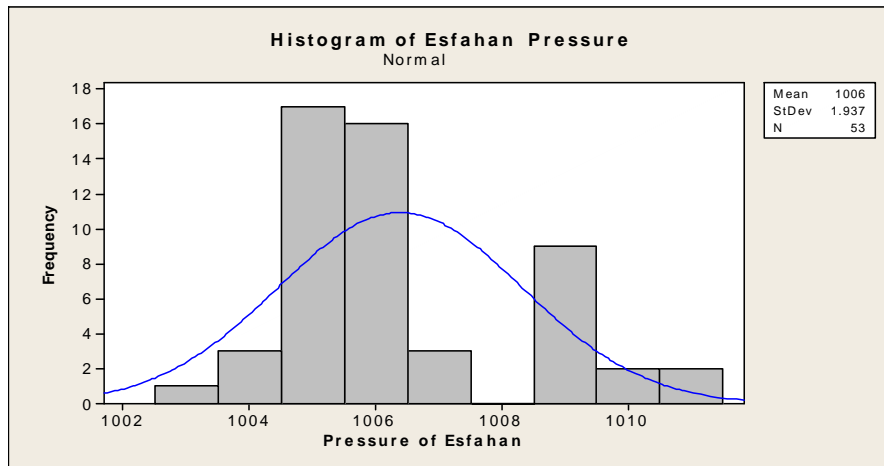
### 3- نمودار هیستوگرام

نمودار هیستوگرام<sup>1</sup> نوعی نمودار میله ای است که به کمک آن می توان داده ها را نمایش داد. نمودار مذکور شامل ستون های به هم پیوسته ای است که نحوه توزیع داده ها را نشان می دهد. در این نمودار عرض ستون برابر با فاصله طبقات است و ارتفاع ستون برابر با فراوانی همان داده ها می باشد و وسط هر ستون منطبق با حد وسط طبقات می باشد. در این نمودار شکل توزیع فراوانی داده ها ، مکان یا تمایل به مرکزی توزیع و پراکندگی یا گسترش توزیع داده های اقلیمی را می توان نشان داد . نمودار های هیستوگرام براساس نحوه کاربرد و ماهیت داده ها می توان به انواع هیستوگرام ساده و هیستوگرام بهنجاری که وضعیت سری را با شرایط برازش داده ها نشان می دهد ، طبقه بندی نمود. برای مثال سری فشار ایستگاه اصفهان (1951 تا 2003) را می توان با نمودار مذکور مشاهده نمود. یعنی :

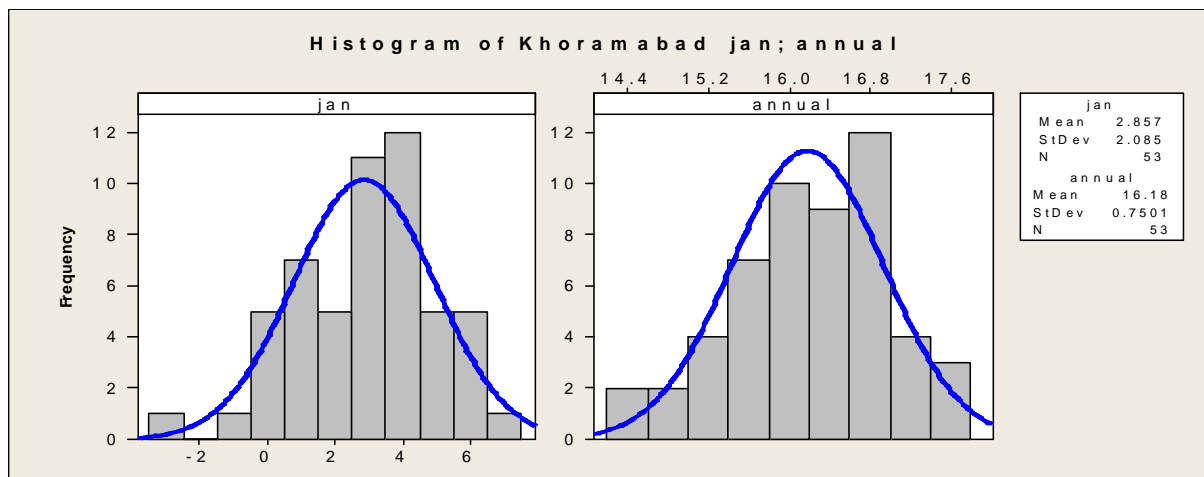


هیستوگرام بهنجاری سری فشار اصفهان :

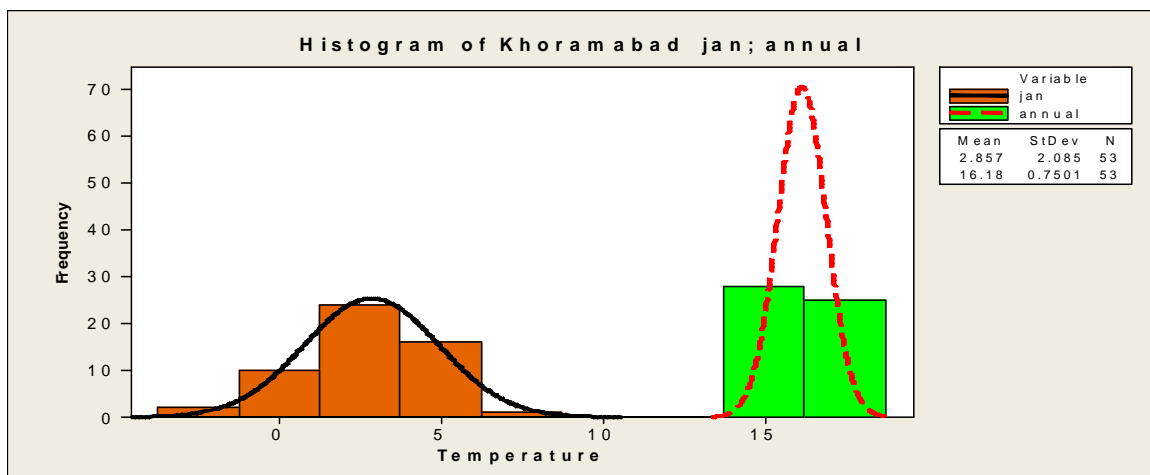
<sup>1</sup> - Histogram Chart



چنانچه داده های اقلیمی را بخواهیم با نمودار هیستوگرام با هم مقایسه کنیم ، می توان از نمودار هیستوگرام چند گانه یا پانلی استفاده کرد. برای مثال اگر بخواهیم دمای ماه ژانویه و میانگین سالانه ایستگاه خرم آباد (1951 تا 2003) را بررسی کنیم می توان نمودار نمودار مذکور به صورت زیر است. یعنی :



یا

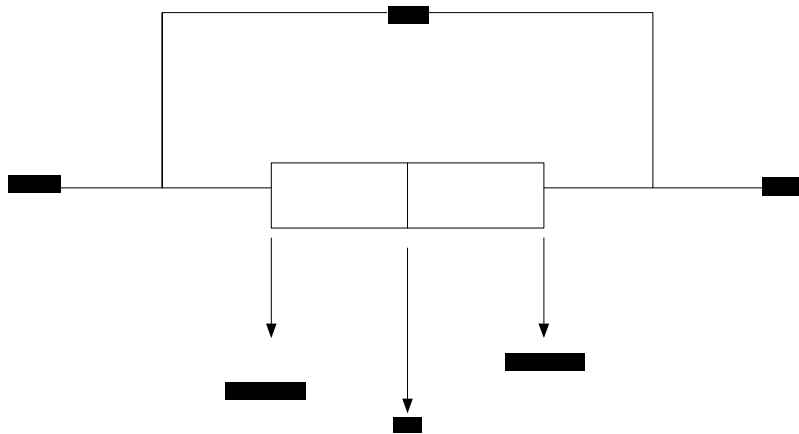




بسته به ماهیت کاربردی موضوع و نوع شاخص می توان نمودار هیستوگرام را برای تحلیل های اقلیمی به شیوه های مختلفی نمایش داد.

#### 4- نمودار جعبه ای

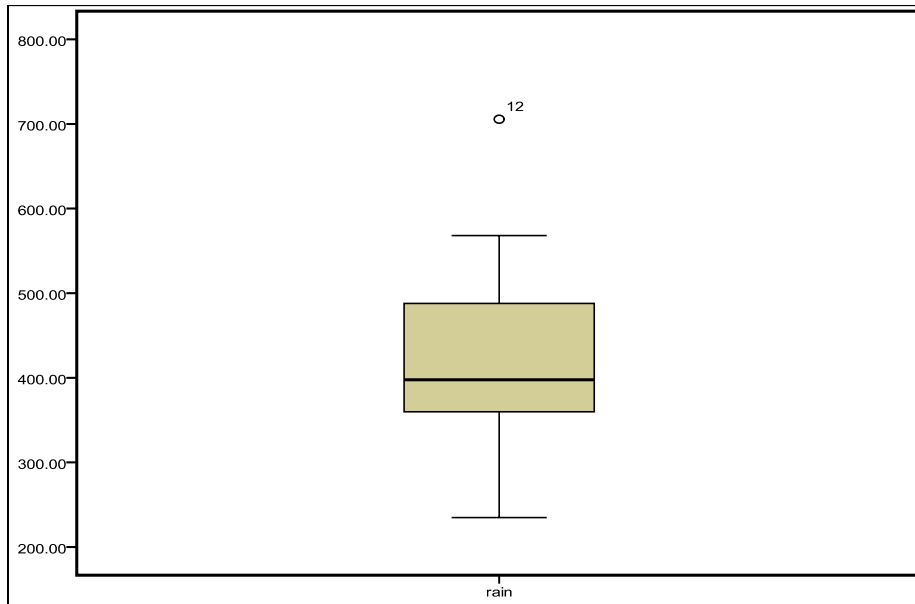
نمودار جعبه ای<sup>1</sup> برای خلاصه نمودن مجموعه بزرگی از داده ها به کار برده می شود. از طریق این نمودار می توان به تقارن و عدم تقارن داده ها ، پرت و انتهای ، دامنه میان چارکی ، میانه ، تغییرپذیری داده ها ، حداقل و حداکثر داده ها در تحلیل های اقلیمی پی برد. در این نمودار مستطیل نشان دهنده دامنه میان چارکی ، یعنی تفاضل بین چارک اول و سوم است . خط داخل مستطیل نشان دهنده میانه (محل تجمع داده ها) است . اگر خط میانه در وسط مستطیل باشد داده ها در حالت تقارن یا قرینه هستند و اگر خط میانه پایین مستطیل باشد داده ها دارای چولگی مثبت هستند و اگر خط میانه بالای مستطیل باشد داده ها دارای چولگی منفی اند. نقاط دو طرف مستطیل حداقل و حداکثر داده ها را نشان می دهند. در این نمودار نقاط پرت با علامت دایره نشان داده می شود. مقادیر انتهایی هم با علامت ستاره نشان داده می شود. این نمودار را هم نمودار چند خطی می گویند . در این نمودار اضلاع پایین و بالای جعبه یا مستطیل بر چارک اول و سوم دلالت دارند. خط عمود داخل مستطیل خط ویشکر<sup>2</sup> گفته می شود نقاط خارج خط مذکور نقاط پرت می باشند. یعنی :



به هر حال نمودار جعبه ای یکی از نمودار های بسیار مهمی که می توان در تحلیل های اقلیمی از آن استفاده نمود . چون در این نمودار چارک ها ، میانه و شاخص های دیگر را نشان می دهد. برای مثال اگر سری بارش دزفول ( 1988 تا 2003 ) جهت بررسی در دست باشد نمودار جعبه ای آن به شکل زیر می باشد:

Box Chart -<sup>1</sup>

Whisker -<sup>2</sup>



یا



## 5- نمودار چند ضلعی

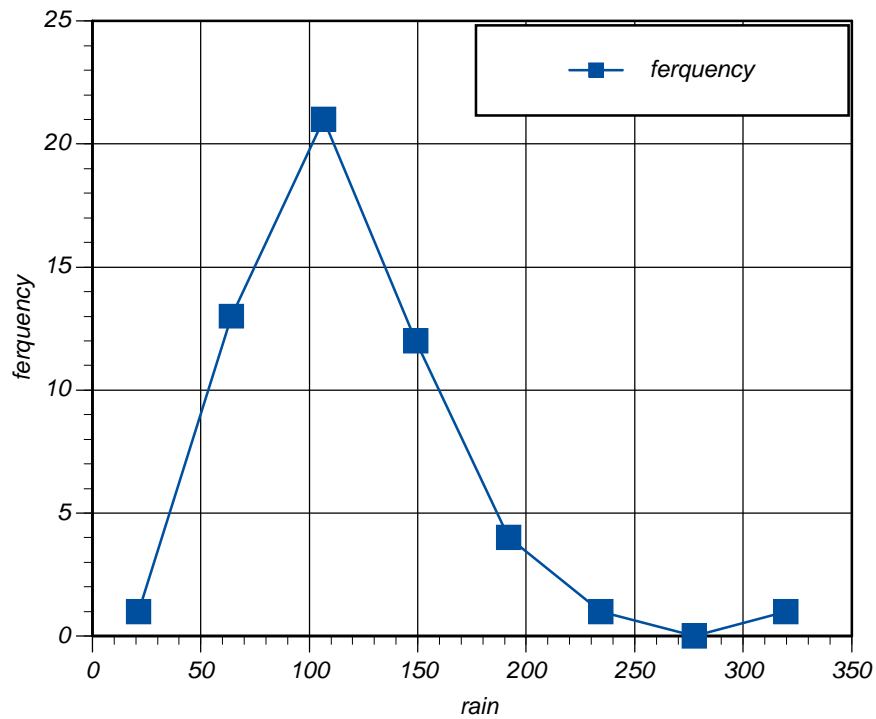
نمودار چند ضلعی<sup>1</sup> از تعداد نقاطی که توسط خطوط راست به یکدیگر وصل می شوند، تشکیل می شود. برای ترسیم این نمودار باید محورهای مختصات را رسم نمود و در محور افقی مقادیر متغیر (اگر سری طبقه بندی شده باشد حد میانی آنها در نظر گرفته می شود) طبقات قرار داده می شود و در محور عمودی مقادیر فراوانی داده ها تنظیم می شود و در نهایت اولین طبقه را در نظر گرفته و نمودار ترسیم می شود. اگر سری بارش اصفهان به صورت جدول زیر در دست باشد، نمودار چند ضلعی آن سری را ترسیم نمایید:

طبقات	حد میانه طبقات	فراوانی مطلق
0-42/64	21/32	1
42/64-85/28	63/96	13
127/92 85/28	106/6	21
170/56 127/92	149/24	12
-213/2 170/56	191/88	4
-255/84 213/2	234/52	1
298/48 255/84	277/16	0
341/12	319/8	1

---

 1 - Polygon Chart

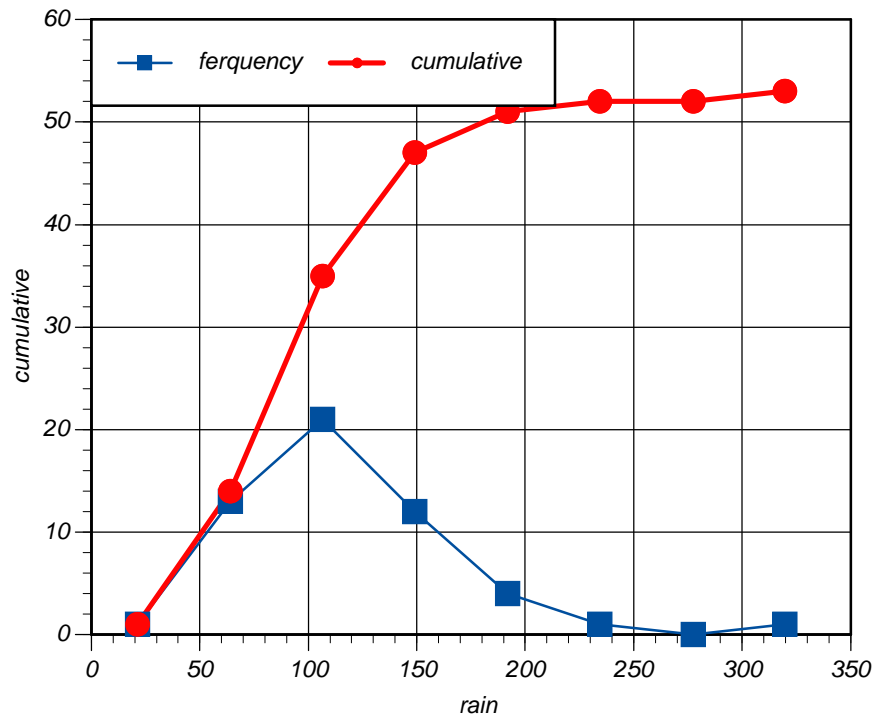
298/48		
جمع		53



6- نمودار تراکمی

نمودار تراکمی یا اجیو<sup>1</sup> برای نمایش سری فراوانی تراکمی داده های اقلیمی قابل استفاده است. برای ترسیم این نمودار محور عمودی را با فراوانی تراکمی یا درصد تراکمی داده ها تنظیم می شود و در محور افقی نمودار مقادیر متغیر ( برای داده های طبقه بندی نشده) و حدود طبقات ( برای داده های طبقه بندی شده ) تنظیم می شود. اگر سری بارش اصفهان در دست باشد. نمودار تراکمی آن به صورت زیر می باشد:

<sup>1</sup> Ogive Chart



7- نمودار شاخه و برگ

نمودار شاخه و برگ<sup>1</sup> برای مجموعه ای از داده ها محدود قابل استفاده است . برای ترسیم این نمودار باید سری را به دو قسمت تقسیم نمود . تعداد مقادیر مربوط به برگ ها به اندازه تعداد فراوانی ها است . برای جدا نمودن شاخه و برگ از یکدیگر خطی عمودی در سمت چپ برگ و سمت راست شاخه ، بین شاخه ها و برگ ها کشیده می شود. اگر سری میانگین سالانه اصفهان در دست باشد نمودار شاخه و برگ آن سری را نمایش دهید:

*tem.esfahan Stem-and-Leaf Plot*

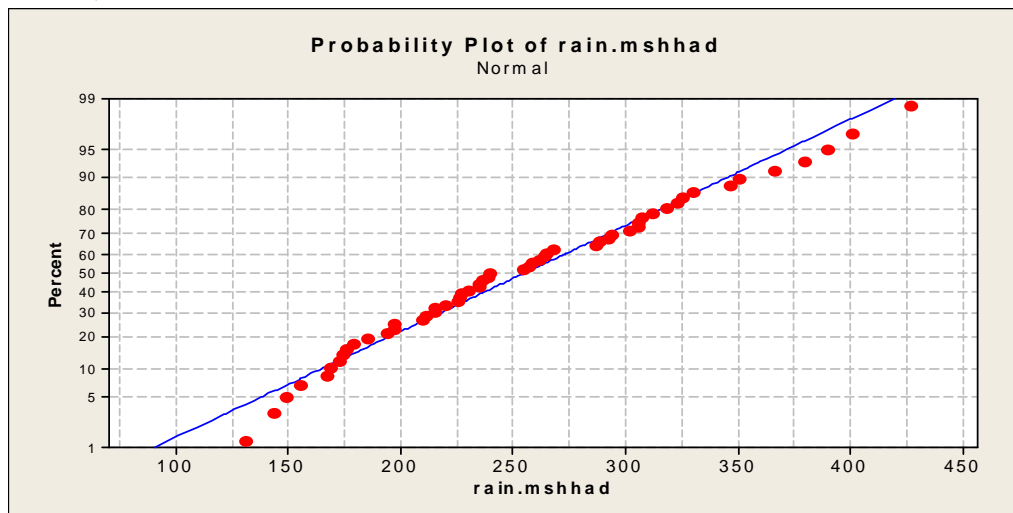
**Frequency Stem & Leaf**

1.00 Extremes (= < 21.1)  
 2.00 21 . 79  
 4.00 22 . 0144  
 8.00 22 . 66777889  
 16.00 23 . 0000000111124444  
 6.00 23 . 568899  
 2.00 24 . 12

Stem width: 1.00  
 Each leaf: 1 case(s)

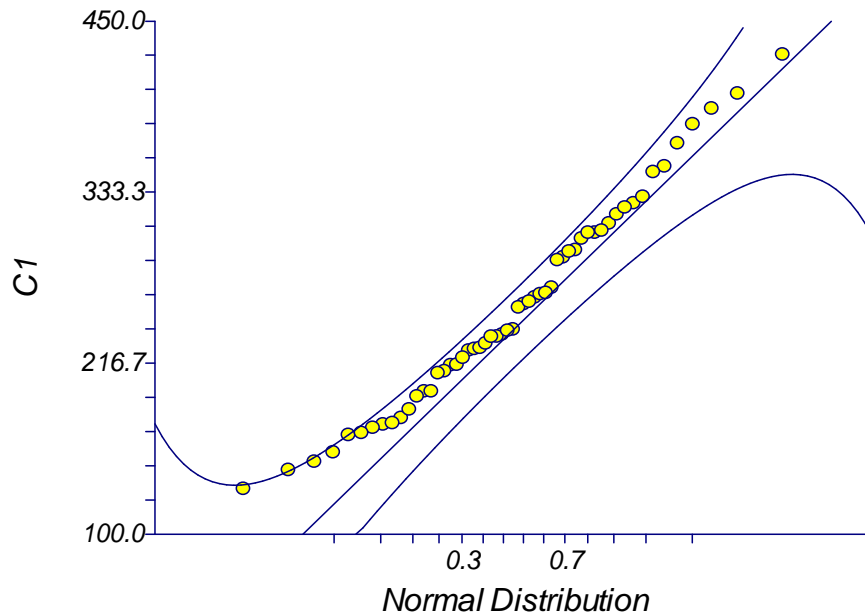
## 8- نمودارهای توزیع بهنجاری

این نمودار اهمیت زیادی در تحقیقات اقلیمی مخصوصاً برای سنجش سری دارد. بر این اساس جهت بررسی بهنجاری یا نرمال بودن سری ها یا متغیرهای اقلیمی از این نمودار باید استفاده کرد. در این نمودار مقادیر مشاهده شده را بر روی محور افقی نمایش داده می شود و مقادیر پیش بینی شده آن از یک توزیع نرمال معیار در محور عمودی نشان می دهند. این نمودار توزیع سری های اقلیمی در مقابل توزیع بهنجاری یا نرمال نشان می دهد. اگر سری به خط مبناء که معمولاً خط مستقیمی در قالب های مختلف می باشد، نزدیک تر باشند آن سری از شرایط بهنجاری برخوردار است. برای مثال اگر بارش ایستگاه مشهد (1951 تا 2005) در دست باشد نمودار بهنجاری سری مذکور را ترسیم نمایید:



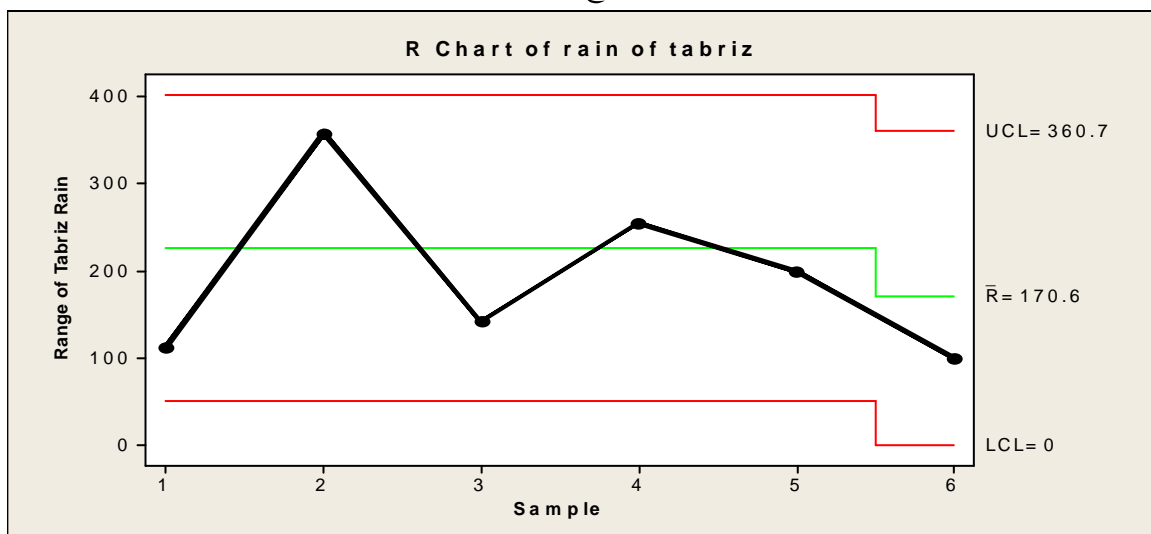
یا

### Rain of Mashhad

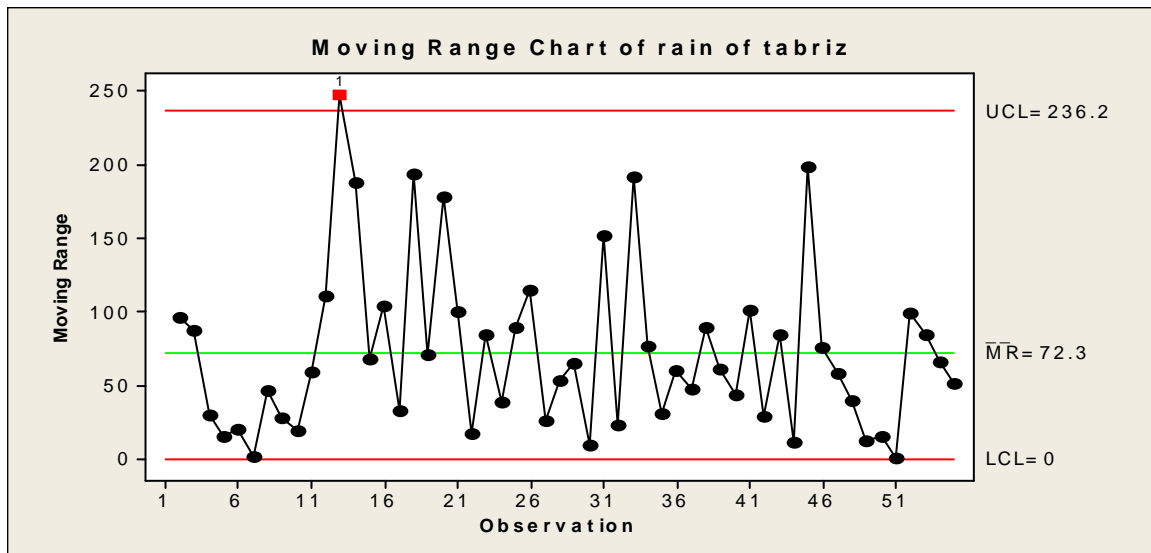


### 9- نمودار های کنترل :

نمودارهایی در تحقیقات اقلیمی برای شناسایی عوامل مؤثر بر شرایط اقلیمی مورد استفاده قرار می گیرند نمودار های کنترلی است . وقتی متغیر های اقلیمی روی شاخص های مختلف مانند شاخص های آماری مورد بررسی قرار می گیرند ، معمولاً کنترل متغیرها مورد نظر می باشد. در تحقیقات آماری اقلیم بررسی متغیر ها یا سری در برابر میانگین ، واریانس ، انحراف معیار و آماره های دیگر از نمودار های کنترلی استفاده می شود . مطالعه تغییرپذیری متغیر ها و سری های اقلیمی با استفاده از نمودار های کنترلی اهمیت زیادی دارند. فرض کنید هدف بررسی سری بارش تبریز در برابر دامنه بارش آن سری باشد ، نمودار زیر مؤید موضوع مذکور می باشد:



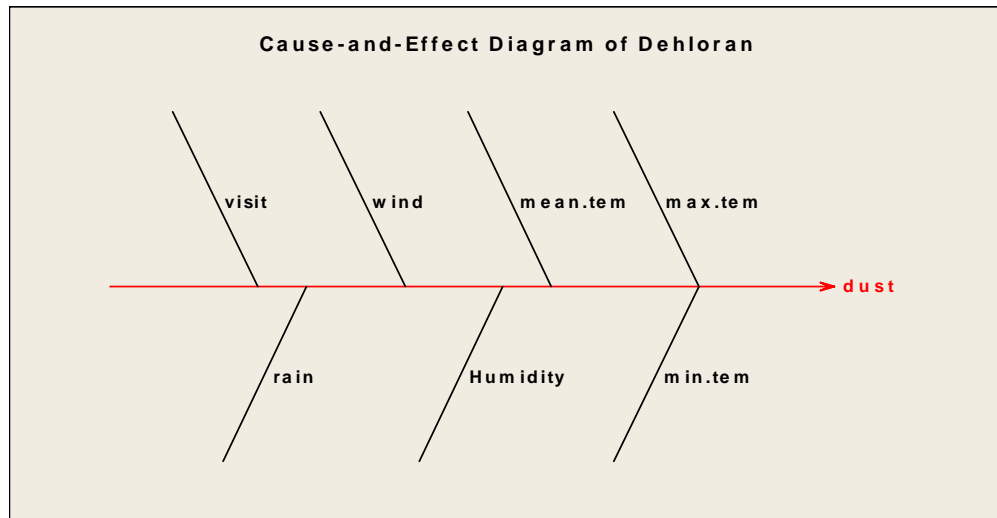
یا



### 10 نمودار علت و معلولی

نمودار علت و معلولی<sup>1</sup> یا نمودار استخوان ماهی علت های بالقوه یک موضوع را رسم می نماید. لذا یک موضوع یا مشکل (معلول) در سمت راست و فهرستی از علت سمت چپ در یک ساختار درختی به نمایش در می آید. نمودار علت و معلولی به انواع ساده و مرکب قابل تقسیم می باشد. اگر هدف محقق بررسی علت باد های گرد و غبار در شهرستان دهلران در استان ایلام (1986 تا 2005) باشد، داده هایی که در این تحلیل در نظر گرفته شده است دمای حداقل، حداکثر و میانگین، رطوبت نسبی، باران و میدان دید می باشند، نمودار علت و معلولی آن به صورت زیر می باشد:





### 11 نمودار پارتو

نمودار پارتو<sup>1</sup> یک نمودار میله ای است که علل تغییرات به وجود آمده را با فراوانی آن مقایسه می کند، در نمودار پارتو از زوایای مختلف می توان به یک موضوع نگاه کرد و سپس به حل آن پرداخت .

چگونگی رسم نمودار پارتو:

برای رسم نمودار پارتو انجام مراحل زیر ضروری است :

مرحله اول : ابتدا مشخص کنید که کدام عنصر اقلیمی مورد بررسی باید قرارگیرد و چطور اطلاعات و داده ها را باید جمع آوری کرد.  
الف تصمیم بگیرید کدام موضوع اقلیمی باید بررسی شود .

ب - مشخص کنید که چه داده هایی لازم است جمع آوری شود و چطور آنها را طبقه بندی کنید.

ج روش جمع آوری داده ها و دوره زمانی جمع آوری آن را مشخص کنید.

مرحله دوم : یک برگه کنترل "برگه ثبت اطلاعات" مناسب طراحی کنید.

مرحله سوم : بعد از علامت گذاری و ثبت موارد مشاهده شده در برگه کنترل فراوانی آنها را به دست آورید.

مرحله چهارم : یک جدول توزیع فراوانی شامل تمام موارد فهرست شده ، ستونهای فراوانی ، فراوانی تجمعی ، درصد فراوانی و درصد فراوانی تجمعی تهیه کنید.

مرحله پنجم : جدول توزیع فراوانی را برحسب تعداد به ترتیب غیرنزولی مرتب کنید.

مرحله ششم : یک محور افقی و دو محور عمودی رسم کنید:

الف - محور افقی : این محور را به تعدادی فواصل یکسان شامل تمام موارد، تقسیم بندی کنید.

ب - محور عمودی سمت چپ : این محور را از صفر تا "n جمع کل داده ها" مدرج کنید.

ج - محور عمودی سمت راست : این محور را از صفر تا "100 درصد کل" مدرج کنید.

<sup>1</sup> Pareto Chart

مرحله هفتم: یک نمودار ستونی رسم کنید.

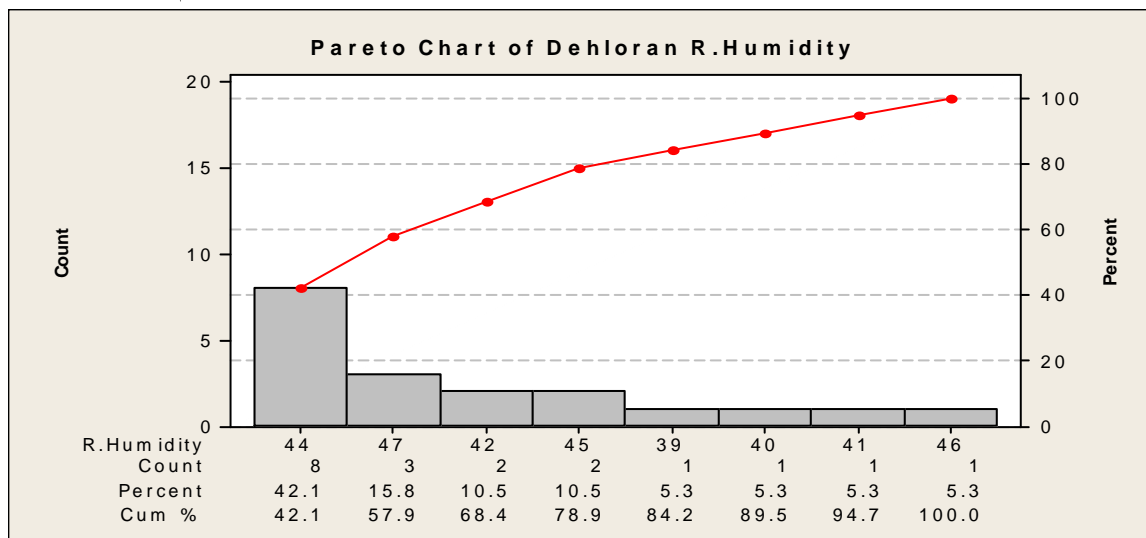
مرحله هشتم: منحنی فراوانی تجمعی، منحنی پارتو را رسم کنید، برای این کار ارزشهای تجمعی را در بالای سمت راست ستون مربوط به هر طبقه با نقطه ای مشخص کرده و سپس این نقاط را به یکدیگر وصل کنید.

مرحله نهم: تمام اطلاعات ضروری را روی نمودار ثبت کنید:

الف - اطلاعات مربوط به نمودار شامل عنوان، واحد، نام رسم کننده نمودار، موارد بامعنی و...

ب - اطلاعات مربوط به داده ها شامل دوره زمانی، محل جمع آوری داده ها، موضوع، جمع کل داده ها و...

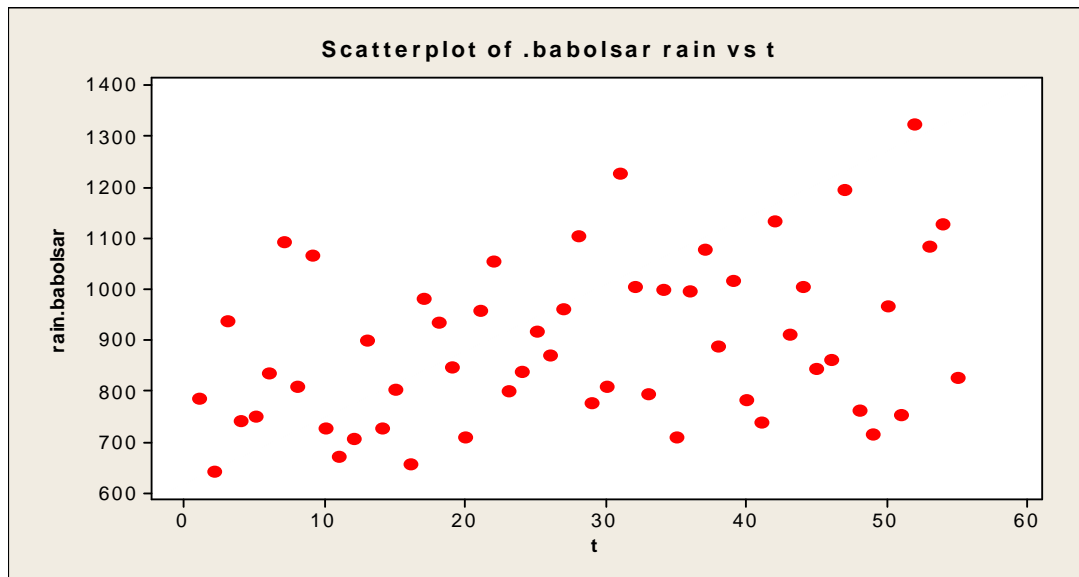
برای نمونه اگر سری رطوبت نسبی ایستگاه دهلران (1987 تا 2005) در دست باشد، نمودار پارتوی آن را ترسیم نمایید:



## 12 نمودار پراکنش

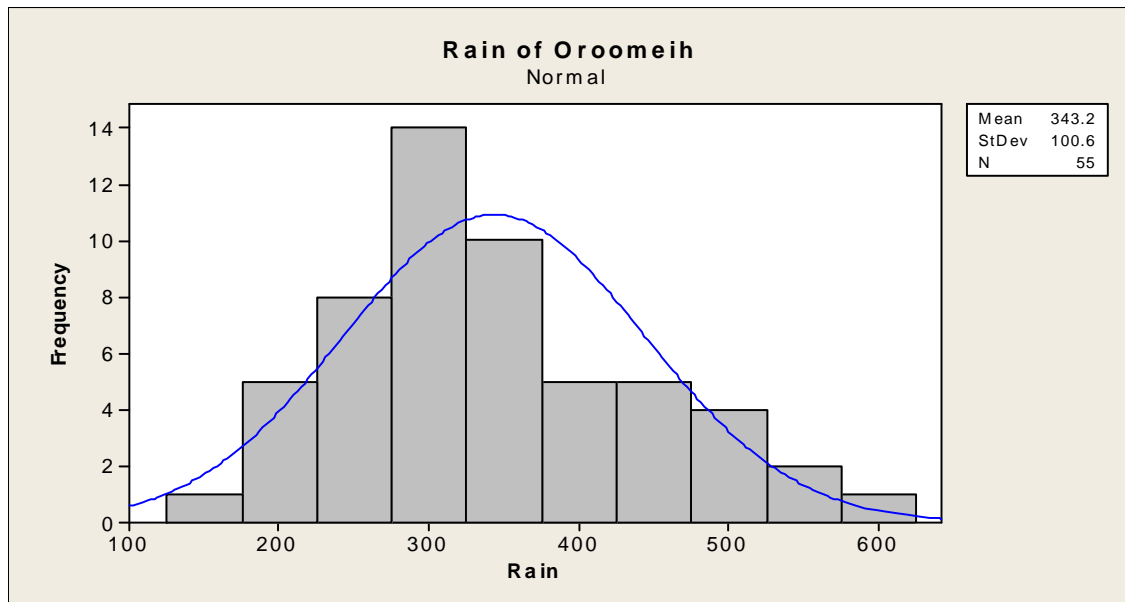
در مطالعه رابطه بین دو متغیر اقلیمی، اولین قدم رسم داده ها به صورت نقاطی بر روی یک صفحه نمودار است. شکل حاصل که نمودار پراکنش<sup>1</sup> نامیده می شود، چگونگی خوشه ای شدن نقاط در اطراف یک خط مستقیم یا یک نوع منحنی مشخص را نشان می دهد و نیز برداشتی عینی از میزان پراکندگی داده ها پیرامون خط یا منحنی را فراهم می کند در بیشتر موقعیتها رابطه نظری پیشین معلومی وجود ندارد که آن را به کار ببریم، بنابراین، اطلاع منعکس در نمودار پراکنش برای جستجو یک مدل ریاضی مناسب مفید است. رسم نمودار پراکنش در تحلیل رگرسیون ساده اهمیت زیادی دارد و بین دو مجموعه داده استفاده می شود تا نوع رابطه بین دو داده را حدس بزنیم در حقیقت حدس می زند که یک رابطه به شکل یک خط بین دو متغیر وجود دارد و سپس به جمع آوری اطلاعات کمی از دو متغیر می پردازد و این داده ها را به صورت نقاطی در یک نمودار دو بعدی رسم می کند. این نمودار که به آن نمودار پراکندگی گفته می شود نقش بسیار مهمی را در تحلیل های رگرسیونی و نمایش ارتباط بین متغیرها ایفا می کند. در صورتی که نمودار نشان دهنده این باشد که داده ها تقریباً (نه لزوماً دقیق) در امتداد یک خط مستقیم پراکنده شده اند، حدس تحلیل گر تأیید شده است. اگر سری بارش بابلسر (1951 تا 2005) در دست باشد، نمودار پراکنش آن به صورت زیر می باشد:

Scatter Chart -<sup>1</sup>



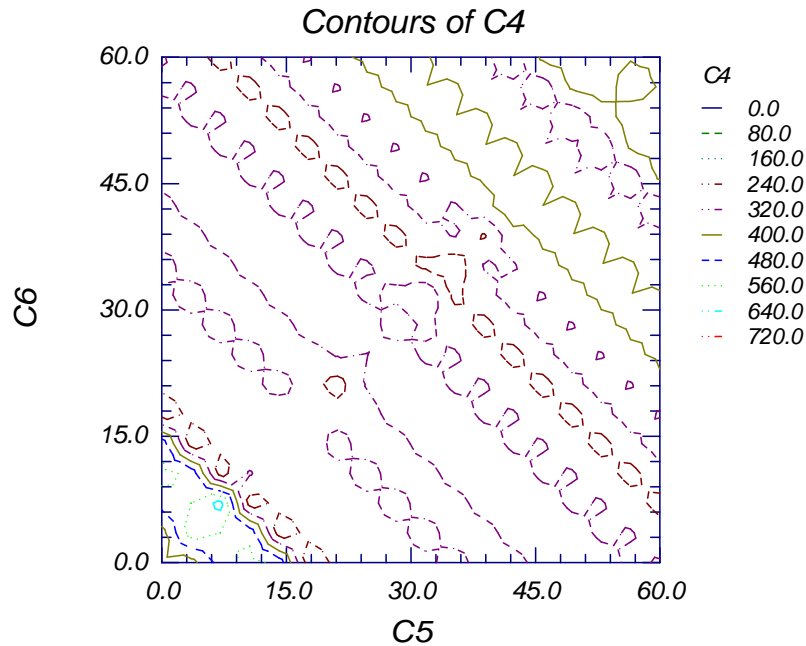
### 13 نمودار تابع چگالی احتمال

نمودار تابع چگالی احتمال<sup>1</sup> در تعیین شرایط بهنجاری (نرمال) و نابهنجاری (غیرنرمال) بودن توزیع داده های اقلیمی بکار می رود. اگر نمودار مذکور متقارن (حالت زنگوله ای شکل) باشد، سری اقلیمی دارای توزیع بهنجاری دارد و بر عکس از شرایط نابهنجاری برخوردار می باشد. برای ترسیم این نمودار سری یا داده های اقلیمی را بر روی محور عمومی تنظیم می شود. برای ترسیم نمودار مذکور به طور ساده فقط توزیع نشان داده می شود و به صورت مرکب علاوه بر توزیع نمود سری به صورت میله ای یا هیستوگرام نمایش داده می شود. برای مثال اگر سری بارش ارومیه (1951 – 2005) در دست باشد، نمودار تابع چگالی آن به صورت زیر می باشد:



#### 14 نمودار خطوط هم ارزش

نمودار خطوط هم ارزش<sup>1</sup> در اقلیم معمولاً دلالت دارد بر نقاطی که مقدار هر عنصر با هم برابر باشند. خطی که نقاط دارای ارزش یکسان را به همدیگر وصل می کند خط هم ارزش نامیده می شود. خطوط هم ارزش در اقلیم مانند خطوط همدم، همبارش، همفشار و... می باشند. برای ترسیم نمودار خطوط هم ارزش معمولاً مقادیر محورهای  $X$ ،  $Y$  و  $Z$  لازم است. محورهای مذکور به ترتیب دلالت بر طول، عرض و ارتفاع (عنصر) اقلیمی دارند. نمایش عناصر اقلیمی با خطوط هم ارزش از اهمیت زیادی در تحقیقات اقلیمی دارد. برای نمونه اگر سری بارش ایستگاه شیراز (1951 تا 2005) در دست باشد نمودار همبارش آن را ترسیم نمایید:



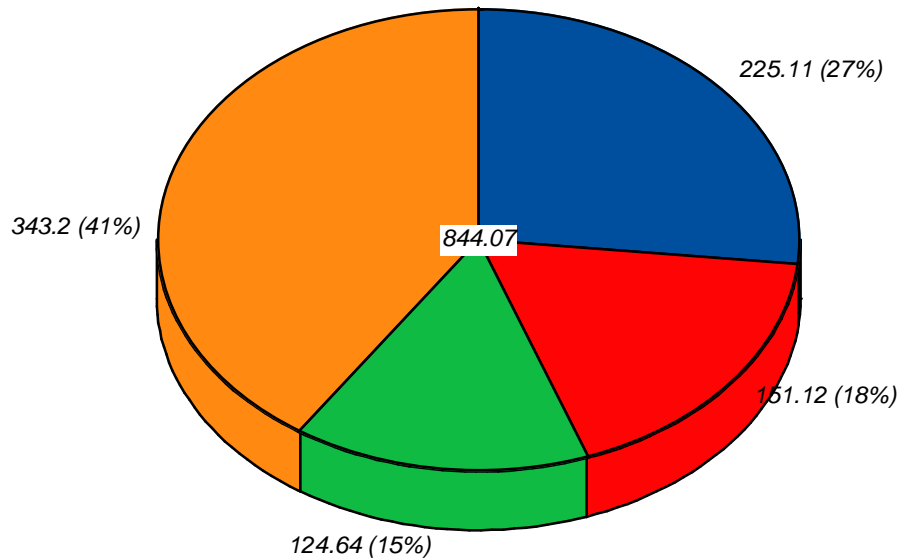
15 نمودار دایره ای :

نمودار دایره ای<sup>1</sup> تصویری از توزیع فراوانی یک متغیر اقلیمی که غالباً "کیفی می باشد را به نمایش می گذارد. برای ترسیم نمودار دایره ای، دایره ای بر روی داده های طبقات ترسیم می شود. این نمودار قطاع های دایره ای را متناسب با مساحت آن برای متغیر های اقلیمی نمایش می دهد. برای مثال اگر میانگین بارش ایستگاه های مشهد، کرمان، اصفهان و ارومیه در قالب منطقه ای به ترتیب 225/11، 151/12، 124/64 و 343/2 میلی متر باشد مقدار زاویه های قطاع های نمودار دایره ای آنها به شکل زیر محاسبه می شود:

مجموع میانگین های بارش ایستگاه ها برابر با 844/07 می باشد.

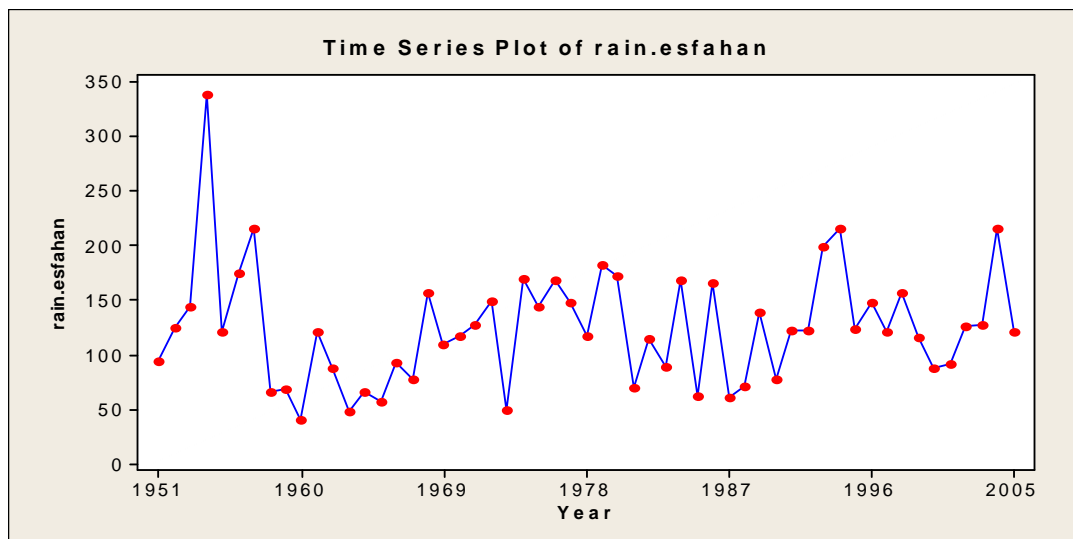
$$\frac{225.11}{844.07} \times 360 = 96.01$$

والی آخر

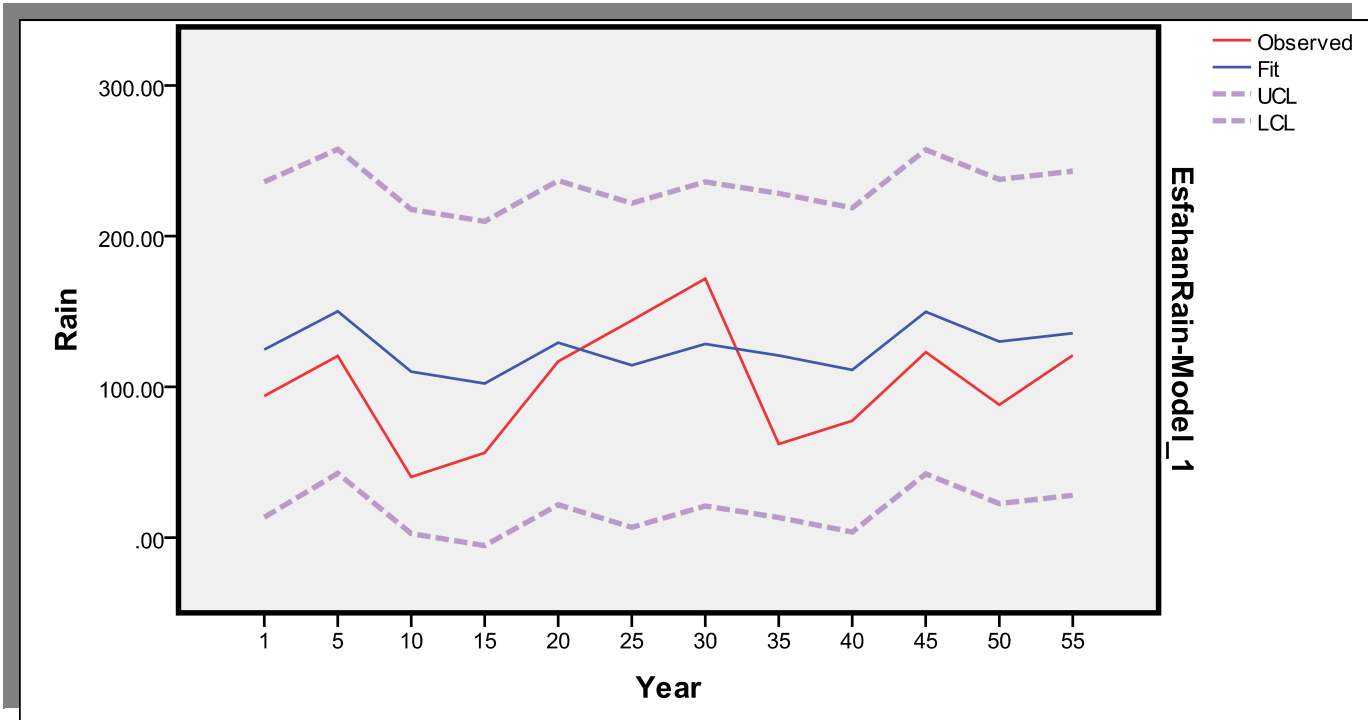


### 16 نمودار سری زمانی

در تحقیقات اقلیمی وقتی سری های زمانی عناصر اقلیمی تهیه شود و عناصر اقلیمی را در قالب زمان بررسی می شود در واقع تحلیل از سری زمانی عناصر اقلیمی را فراهم شده است. در این راستا نمایش تغییرات عناصر اقلیمی در قالب مدل های روند ، فصلی ، دوره ای و تصادفی دلالت بر نمودار های سری زمانی دارد. اگر سری بارش ایستگاه اصفهان (1951 تا 2005) در دست باشد نمونه ای نمودار های سری زمانی آن را ترسیم نمایید:

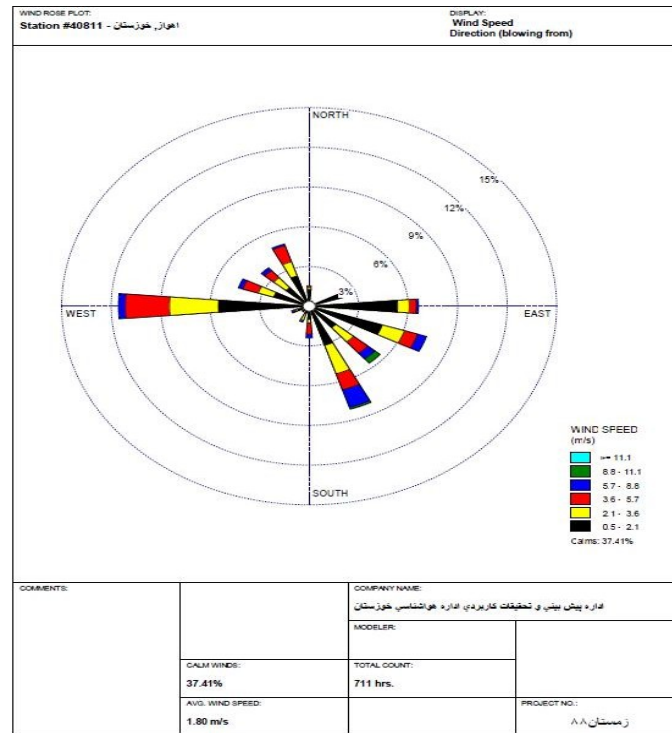


یا



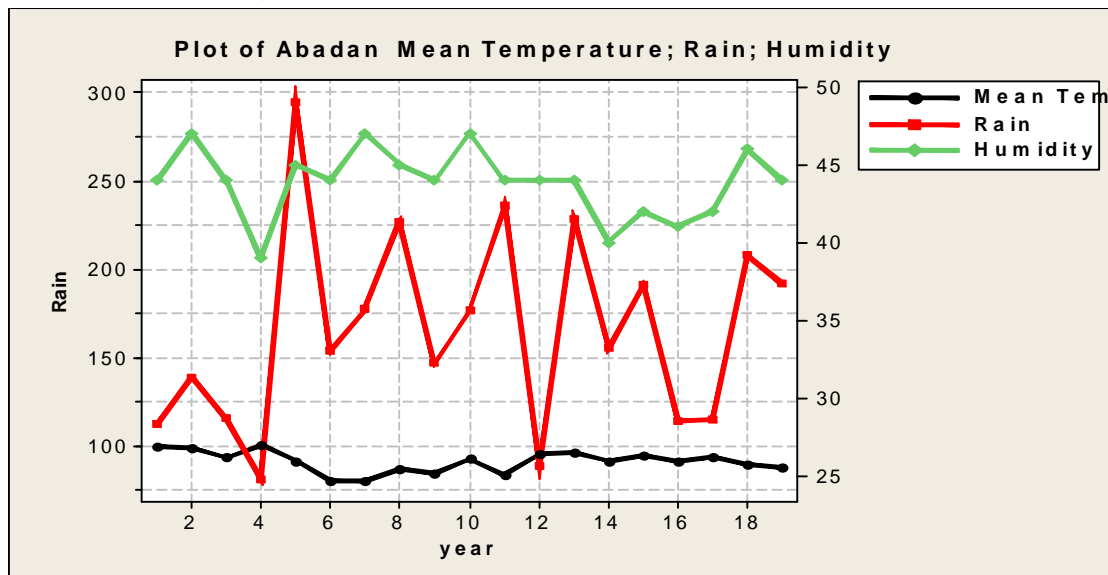
### 17 نمودار برداری

نمایش کمیت های برداری اقلیم مانند باد با استفاده از نمودار های برداری انجام می گیرد . باد به عنوان عنصر اقلیمی قالب برداری دارد یعنی از دو مؤلفه سرعت و جهت برخوردار است. بر این اساس نمایش عنصر باد با نمودار برداری که مشخص کننده سرعت و جهت باد باشد از اهمیت زیادی برخوردار است . برای نمونه گلباد ایستگاه اهواز نمونه ای از نمودار های برداری می باشد .



### 18 کلیماتوگرام ها

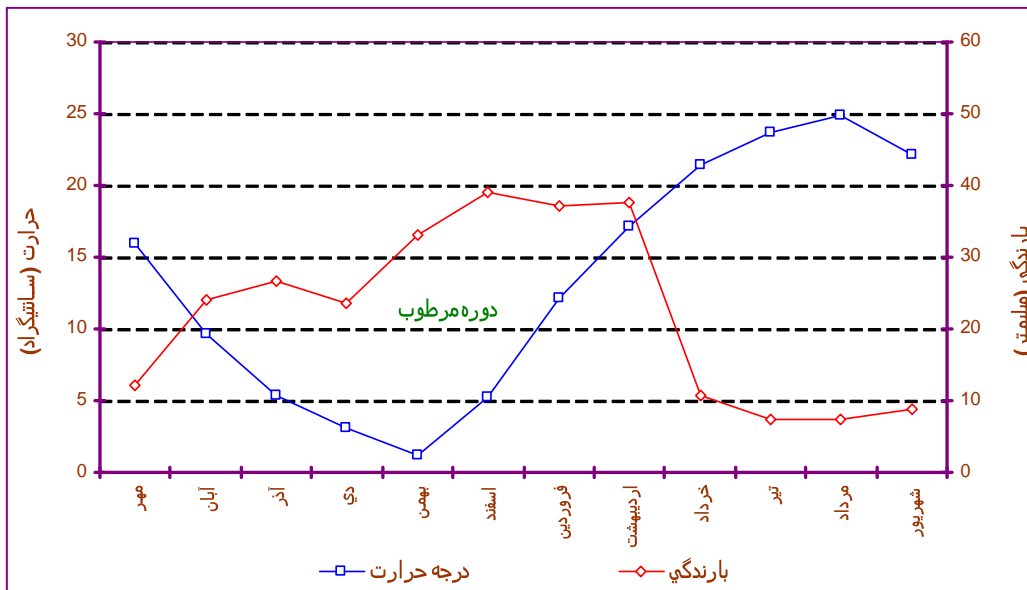
کلیماتوگرام ها از جمله نمودار هایی است که عناصر مختلف اقلیمی مانند دما و بارش را می توان برای ماههای مختلف سال در هر ایستگاه نشان داد. نمود این نمودار به این صورت است که عناصر دما و بارش مربوط به هرماه را به صورت متناظر روی محورهای افقی و عمودی نشان می دهد. دامنه نمودار دلالت بر میزان دامنه تغییرات عناصر اقلیمی دارد. دامنه نمودار بالا نشانگر دامنه شدید تغییرات عناصر اقلیمی است. برای مثال اگر سری های رطوبت نسبی، دما و بارش سالانه ایستگاه آبادان در دست باشد، نمودار کلیماتوگرام آن به صورت زیر می باشد:





## 19 نمودار آمبروترمیک

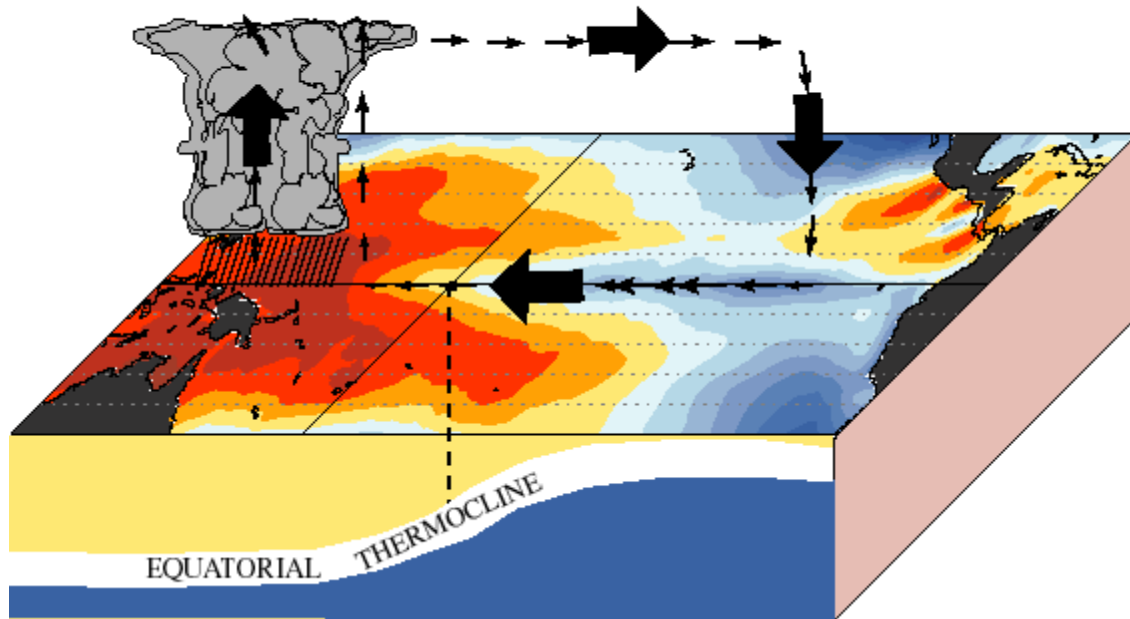
نمودار آمبروترمیک برای نشان دادن ماههای خشک در هر ایستگاه اقلیمی استفاده می شود. در این نمودار بر روی محور عمودی یک طرف بارش و طرف دیگر دما اختصاص داده می شود و بر روی محور افقی ماههای سال تعیین می شود. معمولاً برای ترسیم آن نمودار محور بارش را نسبت به محور دما دو برابر در نظر گرفته می شود. در نمودار ترسیم شده ماههایی که در آنها محورهای بارش و دما همدیگر قطع می کنند ماههای خشک مشخص می شود. برای مثال نمودار آمبروترمیک ایستگاه بجنورد به صورت زیر می باشد:



## 20 نمودار شماتیک

نمودار شماتیک در واقع طرح نمودی مراحل فرآیندی یک جریان در اقلیم شناسی می باشد. نمودارهای شماتیک دلالت بر نمودارهای مدل های اقلیمی دارند که مراحل فرآیندی آن مدل را نشان می دهند برای نمونه نمودار شماتیک لاینو را می توان به صورت زیر در نظر گرفت:

## December - February La Niña Conditions



### فصل سوم: آماره های توصیفی و تغییرپذیری آنها:

اقلیم شناسی ، علم سنجش ، تحلیل و پیش بینی شرایط جوی بلند مدت در یک مکان می باشد. در اقلیم شناسی تجزیه و تحلیل به سه شیوه تبیینی<sup>1</sup>، اکتشافی<sup>2</sup> و تأییدی<sup>3</sup> می باشد. در تحلیل تبیینی به چونی و چرایی و علت یابی شرایط و تغییرپذیری اقلیم و در تحلیل اکتشافی به شناسایی و پیش بینی شرایط اقلیم و در تحلیل تأییدی بر روند پیدایش شرایط اقلیمی تأکید دارد. در شیوه تبیینی توجه خود را به روند دوره ای و مرحله عناصر اقلیمی هر قلمرو در قالب توصیف و قیاس معطوف می کند و در شیوه اکتشافی بر فرآیندهای تکاملی سیستم های اقلیمی در گذشته ، حال و آینده تأکید دارد و در شیوه تأییدی علاوه بر شرایط تکامل سیستم های اقلیمی بر استنتاج شرایط حاکم اقلیمی بر هر قلمرو توجه می شود. در این شیوه علاوه بر مقایسه سیستم به فرض های مختلف توجه دارد. اگر بخواهند ویژگی های یک متغیر اقلیمی مانند دما را

1 - Explanation  
2 - exploration  
3 - confirmation

توصیف کنند باید به توصیف و تحلیل آن متغیر بپردازند. اولین قدم برای توصیف و تحلیل هر متغیر اقلیمی بررسی فراوانی ها و طبقه بندی آن می باشد. فراوانی ها در تحلیل های اقلیمی بر توزیع آنها در قالب های زمانی و مکانی تأکید دارد. توزیع فراوانی نشان می دهد که هر یک از نمره های ممکن در داخل توزیع (در قالب های زمانی و مکانی) توسط چند تا از آزمودنی ها بدست آمده است. برای بررسی فراوانی های متغیرهای اقلیمی باید با توجه به نوع و فراوانی آنها را در قالب های توزیع فراوانی ساده (طبقه بندی نشده) و طبقه بندی شده بررسی نمود. برای بررسی فراوانی متغیرهای اقلیمی در قالب مدل ساده باید انواع فراوانی های مطلق<sup>1</sup>، نسبی<sup>2</sup>، درصد فراوانی نسبی<sup>3</sup>، فراوانی تجمعی یا تراکمی<sup>4</sup>، درصد فراوانی تجمعی<sup>5</sup> اشاره نمود. فراوانی مطلق عبارت است از مقدار معین در مقیاس اعداد که فراوانی وقوع هر نمره در آن مشخص شده است. فراوانی نسبی در تحلیل های اقلیمی دلالت بر مقایسه فراوانی سری هایی از داده ها یا متغیرهای اقلیمی مختلف با حجم متفاوت دارد. توزیع فراوانی نسبی (f) در واقع فراوانی مطلق (F) یک متغیر اقلیمی (در قالب زمانی و مکانی) به صورت طبقه بندی شده یا نشده به تعداد کل فراوانی های موجود (N) را مشخص می کند، یعنی  $f = F / N$ . لذا فراوانی های نسبی همان نسبت ها هستند که مقدار آنها از صفر تا یک می باشد. بر این اساس توزیع فراوانی نسبی به علت مقایسه داده های اقلیمی از اهمیت زیادی برخوردار است. نسب درصدی فراوانی نسبی را درصد فراوانی نسبی می خوانند  $f = (F / N) \times 100$ . فراوانی تجمعی یا تراکمی، فراوانی است که تعداد داده های قرار گرفته شده در زیر نقطه معین در مقیاس معین را نشان می دهد. فراوانی تراکمی یک متغیر در تحلیل های اقلیمی نشانگر تعداد متغیرهایی که پایین تر از حد واقعی بالای آن متغیر است، از نکاتی که در این تحلیل های اقلیمی مورد توجه قرار می گیرد توزیع زمانی و مکانی فراوانی تراکمی یا تجمعی متغیرها است. در تحلیل های اقلیمی بهتر است جهت تفسیر داده ها یا متغیرهای اقلیمی توزیع فراوانی تراکمی را به نسبت های تراکمی یا درصد ها تبدیل نمود. نسبت تراکمی یا تجمعی عبارت است از نسبت فراوانی تراکمی به تعداد کل متغیرها  $f_r = (f / N)$ . درصد تراکمی یا تجمعی متغیرهای اقلیمی عبارت است از نسبت تراکمی یا تجمعی ضربدر 100، یعنی  $f_r = (f / N) \times 100$ . نسبت تراکمی، نسبت متغیرهایی را که مقدار آنها از یک حد معین کمتر است، مشخص می سازد. برای فهم بهتر به موقعیت های شماره 1 و 2 توجه شود.

موقعیت شماره 1:

- 
- Absolute Ferquency - 1
  - Relative Ferquency - 2
  - Relative Ferquency Percentage - 3
  - Cumulative Ferquency - 4
  - Cumulative Ferquency Percentage - 5

برای مثال اگر بارش ایستگاه اصفهان به صورت جدول زیر در دست باشد ، فراوانی های آن چگونه است؟

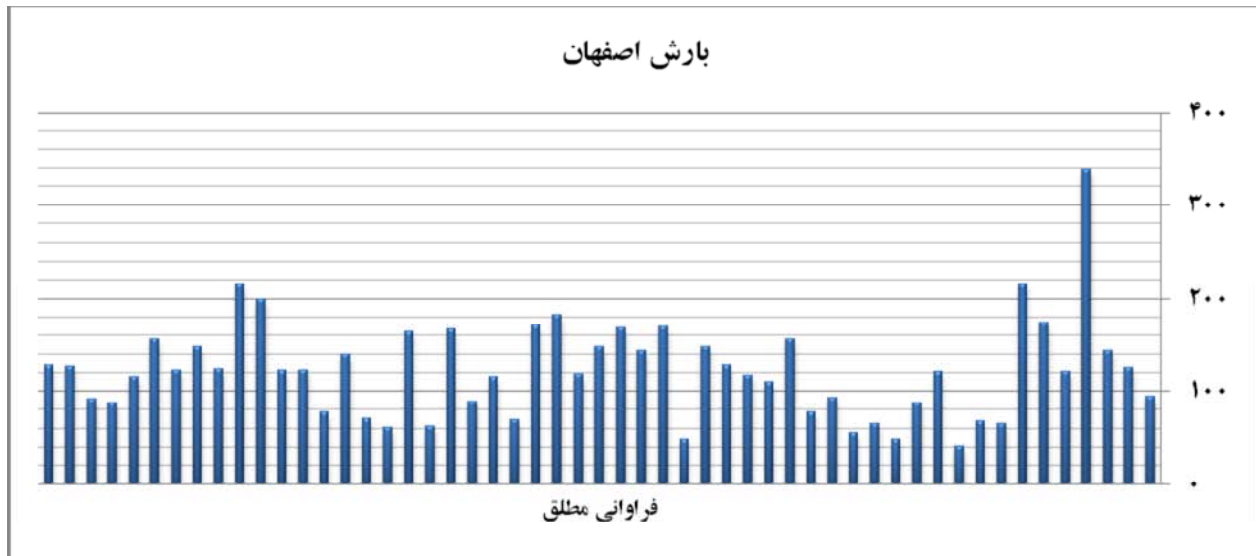
سال آماری	بارش	فراوانی مطلق (F)	فراوانی نسبی (f)	درصد فراوانی نسبی	فراوانی تجمعی (Fc)
1951	94.00	1	0/01887	1/89	1
1952	124.40	1	0/01887	1/89	2
1953	143.90	1	0/01887	1/89	3
1954	338.70	1	0/01887	1/89	4
1955	120.50	1	0/01887	1/89	5
1956	174.50	1	0/01887	1/89	6
1957	215.10	1	0/01887	1/89	7
1958	65.20	1	0/01887	1/89	8
1959	68.30	1	0/01887	1/89	9
1960	40.20	1	0/01887	1/89	10
1961	121.00	1	0/01887	1/89	11
1962	87.30	1	0/01887	1/89	12
1963	48.20	1	0/01887	1/89	13
1964	65.90	1	0/01887	1/89	14
1965	56.30	1	0/01887	1/89	15
1966	92.90	1	0/01887	1/89	16
1967	77.70	1	0/01887	1/89	17

1968	156.80	1	0/01887	1/89	18
1969	109.90	1	0/01887	1/89	19
1970	116.90	1	0/01887	1/89	20
1971	127.20	1	0/01887	1/89	21
1972	148.60	1	0/01887	1/89	22
1973.	49.00	1	0/01887	1/89	23
1974.	169.80	1	0/01887	1/89	24
1975	144.10	1	0/01887	1/89	25
1976	168.30	1	0/01887	1/89	26
1977	148.30	1	0/01887	1/89	27
1978	117.30	1	0/01887	1/89	28
1979	182.30	1	0/01887	1/89	29
1980	171.80	1	0/01887	1/89	30
1981	69.20	1	0/01887	1/89	31
1982	114.60	1	0/01887	1/89	32
1983	88.50	1	0/01887	1/89	33
1984	167.70	1	0/01887	1/89	34
1985	62.10	1	0/01887	1/89	35
1986	165.10	1	0/01887	1/89	36
1987	61.00	1	0/01887	1/89	37

1988	70.90	1	0/01887	1/89	38
1989	139.30	1	0/01887	1/89	39
1990	77.50	1	0/01887	1/89	40
1991	122.40	1	0/01887	1/89	41
1992	122.60	1	0/01887	1/89	42
1993	198.80	1	0/01887	1/89	43
1994	75.90	1	0/01887	1/89	44
1995	123.00	1	0/01887	1/89	45
1996	147.90	1	0/01887	1/89	46
1997	121.50	1	0/01887	1/89	47
1998	157.00	1	0/01887	1/89	48
1999	115.30	1	0/01887	1/89	49
2000	88.10	1	0/01887	1/89	50
2001	91.20	1	0/01887	1/89	51
2002	126.60	1	0/01887	1/89	52
2003	127.80	1	0/01887	1/89	53
جمع		53	53	100	

نمود فراوانی های بارش ایستگاه اصفهان :

نمود فراوانی مطلق بارش ایستگاه اصفهان :



برای نمایش بهتر فراوانی های بارش ایستگاه اصفهان به علت طولانی بودن لازم است سری را به صورت طبقه بندی شده نشان داد . برای تحلیل و توصیف فراوانی های طبقه بندی شده هر متغیر اقلیمی لازم است به مراحل زیر توجه نمود :

- مرتب کردن فراوانی متغیر اقلیمی (از بزرگ یا از کوچک).
- محاسبه دامنه تغییرات (تفاضل بین بزرگترین و کوچکترین متغیر به اضافه یک) .
- تعیین تعداد طبقات .
- تعیین فاصله طبقات
- تنظیم طبقات با در نظر گرفتن کرانه ها
- محاسبه نقطه میانه طبقات

برای انجام مراحل مذکور بعد از مرتب کردن متغیرهای اقلیمی باید ابتدا دامنه تغییرات سری را محاسبه نمود. برای محاسبه دامنه تغییرات می توان از فرمول زیر استفاده کرد:

$$R = (\text{Max}_v - \text{min}_v) + 1$$

دامنه تغییرات دلالت بر تفاضل بین حداکثر (Max) و حداقل (Min) متغیرها دارد . بعد محاسبه دامنه تغییرات باید تعداد طبقات را با یکی از فرمول ها زیر بدست آورد.

$$1k = 1 + 3.332 \text{Log}n$$

یا

$$k = \sqrt{N}$$

یا

$$^1k = 2.5 \times N^{1/4}$$

بعد از آن فاصله طبقات (C = R / K) باید محاسبه نمود. با تعیین فاصله طبقات متغیرهای اقلیمی باید با توجه به کرانه های پایین و بالا تنظیم

نماید و سپس نقطه میانه طبقات را با توجه به کرانه پایین و بالا ( $M_p = \frac{a_i + b_i}{2}$ ) محاسبه نمود. نقطه قابل اشاره این است که چون در

مطالعات اقلیمی باید همیشه سری طولانی را برای تحقیق خود انتخاب نماید، لازم است برای بررسی سریع تر، سری متغیرهای اقلیمی را در به

صورت طبقات محدودی (بدون در نظر گرفتن کرانه ها) طبقه بندی نمود. بر این اساس مرتب کردن سری ها در مطالعات اقلیمی اهمیت زیادی

دارد. هرچند که امروز با استفاده از نرم افزارهای متعدد آماری ضرورتی به این امر نیست ولی دانشجو باید فرآیندی که در جریان استفاده از نرم

افزارهای مذکور اتفاق می افتد را بداند.

متناسب با موقعیت شماره 1:

- تعیین دامنه تغییرات (R = (338.7 - 40.2) = 298.5)

- تعیین تعداد طبقات (k =  $\sqrt{N} = \sqrt{53} = 7.28 / 8$ )

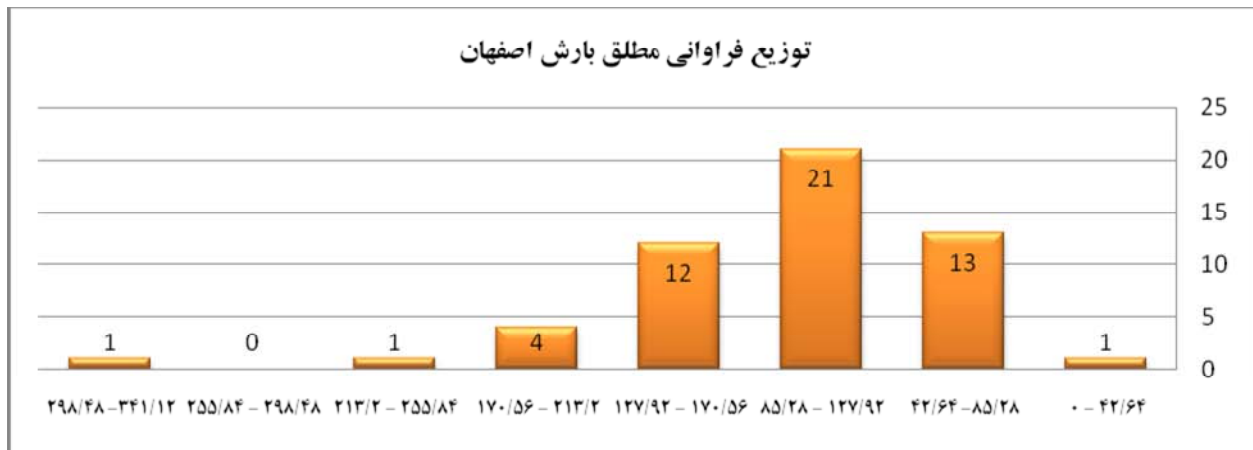
- فاصله طبقات (C = 298.5 / 7 = 42.64)

طبقات	حد میانه طبقات	فراوانی مطلق	فراوانی نسبی	درصد فراوانی نسبی	فراوانی تجمعی	نسبت تراکمی	درصد فراوانی تجمعی
0-42/64	21/32	1	0/0189	1/89	1	0/0189	1/89
42/64-85/28	63/96	13	0/2454	24/54	14	0/264	26/4
85/28-127/92	106/6	21	0/396	39/6	35	0/66	66
-170/56	149/24	12	0/2264	22/64	47	0/887	88/7



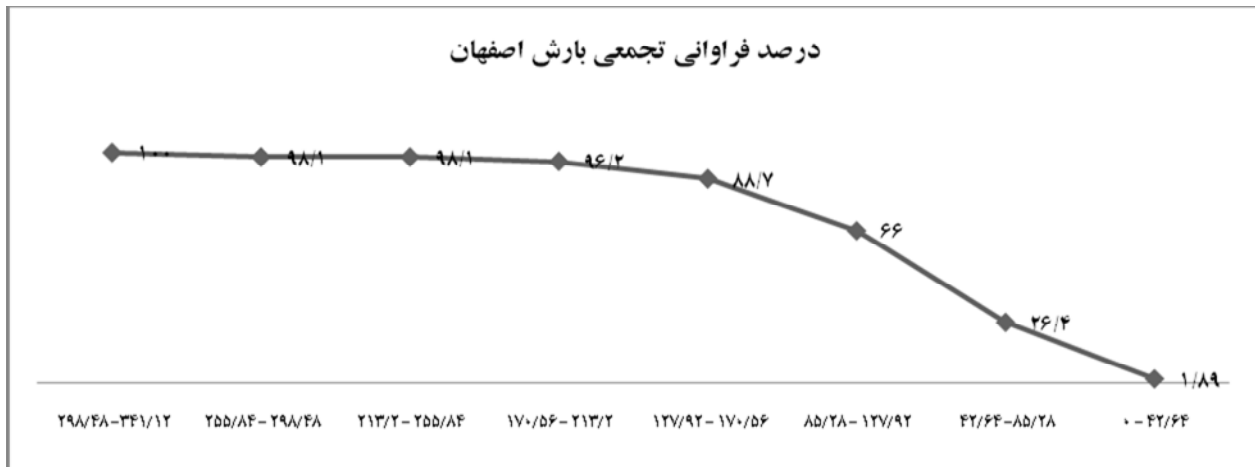
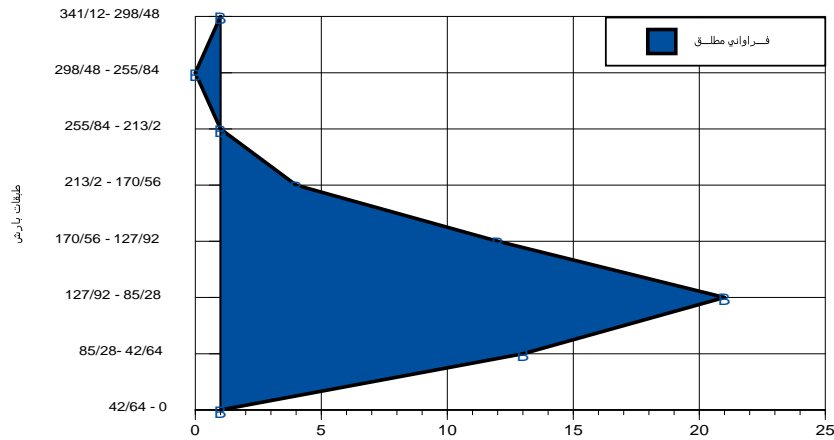
127/92							
-213/2	191/88	4	0/0755	7/55	51	0/962	96/2
170/56							
-255/84	234/52	1	0/0189	1/89	52	0/981	98/1
213/2							
-298/48	277/16	0	0	0	52	0/981	98/1
255/84							
-341/12	319/8	1	0/0189	1/89	53	1	100
298/48							
جمع		53	1	100			

نمود فراوانی های بارش ایستگاه اصفهان :



با توجه به نمودار شماره مشخص می سازد که فراوانی بارش در کدام طبقات توزیع شده است . لذا بیشترین فراوانی بارش سالانه ایستگاه اصفهان در حدود طبقات بارش 127/92 تا 85/28 میلی متر می باشد. از طرفی حدود 39/6 درصد بارش سالانه اصفهان در حدطبقاتی بارش مذکور قرار دارد و همچنین حدود 98/1 درصد بارش سالانه اصفهان در دوره مذکور (2003 تا 1951) از 298 /48 میلی متر کمتر است . یعنی

:



موقعیت شماره 2:

چنانچه سری رطوبت نسبی ایستگاه اصفهان 53 سال (1951 تا 2003) بصورت جدول شماره 2-2 در دست باشد سری مذکور را طبقه بندی

نمایید؟

- تعیین دامنه تغییرات ( $R = (57 - 28) = 29$ )

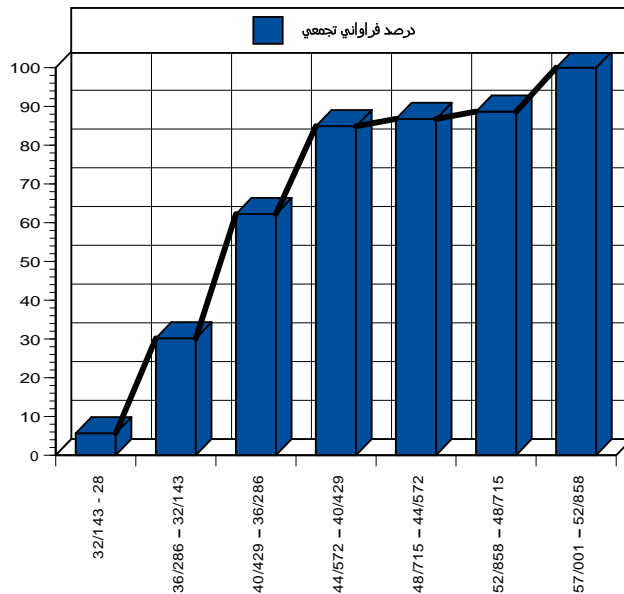
- تعیین تعداد طبقات ( $k = \sqrt{N} = \sqrt{53} = 7.28$ )

- فاصله طبقات ( $C = 29 / 7 = 4.143$ )

جدول شماره 2-2: جدول فراوانی های رطوبت نسبی ایستگاه اصفهان:

طبقات	حد میانه طبقات	فراوانی مطلق	فراوانی نسبی	درصد فراوانی نسبی	فراوانی تجمعی	درصد فراوانی تجمعی
28 32/143	30/07	3	0/057	5/7	3	5/7
- 36/286 32/143	34/214	13	0/245	24/5	16	30/19
- 40/429 36/286	38/357	17	0/321	32/1	33	62/26
- 44/572 40/429	42/5	12	0/226	22/6	45	84/9
- 48/715 44/572	46/64	1	0/0189	1/89	46	86/79
- 52/858 48/715	50/786	1	0/0189	1/89	47	88/68
- 57/001 52/858	54/929	6	0/1132	11/32	53	100
جمع		53	1	100		

متناسب با جدول شماره 32/1 درصد فراوانی سالانه رطوبت نسبی بین 40/429 تا 36/286 درصد می باشد. همچنین حدود 88/68 درصد میزان رطوبت نسبی سالانه اصفهان در طول دوره 53 ساله کمتر از 52/858 درصد می باشد. . به طریقی میزان رطوبت نسبی اصفهان نشان می دهد منبع رطوبتی اصفهان برای بارش بیرونی می باشد. یعنی :



استفاده از نرم افزار SPSS :

یکی از نرم افزار هایی که در تحقیقات اقلیمی از آن استفاده می شود نرم افزار SPSS<sup>1</sup> است . از این نرم افزار نسخه های متعددی جهت استفاده محققین وجود دارد. بر این اساس مراحل اشاره شده را می توان با استفاده از این نرم افزار نشان داد . لذا با توجه به موقعیت شماره 2 می توان مراحل عملیات آماری را با این نرم افزار نشان داد. یعنی :

	فراوانی مطلق	درصد فراوانی نسبی	درصد تجمعی فراوانی نسبی
28.00	1	1.9	1.9

<sup>1</sup> - Statistical Package for the Social Sciences

31.60	1	1.9	3.8
32.00	1	1.9	5.7
33.00	1	1.9	7.5
34.00	2	3.8	11.3
35.00	3	5.7	17.0
36.00	6	11.3	28.3
36.10	1	1.9	30.2
37.00	6	11.3	41.5
38.00	5	9.4	50.9
39.00	5	9.4	60.4
40.00	2	3.8	64.2
41.00	3	5.7	69.8
42.00	5	9.4	79.2
43.00	2	3.8	83.0
44.00	1	1.9	84.9
45.00	1	1.9	86.8
52.00	1	1.9	88.7
53.00	1	1.9	90.6
54.00	2	3.8	94.3
56.00	2	3.8	98.1
57.00	1	1.9	100.0
جمع	53	100.0	

نکته اساسی تفسیر توزیع فراوانی ها در تحلیل آماری اقلیم می باشد. با توجه به جدول رطوبت نسبی بالا نتایج نشان می دهد که 83 درصد رطوبت نسبی ایستگاه اصفهان در دوره 53 ساله حدود 43 درصد و کمتر است و حدود 98 درصد رطوبت نسبی ایستگاه اصفهان حدود 56 درصد و کمتر می باشد. لذا اهمیت میزان رطوبت نسبی در فرآیند های مختلف اقلیمی شامل تراکم، بارش و ... در این قلمرو قابل توجه می باشد.

## استفاده از نرم افزار NCSS :

یکی از نرم افزار های مهمی که می توان در مطالعات اقلیمی از آن استفاده کرد ، نرم افزار NCSS<sup>1</sup> است . این نرم افزار همراه با بسته آماری PASS<sup>2</sup> قابلیت های قابل توجهی را برای تحلیل های آماری برای محققین فراهم کرده است . لذا متناسب با موقعیت شماره 2 وضعیت رطوبت نسبی ایستگاه اصفهان با نرم افزار مذکور بررسی شده است :

## توزیع فراوانی رطوبت نسبی ایستگاه اصفهان

رطوبت نسبی	فراوانی مطلق	نسبی فراوانی تجمعی	درصد فراوانی نسبی	درصد تجمعی فراوانی نسبی
28	1	1	1.89	1.89
31.6	1	2	1.89	3.77
32	1	3	1.89	5.66
33	1	4	1.89	7.55
34	2	6	3.77	11.32
35	3	9	5.66	16.98
36	6	15	11.32	28.30
36.1	1	16	1.89	30.19
37	6	22	11.32	41.51
38	5	27	9.43	50.94
39	5	32	9.43	60.38
40	2	34	3.77	64.15
41	3	37	5.66	69.81
42	5	42	9.43	79.25
43	2	44	3.77	83.02
44	1	45	1.89	84.91
45	1	46	1.89	86.79
52	1	47	1.89	88.68

<sup>1</sup> - Number Cruncher Statistical System  
<sup>2</sup> - Power Analysis and Sample Size

53	1	48	1.89	90.57
54	2	50	3.77	94.34
56	2	52	3.77	98.11
57	1	53	1.89	100.00

### استفاده از نرم افزار Minitab :

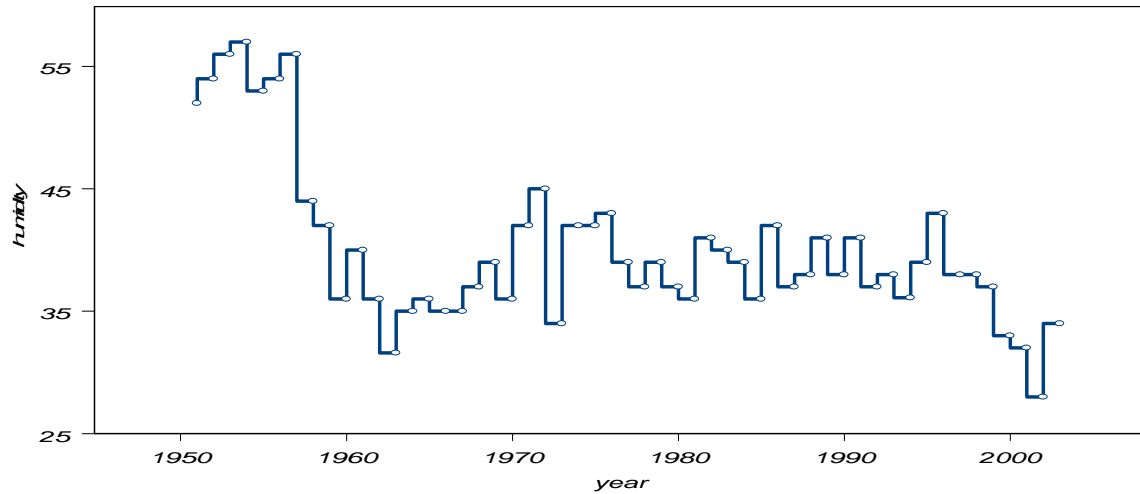
از نرم افزارهای مهم دیگری که می توان برای تحلیل های پیشرفته آماری استفاده نرم افزار مینی تب می باشد. بر این اساس متناسب با موقعیت

شماره 2 رطوبت نسبی ایستگاه اصفهان مورد بررسی قرار گرفت:

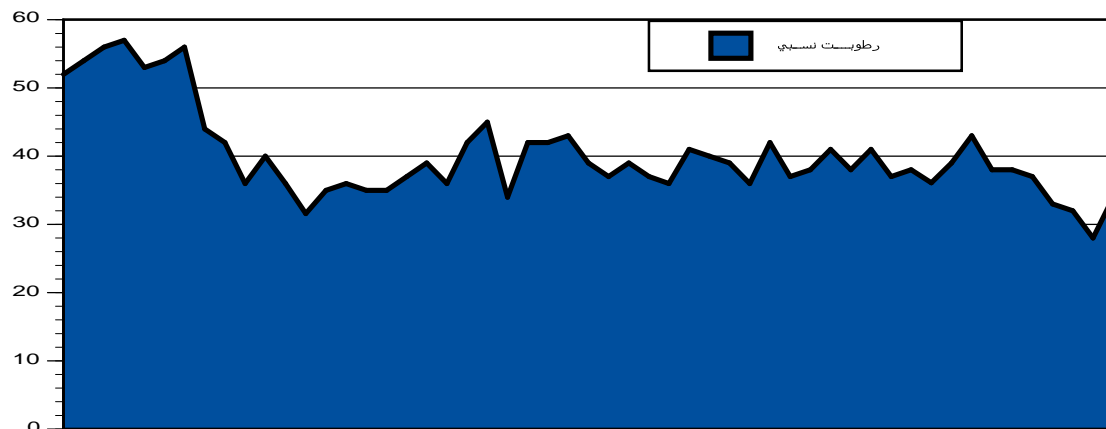
<i>humidity</i>	<i>Count</i>	<i>CumCnt</i>	<i>Percent</i>	<i>CumPct</i>
28.0	1	1	1.89	1.89
31.6	1	2	1.89	3.77
32.0	1	3	1.89	5.66
33.0	1	4	1.89	7.55
34.0	2	6	3.77	11.32
35.0	3	9	5.66	16.98
36.0	6	15	11.32	28.30
36.1	1	16	1.89	30.19
37.0	6	22	11.32	41.51
38.0	5	27	9.43	50.94
39.0	5	32	9.43	60.38
40.0	2	34	3.77	64.15
41.0	3	37	5.66	69.81
42.0	5	42	9.43	79.25
43.0	2	44	3.77	83.02
44.0	1	45	1.89	84.91
45.0	1	46	1.89	86.79
52.0	1	47	1.89	88.68
53.0	1	48	1.89	90.57
54.0	2	50	3.77	94.34
56.0	2	52	3.77	98.11
57.0	1	53	1.89	100.00

N= 53

تحلیل فراوانی ها در تحقیقات اقلیمی با ترسیم نمودارها از اهمیت زیادی برخوردار است . در این راستا نمایش نمودارها با استفاده از نرم افزارهای مختلف از اهمیت زیادی برخوردار است . برای نمونه رطوبت نسبی ایستگاه اصفهان با نمودارهای مختلف به نمایش گذاشته می شود. یکی از نرم افزارهای مهم برای نمایش نموداری متغیرهای اقلیمی نرم افزار **Splus** می باشد.



نمودار بالا میل کاهش کلی رطوبت نسبی را در دوره 53 ساله در ایستگاه اصفهان نشان می دهد.



نمودار بالا میل کاهش کلی رطوبت نسبی را در دوره 53 ساله در ایستگاه اصفهان نشان می دهد.

آماره های توصیفی و تغییر پذیری آنها:



عامل های اقلیمی بعنوان کمیت هایی قلمداد می شوند که از داده ها و اطلاعات مورد مطالعه در جامعه مورد بررسی بدست می آیند و با در دست داشتن آنها می توان خصوصیات هر قلمرو اقلیمی را مشخص و تحلیل نمود. آماره هایی که در مطالعات اقلیمی مورد توجه می باشند تحت عناوین زیر مدنظر قرار می گیرند:

- آماره هایی که در قالب بررسی وضعیت میانگین های اقلیمی مطرح می شوند.
- آماره هایی که در قالب بررسی وضعیت توزیع و پراکنش متغیرهای اقلیمی مطرح هستند
- آماره هایی که در قالب میزان انحرافات از شرایط بهنجاری مطرح می شوند .
- آماره هایی که در قالب میزان کشیدگی متغیرها را از شرائط بهنجاری نشان می دهند.

با توجه به دامنه مطالعات اقلیم شناسی آماره های مختلف به طور خاص مورد بررسی قرار خواهند گرفت .

آماره های بررسی وضعیت گرایش به مرکز متغیر های اقلیمی :

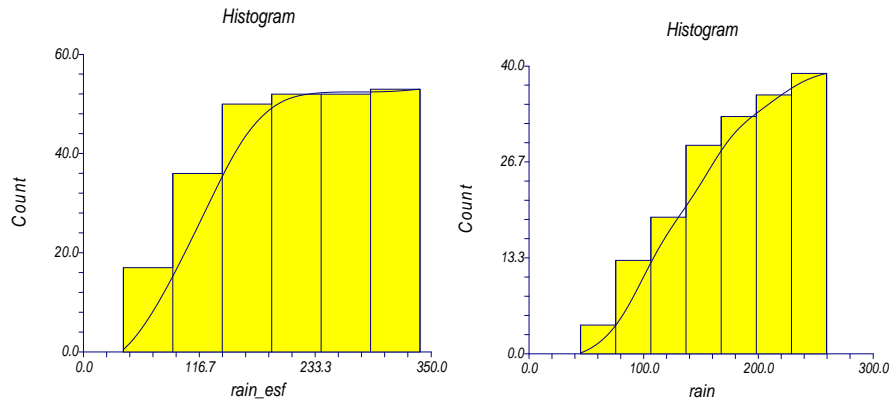
گرایش به مرکز در تحلیل های آماری مشخص کننده یک مقداری است که میانگین مقادیر یک نمونه را تعیین می کند . گرایش به مرکز از طریق تعیین مقداری که معرف مرکز توزیع یا میانگین مقادیر داده ها است ، بدست می آید. آماره های بسیار متداول در این طیف میانگین ها ، مد و میانه می باشد. نکته مهمی که در مطالعات اقلیمی باید به آن توجه شود تغییرپذیری این آماره ها در تحقیقات اقلیمی است. تغییر پذیری توزیع زمانی و مکانی متغیرهای اقلیمی است که نشان دهنده انحرافات متغیرها از اندازه مرکزی آنها است . این انحرافات را می توان در قالب دامنه تغییرات ، دامنه چارکی متوسط ، پراکنش ، انحراف معیار و ضریب تغییرات مورد بررسی قرار داد . برای توجیه مطلب به موقعیت شماره 1 توجه شود.

موقعیت شماره 1 :

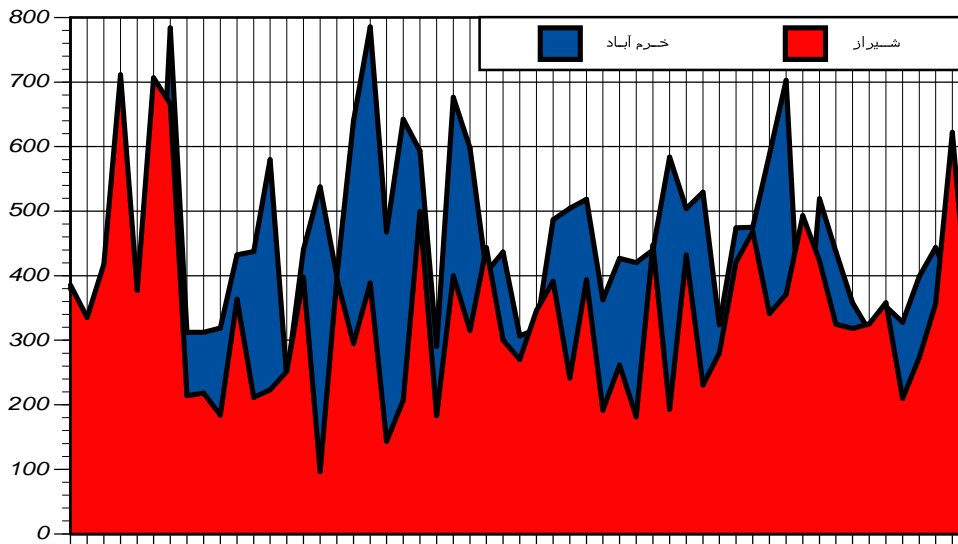
سری بارش ایستگاه های اصفهان و کاشان در دست بررسی است میزان توزیع یا پراگندگی یا تغییرپذیری کدام بیشتر است ؟

پراکنش بارش اصفهان

پراکنش بارش کاشان

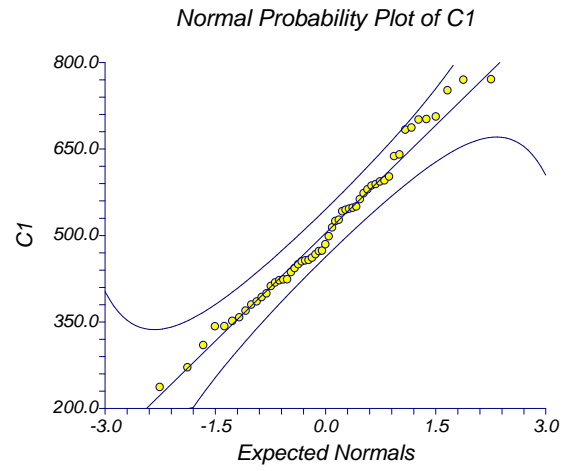
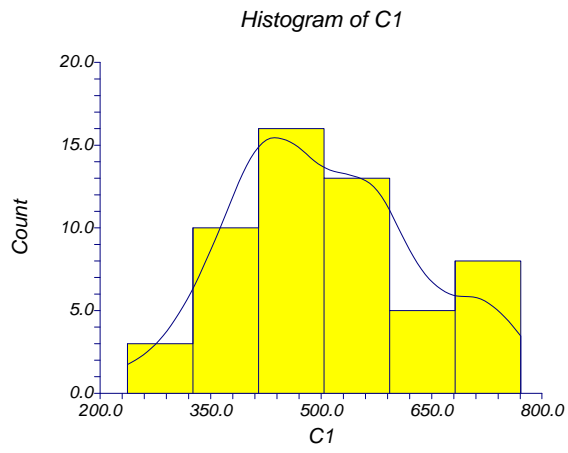


با توجه به نمودار های پراکنش بارش اصفهان و کاشان مشخص است که پراکنندگی یا تغییر پذیری بارش اصفهان بیشتر از بارش ایستگاه کاشان است . از نظر کمی هم می توان با محاسبه دامنه تغییرات یا سایر شاخص های پراکنندگی به این موضوع پی برد. در قسمت های بعدی کتاب به شاخص های پراکنندگی اشاره می شود. همچنین به نمودار های بارش ایستگاه های خرم آباد و شیراز (2005 تا 1951) در زیر توجه شود:

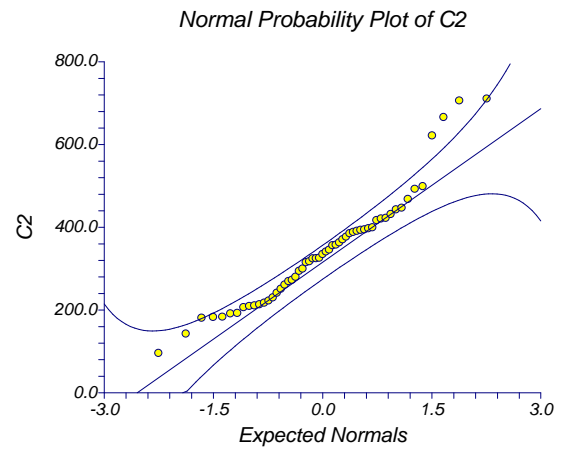
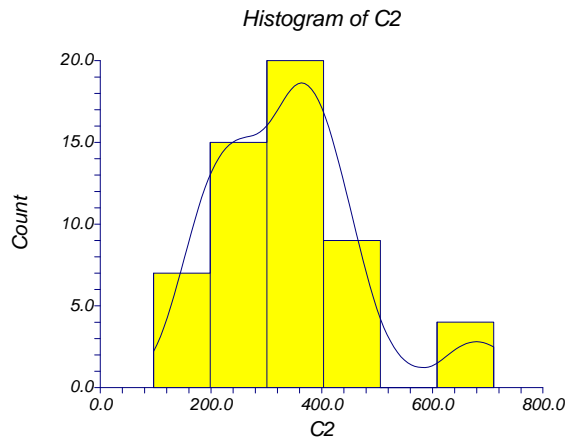


با توجه نمودار بالا هر چند آماره های توصیفی آن به خصوص میانگین آنها تفاوت دارند . یعنی میانگین بارش سالانه شیراز حدود  $342/209$  میلی متر و خرم آباد حدود  $506/02$  میلی متر می باشد . اما نمودار میزان پراکنندگی و تغییرپذیری بارش سالانه شیراز بیشتر می باشد. براین اساس هر چند توجه به آماره های توصیفی در تحقیقات اقلیمی ضروری است اما توجه به تغییرپذیری متغیرها در رابطه با آماره های توصیفی اهمیت بیشتر دارد . یعنی :

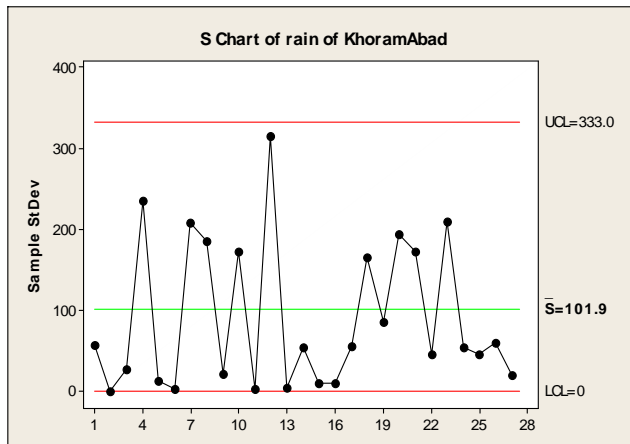
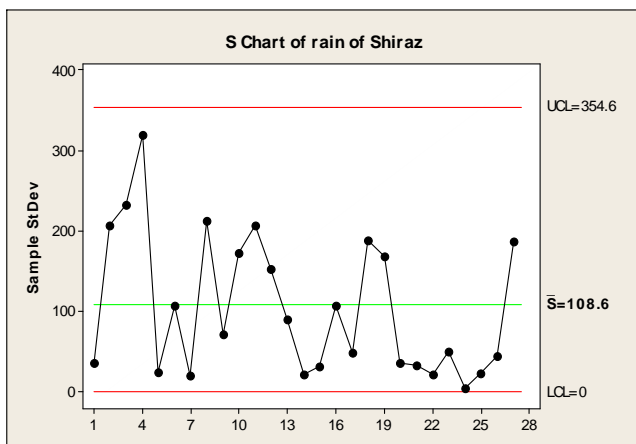
بارش خرم آباد



بارش شیراز



همچنین :



نمودار های بالا هم میزان پراکنش بیشتر بارش سالانه ایستگاه شیراز را نسبت به بارش سالانه ایستگاه خرم آباد نشان می دهد . بر این اساس توجه به آماره های توصیفی و تغییرپذیری آنها در مطالعات اقلیمی از اهمیت زیادی برخوردار است.

محاسبه میانگین حسابی متغیرهای اقلیمی :

یکی از مهمترین آماره هایی که در تحقیقات اقلیمی مورد استفاده قرار می گیرد میانگین حسابی است. میانگین حسابی<sup>1</sup> مقداری از متغیر است که بیشتر متغیرها حول آن تمرکز یافته اند. میانگین حسابی حاصل تقسیم جمع متغیرها بر تعداد آزمودنی ها (N) است . برای محاسبه میانگین از نماد یونانی زیگما (  $\sum$  ) برای نشان دادن مجموع نمره در متغیر (X) استفاده می شود . میانگین حسابی به ساده و دارای فراوانی تقسیم می شود. برای محاسبه میانگین حسابی ساده می توان از رابطه شماره 1-1 استفاده کرد:

$$\bar{X} = \frac{\sum X_i}{N} \quad (1-1)$$

$$\sum X_i = X_1 + X_2 + \dots + X_N \quad (2-1)$$

اگر نمره ها در مقیاس نمره ها به صورت منظم تنظیم شوند ، مقدار میانگین دقیقاً در مرکز ثقل توزیع قرار می گیرد . بر این اساس میانگین را می توان نقطه ای در توزیع نمره ها دانست که مجموع جبری تفاوت نمره ها از آن نقطه برابر با صفر است ، یعنی  $(\sum (X - \bar{X}) = 0)$  . برای تبیین بهتر به موقعیت شماره 2 توجه شود.

موقعیت شماره 2:

Arithmetic Mean - 1

اگر سری فشار اصفهان در دست باشد ، میانگین فشار در طول دوره 53 ساله چقدر می باشد.

جدول شماره 2-3: سری فراوانی های فشار اصفهان:

مقدار فشار به هکتوپاسکال	فراوانی مطلق	فراوانی نسبی	درصد تجمعی فراوانی نسبی
1003.00	1	1.9	1.9
1003.60	1	1.9	3.8
1004.40	2	3.8	7.5
1004.68	1	1.9	9.4
1004.80	3	5.7	15.1
1004.90	4	7.5	22.6
1005.00	1	1.9	24.5
1005.10	2	3.8	28.3
1005.20	2	3.8	32.1
1005.30	2	3.8	35.8
1005.40	2	3.8	39.6
1005.50	2	3.8	43.4
1005.60	3	5.7	49.1
1005.70	3	5.7	54.7
1005.80	2	3.8	58.5
1006.00	2	3.8	62.3
1006.30	1	1.9	64.2
1006.40	3	5.7	69.8
1006.50	1	1.9	71.7
1006.70	1	1.9	73.6
1006.90	1	1.9	75.5
1008.70	3	5.7	81.1
1009.20	1	1.9	83.0

1009.30	3	5.7	88.7
1009.40	2	3.8	92.5
1009.50	1	1.9	94.3
1010.20	1	1.9	96.2
1010.50	1	1.9	98.1
1011.00	1	1.9	100.0
کل	53	100.0	

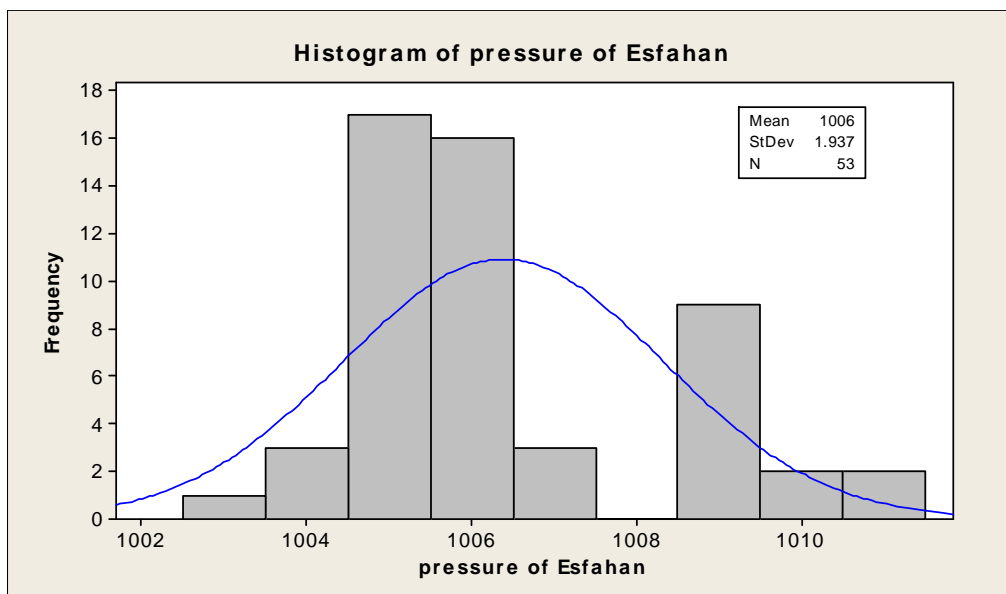
$$\bar{X} = \frac{\sum x_i}{N} = \frac{53338.38}{53} = 1006.38$$

مقدار میانگین فشار در دوره آماری 53 ساله ایستگاه اصفهان (2003 – 1951) برابر با 1006/38 میلی بار می باشد. همچنین حدود 75

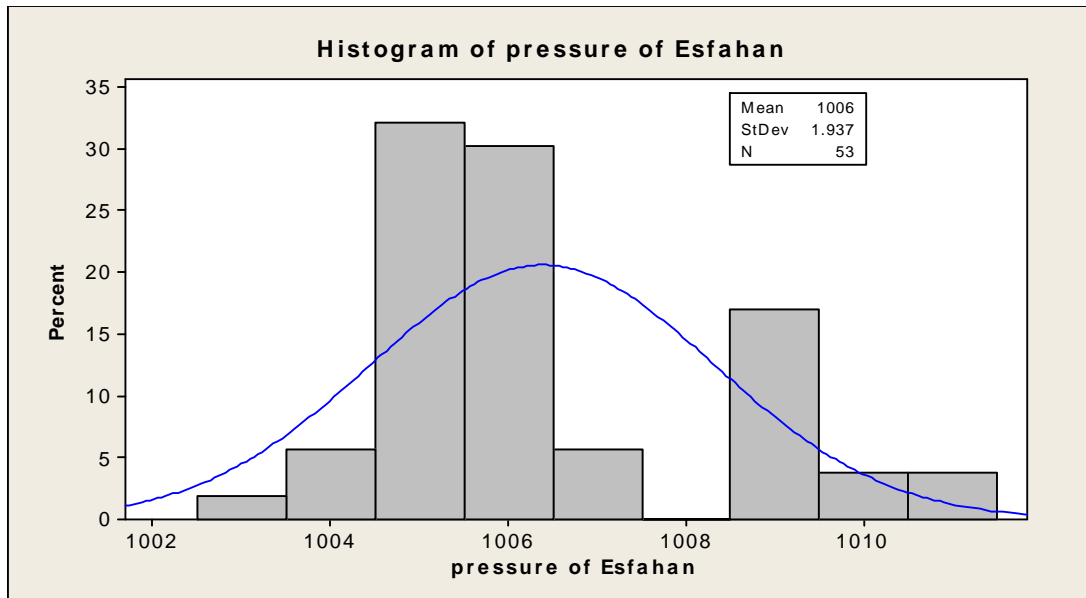
درصد از توزیع فشار کمتر از 1006/9 میلی بار می باشد. بر این اساس هیچ کدام از اندازه های گرایش به مرکز را نمی توان بهترین اندازه

در مطالعات و تحقیقات اقلیمی دانست. چون بیشتر سری های اقلیمی توزیع متقارن در دوره های زمانی ندارند. به شکل توزیع فشار

ایستگاه اصفهان نگاه کنید:



عدم متقارن بودن سری فشار اصفهان در دوره 53 ساله مؤید موضوع بالا است. از نکاتی که در تحقیقات اقلیمی باید به آن توجه نمود این است که میل اثر داده ها انتهای بر میانگین بیشتر است. به طوری که در فراوانی توزیع فشار ایستگاه اصفهان نشان می دهد که در طول دوره مذکور تعداد 37 سال آن فشار مساوی یا کمتر از میانگین کل دوره می باشد و تعداد 16 دوره آن بیشتر از میانگین کل دوره است. لذا میل اثر داده های انتهای دوره در میانگین بیشتر می باشد. به نمودار شماره توجه نمایید:



توزیع درصدی فشار اصفهان میل اثر داده های انتهای دوره در میانگین بهتر نشان می دهد. از نکات دیگری که باید در استفاده از میانگین در تحقیقات اقلیمی به آن توجه نمود این است که در توزیع نامتقارن داده ها از شاخص های دیگر گرایش به مرکز برای معرف آن توزیع باید استفاده کرد.

محاسبه میانگین با متغیرهای طبقه بندی شده:

در بیشتر مواقع به علت حجم زیادی داده ها در تحقیقات اقلیمی لازم است اقلیم شناس سری یا سری های مورد مطالعه را طبقه بندی نماید و بعد از طبقه بندی به محاسبه آماره های آنها پردازد. بر این اساس یکی از شیوه های محاسبه میانگین با استفاده از فراوانی ها ( $F_i$ ) می باشد. برای محاسبه میانگین با متغیرهای طبقه بندی شده می توان از رابطه (3-1) می توان استفاده نمود:

$$\bar{X} = \frac{1}{N} \sum F_i X_i \quad (3-1)$$

مراحل محاسبه :

- تعیین دامنه تغییرات (  $R = (1011 - 1003) + 1 = 9$  )
- تعیین تعداد طبقات (  $k = 1 + 3.322 \log 53 = 6.728$  ) : 7
- فاصله طبقات (  $C = 9 / 7 = 1.29$  )

جدول شماره 2-3: جدول توزیع فراوانی های فشار ایستگاه اصفهان :

حدود طبقات	فراوانی مطلق (Fi)	میانه طبقات (Mp)	Fi Mp
1003- 1004/29	2	1003/645	2007/29
- 1005/58 1004/29	21	1004/935	21103/635
- 1006/87 1005/58	16	1006/225	16099/6
- 1008/16 1006/87	1	1007/515	1007/515
- 1009/45 1008/16	9	1008/805	9079/245
- 1010/74 1009/45	3	1010/095	3030/285
- 1012/03 1010/74	1	1011/385	1011/385



جمع	53		53338/955
-----	----	--	-----------

$$\bar{X} = \frac{1}{53} \sum 53338.955 = 1006.395$$

با استفاده از نرم افزار SPSS:

*Descriptive Statistics*

	<i>N</i>	<i>Range</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Sum</i>	<i>Mean</i>
<i>pressure</i>	53	8.00	1003.00	1011.00	53338.38	1006.3845
<i>Valid N (listwise)</i>	53					

محاسبه میانگین وزنی متغیر های اقلیمی :

اگر در یک مجموعه از داده های اقلیمی مقادیر اندازه های متغیر مورد نظر مانند دما را نشان دهند و هر یک از این اندازه ها دارای اهمیت معینی باشند به عبارت دیگر هر یک بر حسب اهمیتی که دارند ، ضریب یا وزنی مانند  $W_1$ ،  $W_2$ ، ... و  $W_k$  نسبت داده شود یا به هر یک از متغیرها ضریبی داده شود لذا می توان میانگین وزنی<sup>1</sup> آنها را با استفاده از رابطه (4-2) محاسبه نمود:

$$\bar{X}_w = \frac{\sum w_i x_i}{\sum w_i} \quad (4-2)$$

### موقعیت شماره 3

اگر وضعیت زیر حوضه های آبریز اورگان در غرب استان اصفهان به صورت جدول زیر باشد میانگین وزنی بارش حوضه آبریز مذکور را محاسبه کنید ؟

زیر حوضه	مساحت (km2) (W)	بارش (mm) (X <sub>i</sub> )	بارش در مساحت	نسبت درصد مساحت	نسبت درصد مساحت به بارش
1	11/1	392/5	4356/75	21/876	8583/97
2	5/93	348/4	2066/012	11/687	4002/07
3	19/95	441/5	8807/93	39/32	17359/78
4	1/03	397/4	409/322	2/03	806/722
5	12/73	490/5	6244/065	25/09	12306/645
جمع	50/74	میانگین: 414/06	21884/074	100	43059/187

<sup>1</sup> - Weighted Mean

$$\bar{X}_w = \frac{\sum w_i x_i}{\sum w_i} = \frac{21882.074}{50.74} = 431.3 \text{ میلی متر}$$

$$\bar{X}_{Rw} = \frac{\sum w_i x_i}{\sum w_i} = \frac{43059.187}{50.74} = 848.624$$

همانطوری که مشاهده می شود میانگین بارش حوضه برابر با **414/06** میلی متر می باشد ولی میانگین وزنی بارش در حوضه ها **431/3** میلی متر می باشد.

محاسبه میانگین هندسی متغیرهای اقلیمی:

اگر سری داده های جمع آوری شده به صورت نسبی یا درصدی باشند مانند نسبت ها ، درصدها، شاخص ها و ... از میانگین هندسی<sup>1</sup> استفاده می شود. میانگین هندسی متغیرها از جذر **n**ام حاصلضرب آنها بدست می آید . برای محاسبه میانگین هندسی از رابطه (5-2) استفاده می شود.

$$\bar{X}_G = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} \quad (5-2)$$

$$\bar{X}_G = (x_1 \times x_2 \times \dots \times x_n)^{1/n}$$

همچنین می توان از رابطه زیر استفاده کرد:

$$\bar{X}_G = \text{antilog} \frac{1}{n} \sum \log x_i \quad (6-2)$$

موقعیت شماره 4:

اگر سری دمای سالانه ایستگاه اصفهان 53 (2003 - 1951) در دست باشد میانگین هندسی آن را حساب کنید :

ردیف	دما	لگارتیم دما	ردیف	دما	لگارتیم دما
1	15.30	1.185	28	16.40	1.215
2	14.90	1.173	29	16.30	1.212

<sup>1</sup> Geometric Mean

3	15.20	1.182	30	16.50	1.217
4	15.30	1.185	31	16.40	1.215
5	15.40	1.188	32	15.80	1.199
6	14.90	1.173	33	16.10	1.207
7	14.30	1.155	34	15.90	1.201
8	16.90	1.228	35	16.90	1.228
9	16.60	1.220	36	15.90	1.201
10	16.70	1.223	37	17.00	1.230
11	16.10	1.207	38	17.20	1.236
12	16.00	1.204	39	16.50	1.217
13	16.90	1.228	40	17.40	1.241
14	15.60	1.193	41	16.80	1.225
15	16.10	1.207	42	15.40	1.188
16	16.70	1.223	43	16.30	1.212
17	15.60	1.193	44	18.74	1.273
18	16.00	1.204	45	15.90	1.201
19	16.00	1.204	46	15.60	1.193
20	17.00	1.230	47	16.20	1.210
21	16.20	1.210	48	16.60	1.220
22	14.50	1.161	49	16.90	1.228
23	16.20	1.210	50	16.90	1.228
24	15.30	1.185	51	17.60	1.246
25	15.70	1.196	52	17.40	1.241

<b>26</b>	<b>15.70</b>	<b>1.196</b>	<b>53</b>	<b>17.00</b>	<b>1.230</b>
<b>27</b>	<b>16.80</b>	<b>1.225</b>			میانگین <b>1/20944</b>

$$\bar{X}_G = \text{antilog} \frac{1}{53} \sum 64.1 = 16.19$$

یا

محاسبه میانگین هندسی با نرم افزار SPSS :

ردیف	دما	ردیف	دما	ردیف	دما
1	15.30	19	16.00	37	17.00
2	14.90	20	17.00	38	17.20
3	15.20	21	16.20	39	16.50
4	15.30	22	14.50	40	17.40
5	15.40	23	16.20	41	16.80
6	14.90	24	15.30	42	15.40
7	14.30	25	15.70	43	16.30
8	16.90	26	15.70	44	18.74
9	16.60	27	16.80	45	15.90
10	16.70	28	16.40	46	15.60
11	16.10	29	16.30	47	16.20
12	16.00	30	16.50	48	16.60
13	16.90	31	16.40	49	16.90
14	15.60	32	15.80	50	16.90

15	16.10	33	16.10	51	17.60
16	16.70	34	15.90	52	17.40
17	15.60	35	16.90	53	17.00
18	16.00	36	15.90	میانگین هندسی	16/19

محاسبه میانگین هندسی با نرم افزار NCSS:

*Means Section of temperature*

			<i>Geometric</i>
<i>Parameter</i>	<i>Mean</i>	<i>Median</i>	<i>Mean</i>
<i>Value</i>	16.21774	16.2	16.19728
<i>Std Error</i>	0.1130378		
<i>95% UCL</i>	16.44456	16.5	16.42539
<i>T-Value</i>	143.4718		
<i>Prob Level</i>	0.000000		
<i>Count</i>	53		53

محاسبه میانگین هندسی با نرم افزار Statistica7 :

<i>Descriptive Statistics (esf2)</i>			
	<i>Geometric</i>	<i>Minimum</i>	<i>Maximum</i>
<i>temperature.esf</i>	16.19728	14.30000	18.74000

از نکات مهمی که باید در محاسبه میانگین هندسی در مطالعات و تحقیقات اقلیمی به آن اشاره نمود این است که متغیرهای اقلیمی به صورت مشخصه های جبری آنها را مورد توجه قرار می دهد. از طرفی اگر متغیرهای اقلیمی دارای مقادیر منفی و صفر باشد ، استفاده از این میانگین توصیه نمی شود . چون مقادیر منفی و صفر با این شیوه قابل محاسبه نمی باشد.

محاسبه میانگین همساز یا هارمونیک متغیر های اقلیمی :

برای محاسبه میانگین همساز<sup>1</sup> متغیرهای اقلیمی باید ابتدا مقادیر معکوس متغیرها را محاسبه نمود و بعد میانگین آنها را بدست آورد. در اقلیم وقتی از این نوع میانگین استفاده می شود که مقدار و شدت و ارزش متغیرها متفاوت باشد. برای محاسبه میانگین همساز متغیرها از رابطه شماره (7-2) استفاده می شود.

$$\bar{X}_h = \frac{n}{\sum \frac{1}{x_i}} \quad (7-2)$$

برای فهم شیوه میانگین همساز لازم است به موقعیت شماره 5 توجه شود .

موقعیت شماره 5:

اگر سری بارش 20 ساله (1986 تا 2005) ایستگاه دهلران در دست باشد میانگین همساز سری را محاسبه نمایید ؟

بارش	بارش/1	بارش	بارش/1
510.30	0.002	648.20	0.002
228.90	0.004	293.90	0.003
405.30	0.002	256.70	0.004
361.80	0.003	268.10	0.004
212.10	0.005	205.50	0.005
285.40	0.004	230.80	0.004
366.50	0.003	262.80	0.004
375.60	0.003	248.60	0.004
402.00	0.002	257.50	0.004
169.90	0.006	214.50	0.005

$$\bar{X}_h = \frac{20}{0.072} = 277.78$$

محاسبه میانگین همساز با نرم افزار SPSS :

بارش	بارش
510.30	648.20
228.90	293.90
405.30	256.70
361.80	268.10
212.10	205.50
285.40	230.80
366.50	262.80
375.60	248.60
402.00	257.50
169.90	214.50
میانگین همساز	278/998

محاسبه میانگین همساز بارش دهلران با نرم افزار NCSS :

*Means Section of Dehloran\_rain*

<i>Parameter</i>	<i>Mean</i>	<i>Median</i>	<i>Geometric Mean</i>	<i>Harmonic Mean</i>	<i>Sum</i>	<i>Mode</i>
<i>Value</i>	310.22	265.45	293.0873	<u>278.9966</u>	6204.4	
<i>Std Error</i>	26.08392				521.6785	
<i>95% LCL</i>	255.6257	230.8	250.6346	244.2473	5112.515	
<i>95% UCL</i>	364.8143	366.5	342.7307	325.2735	7296.286	
<i>T-Value</i>	11.8931					



Prob Level 0.000000

20 20

20 Count

محاسبه میانگین متحرک متغیرهای اقلیمی :

یکی از شاخص های گرایش به مرکز بسیار مهم در اقلیم شناسی میانگین متحرک<sup>1</sup> می باشد . با توجه به ماهیت زمانی متغیرهای اقلیمی استفاده از این آماره اهمیت زیادی دارد . میانگین متحرک با رتبه های مختلف در نظر گرفته و محاسبه می شود . برای محاسبه میانگین متحرک می توان از رابطه شماره (8-2) استفاده کرد:

$$\begin{aligned}\bar{X}_{m(1)} &= \frac{X_1 + X_2 + \dots + X_n}{n} \\ \bar{X}_{m(2)} &= \frac{X_2 + X_3 + \dots + X_{n+1}}{n} \\ \bar{X}_{m(3)} &= \frac{X_3 + X_4 + \dots + X_{n+2}}{n}\end{aligned} \quad (8-2)$$

برای سری های زمانی می توان نیمه میانگین ها<sup>2</sup> را نیز محاسبه نمود. در این روش ابتدا باید نقطه میانی سری را مشخص نمود و بعد میانگین را برای نیمه های اول و دوم محاسبه نمود. در این صورت جامعه دارای دو میانگین است . برای فهم بهتر به موقعیت شماره 6 اشاره می شود.

موقعیت شماره 6 :

اگر بارش 20 ساله ( 2005 - 1986 ) ایستگاه کرمانشاه در دست باشد ، میانگین متحرک رتبه 4 آن را حساب کنید؟

1 - Moving averages  
2 - Semi Averages

بارش	میانگین متحرک رتبه 4
440.00	
583.80	
504.00	$\bar{X}_{m(1)} = \frac{440 + 583.8 + 504 + 529.5}{4} = 514.325$
529.50	485/35
324.10	457/975
474.30	450/725
475.00	465/3
587.80	559/9
702.50	495/275
215.80	506/3
519.10	468/725
437.50	382/575
357.90	407/3
314.70	365/625

352.40	338/2
327.80	348/425
398.80	380/65
443.60	388/425
383.50	402/75
384.80	

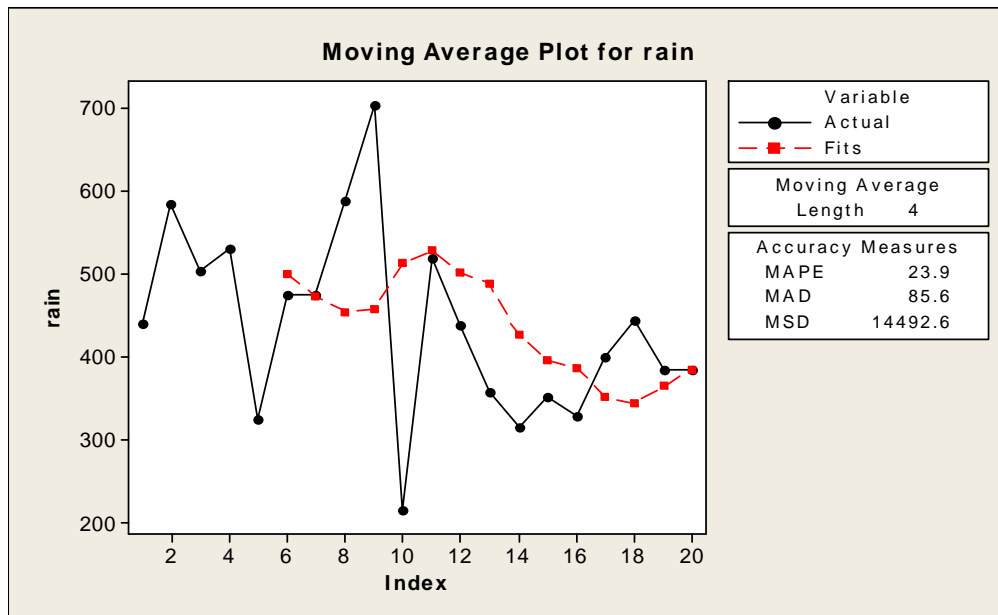
محاسبه میانگین متحرک رتبه 4 با نرم افزار Minitab:

#### Accuracy Measures

MAPE 26.2  
MAD 91.4  
MSD 16840.2

Time	rain	MA	Predict	Error
1	440.0	*	*	*
2	583.8	*	*	*
3	504.0	*	*	*
4	529.5	514.325	*	*
5	324.1	485.350	514.325	-190.225
6	474.3	457.975	485.350	-11.050
7	475.0	450.725	457.975	17.025
8	587.8	465.300	450.725	137.075
9	702.5	559.900	465.300	237.200
10	215.8	495.275	559.900	-344.100
11	519.1	506.300	495.275	23.825
12	437.5	468.725	506.300	-68.800
13	357.9	382.575	468.725	-110.825
14	314.7	407.300	382.575	-67.875
15	352.4	365.625	407.300	-54.900
16	327.8	338.200	365.625	-37.825
17	398.8	348.425	338.200	60.600
18	443.6	380.650	348.425	95.175

19	383.5	388.425	380.650	2.850
20	384.8	402.675	388.425	-3.625



از نکات لازم در کاربرد میانگین ها این است که اگر هر عددی در متغیرها ضرب یا تقسیم کنیم به همان نسبت در میانگین مؤثر می باشد. یا اگر همه مقادیر یک متغیر از یک عدد ثابت کم شود میانگین هم به همان مقدار کم می شود. از طرفی مجموع انحراف مقادیر از میانگین برابر با صفر است. میانگین فقط برای متغیرهای فاصله ای و نسبی به کار برده می شود. در میانگین همه داده ها و توزیع فراوانی آنها اثر دارد. از طرفی باید اقلیم شناس بداند که توزیع های متفاوت دارای میانگین یکسان می باشند و تمایل میانگین به طرف چولگی بیشتر از سایر شاخص های گرایش به مرکز است. به هر حال در تحلیل های اقلیمی استفاده از میانگین اهمیت زیادی دارد زیرا در محاسبه میانگین متغیرهای اقلیمی همه داده ها مؤثر هستند و متغیرها با دامنه های متفاوت در آن تأثیر دارد.

محاسبه میانه متغیرهای اقلیمی :

یکی از آماره های گرایش به مرکز وضعیت میانه<sup>1</sup> سری ها می باشد . اگر سری اقلیمی را به صورت نزولی یا به صورت صعودی مرتب کنند مقدار متغیر یا صفتی که در وسط قرار می گیرد میانه سری یا متغیرها نامیده می شود. لذا میانه مقداری از متغیر است که فراوانی مقادیر بزرگتر و کوچکتر از آن مساوی باشد . روش محاسبه میانه به دو شکل داده های طبقه بندی نشده و طبقه بندی شده متفاوت خواهد بود.

برای محاسبه میانه ابتدا باید سری را مرتب و نقطه میانی آن با  $m_e = \frac{n}{2}$  بدست آورد . برای محاسبه میانه داده های طبقه بندی نشده دو

حالت وجود دارد . اگر تعداد سری فرد باشد بعد از مرتب کردن عدد وسطی میانه است . برای تعیین عدد میانه کافی است از رابطه

$m_e = \frac{(n+1)}{2}$  استفاده نمود. اگر سری زوج باشد بعد از مرتب کردن معدل دو عدد میانی میانه می باشند. برای فهم بهتر به موقعیت

شماره 7 اشاره می شود .

موقعیت شماره 7 :

اگر سری ابرناکی ایستگاه اصفهان در دست باشد ، میانه آن سری را محاسبه نمایید؟

ردیف	مقدار ابرناکی	مقدار ابرناکی مرتب شده	ردیف	مقدار ابرناکی	مقدار ابرناکی مرتب شده
1	29.00	13.00	28	15.00	34.00
2	40.00	15.00	29	27.00	34.00
3	51.00	15.00	30	29.00	34.00
4	56.00	15.00	31	15.00	36.00
5	36.00	17.00	32	45.00	36.00
6	47.00	21.00	33	32.00	36.00
7	54.00	22.00	34	34.00	36.00
8	45.00	23.00	35	31.00	37.00
9	34.00	23.00	36	49.00	37.00

<sup>1</sup> 7median -

10	28.00	24.00	37	25.00	37.00
11	36.00	25.00	38	38.00	38.00
12	34.00	25.00	39	36.00	40.00
13	40.00	27.00	40	15.00	40.00
14	32.00	28.00	41	31.00	40.00
15	37.00	28.00	42	25.00	43.00
16	37.00	29.00	43	34.00	45.00
17	31.00	29.00	44	13.00	45.00
18	34.00	30.00	45	40.00	46.00
19	52.00	30.00	46	46.00	47.00
20	30.00	31.00	47	28.00	47.00
21	37.00	31.00	48	17.00	49.00
22	43.00	31.00	49	23.00	51.00
23	24.00	32.00	50	22.00	52.00
24	47.00	32.00	51	23.00	52.00
25	33.00	33.00	52	21.00	54.00
26	52.00	34.00	53	30.00	56.00
27	36.00	34.00			

$$m_e = \frac{(53 + 1)}{2} = 27$$

عدد میانه عدد 27 ام یعنی 34 است .

محاسبه میانه با استفاده از نرم افزار NCSS :

Parameter	Mean	Median	Geometric Mean	Harmonic Mean	Sum	Mode
Value	33.9434	<u>34</u>	32.14198	30.13879	1799	34
Std Error	1.461508				77.45992	
95% LCL	31.01067	30	29.20386	27.13856	1643.565	
95% UCL	36.87613	36	35.37569	33.88484	1954.435	
T-Value	23.2249					
Prob Level	0.000000					
Count	53		53	53		5

محاسبه میانه برای سری های اقلیمی طبقه بندی شده:

برای محاسبه میانه از طریق سری طبقه بندی شده باید با تعیین تعداد طبقات و تنظیم داده ها در قالب فراوانی های مطلق و تجمعی سری را

مورد بررسی قرار داد. برای محاسبه میانه سری طبقه بندی شده از رابطه شماره (9-2) می توان استفاده کرد:

$$M_e = L + \frac{N/2 - F}{f} \times C \quad (9-2)$$

L کرانه پایین که میانه در آن قرار دارد، F فراوانی تجمعی ماقبل طبقه ای که میانه در آن قرار دارد، f فراوانی مطلق طبقه ای که میانه در

آن قرار دارد. برای فهم بهتر به موقعیت شماره 8 توجه شود.

موقعیت شماره 8:

اگر سری فشار ایستگاه اصفهان به صورت جدول زیر در دسترس باشد میانه سری را محاسبه کنید؟

حدود طبقات	فراوانی مطلق	فراوانی تجمعی
1003- 1004/29	2	2
- 1005/58	21	<u>23</u>
1004/29		
- 1006/87	16	39
1005/58		

-1008/16 1006/87	1	40
-1009/45 1008/16	9	49
-1010/74 1009/45	3	52
-1012/03 1010/74	1	53
جمع	53	

- ابتدا مقدار  $53 \div 2 = 26.5$  را بدست می آورند، بر اساس عدد بدست آمده طبقه ای که میانه در آن قرار دارد را مشخص می نمایند .
- عدد  $26/5$  نشان می دهد که میانه در طبقه فراوانی تجمعی بین  $23$  و  $39$  قرار دارد ، لذا طبقه میانه دار طبقه ای است که فراوانی تجمعی آن  $23$  می باشد چون به عدد مبنا نزدیکتر است . کرانه پایین آن برابر است با  $1.29 \div 2 = 0.645$  لذا براساس فرمول :

$$M_e = 1004.29 + \frac{26.5 - 2}{21} \times 1.29$$

$$= 1005.79$$

محاسبه میانه با استفاده از نرم افزار **S-Plus** :

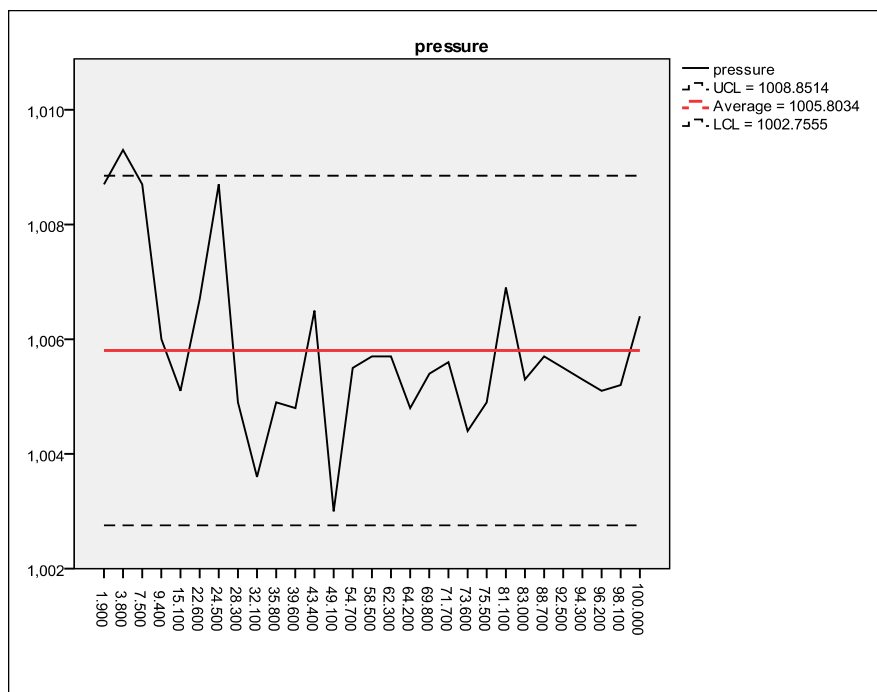
\*\*\* Summary Statistics for data in: DS6 \*\*\*

pressure  
 Min: 1003.000  
 Mean: 1006.385  
 Median: 1005.700  
 Max: 1011.000

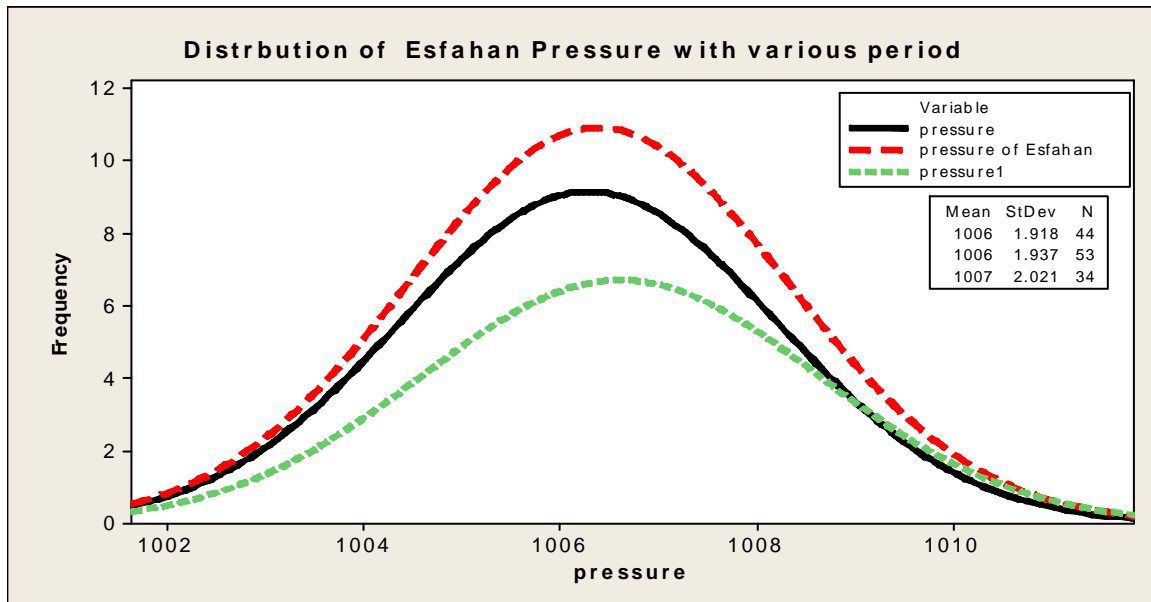


Total N: 53.000  
 NA's : 0.000  
 Sum: 53338.380

میانه در تحلیل های اقلیمی نشان می دهد که 50 درصد داده ها بزرگتر یا کوچکتر از آن می باشد . از طرفی میانه برای آن سری از متغیرهای اقلیمی بکار می رود که توزیع نامتقارن دارند . برای تبیین بهتر مطلب به نمودار شماره توجه نمایید:



به هر حال میانه به عنوان یک اندازه از آماره های گرایش به مرکز توزیع هایی که شکل آنها نامتقارن است استفاده می شود. میانه چون کمتر تحت تأثیر نمره های انتهایی در توزیع قرار می گیرد ، در تحقیقات اقلیمی به خصوص در توزیع فراوانی ها اهمیت زیادی دارد. به نمودارهای شماره توجه نمایید :



با توجه به نمودار شماره ۱ اثر نمره های انتهایی در مقدار میانه و پراکنش سری فشار ایستگاه اصفهان مشخص است. هر چه طول دوره کاهش می یابد توزیع سری نامتقارن تر و میزان پراکنش سری افزایش می یابد. با افزایش نامتقارنی سری اهمیت استفاده از میانه در بین اندازه های گرایش به مرکز بیشتر می شود. به هر حال استفاده از آماره میانه در تحلیل های اقلیمی نشان می دهد که کدام متغیر در وسط سری مورد بررسی قرار دارد و نشانگر این است که 50 درصد سری بالاتر یا پائین تر از آن هستند.

#### محاسبه چارک ها در تحلیل آماری متغیرهای اقلیمی :

چارک ها مقداری از متغیرهای اقلیمی هستند که فراوانی کل را به چهار قسمت مساوی تقسیم می نمایند. به عبارت دیگر چارک ها مقداری از متغیرند که دامنه را به چهار قسمت تقسیم می کنند، بطوری که هر یک از قسمت ها یک چهارم فراوانی کل را دربر داشته باشد. روش محاسبه چارک ها مانند محاسبه میانه است. چارک اول با  $Q_1$ ، چارک دوم با  $Q_2$ ، چارک سوم با  $Q_3$  و چارک چهارم را با  $Q_4$  نشان می دهند. چارک اول مقداری از متغیر است که فراوانی تجمعی نظیر آن برابر با 25 درصد فراوانی کل باشد و چارک دوم همان میانه و مقداری از متغیر است که فراوانی تجمعی آن 50 درصد فراوانی کل باشد و چارک سوم مقداری است که فراوانی تجمعی آن برابر

75 درصد فراوانی کل باشد و چارک چهارم شامل 100 درصد فراوانی ها می شود. برای نمایش چارک ها کافی است از شاخص  $\frac{aN}{4}$

استفاده کرد. مثلاً " اگر چارک اول مورد نظر باشد  $\frac{1N}{4}$ ، اگر چارک دوم مورد نظر باشد  $\frac{2N}{4}$  و اگر چارک سوم مورد نظر باشد  $\frac{3N}{4}$

گرفته می شود. برای محاسبه چارک ها به صورت سری طبقه بندی نشده و سری طبقه بندی شده می توان عمل کرد. برای تبیین بهتر به موقعیت شماره 9 اشاره می شود.

موقعیت شماره 9:

اگر سری بارش 20 ساله (1986 تا 2005) ایستگاه اهواز در دست باشد، چارک اول و سوم آن را محاسبه کنید؟

محاسبه به شیوه سری طبقه بندی نشده:

مراحل:

ابتدا سری را از کوچک به بزرگ مرتب می کنند. سپس میانه یا چارک دوم سری را محاسبه نمایند و در نهایت میانگین سری سمت

چپ (کمتر) چارک اول سری می باشد و سری سمت راست (بیشتر) چارک سوم سری می باشد. یعنی:

	بارش	بارش مرتب شده	ردیف	بارش	بارش مرتب شده
1	327.40	86.40	11	289.90	227.80
2	250.90	136.00	12	468.80	228.20
3	185.00	170.90	13	170.90	234.80
4	224.80	171.30	14	227.80	250.90
5	136.00	185.00	15	234.80	271.80
6	410.30	188.10	16	227.30	289.90
7	302.50	218.80	17	171.30	302.50
8	219.10	219.10	18	218.80	327.40
9	228.20	224.80	19	271.80	410.30
10	86.40	227.30	20	188.10	468.80

$$Q_1 = \frac{N}{4} = \frac{20}{4} = 5$$

$$Q_3 = \frac{3N}{4} = \frac{60}{4} = 15$$

لذا: چارک اول سری 185 و چارک سوم 271/8 می باشد

محاسبه چارک ها به شیوه سری طبقه بندی شده:

مراحل:

- تعیین دامنه تغییرات (  $R = (468.8 - 86.4) = 382.4$  )

- تعیین تعداد طبقات (  $k = \sqrt{N} = \sqrt{20} = 4.47 = 5$  )

- فاصله طبقات (  $C = 382.4 / 5 = 76.48 = 76.5$  )

طبقات	فراوانی	فراوانی تجمعی
86/4- 162/9	2	2
162/9 - 239/4	11	13
239/4- 315/9	4	17
315/9- 392/4	1	18
392/4- 468/9	2	20
جمع	20	

$$Q_1 = L + \frac{\frac{N}{4} - F}{f} \times C$$

برای چارک اول:  $Q_1 = L + \frac{\frac{N}{4} - F}{f} \times C$

$$Q_3 = L + \frac{\frac{3N}{4} - F}{f} \times C \quad \text{برای چارک سوم :}$$

مراحل چارک اول:

$$\frac{1 \times 20}{4} = 5$$

طبقه ای که چارک اول در آن قرار دارد طبقه اول می باشد ، لذا:

$$Q_1 = 162.9 + \frac{5 - 2}{11} \times 76.5 = 183.76$$

مراحل چارک سوم:

$$\frac{3 \times 20}{4} = 15$$

طبقه ای که چارک سوم در آن قرار دارد ، طبقه سوم است ، لذا:

$$Q_3 = 239.4 + \frac{15 - 13}{4} \times 76.5 = 277.65$$

محاسبه چارک ها با استفاده از نرم افزار SPSS :

### Statistics

prec

N	Valid	20
	Missing	0
Percentile	25	185.7750
	50	227.5500

s	75	285.3750
---	----	----------

محاسبه چارک ها با استفاده از نرم افزار NCSS :

### Descriptive Statistics Report

#### Quartile Section of C12

	10th	25th	50th	75th	90th
Parameter	Percentile	Percentile	Percentile	Percentile	Percentile
Value	139.49	185.775	227.55	285.375	402.01
95% LCL		86.4	188.1	227.8	
95% UCL		224.8	271.8	410.3	

محاسبه دهک ها و صدک ها در تحلیل آماری اقلیم :

دهک به این معنی است که یک سری را به ده قسمت مساوی تقسیم نمایند و مقدار هر یک از دهک ها را مشخص کنند.

نحوه محاسبه دهک ها مانند چارک ها می باشد با این تفاوت که برای محاسبه آن می توان از رابطه زیر استفاده کرد:

$$D_x = \frac{x(N+1)}{10}$$

صدک هم مانند دهک می باشد با این تفاوت که یک سری را به صد قسمت تقسیم می کنند و مقدار آن بر اساس رابطه

زیر محاسبه می شود :

$$p_x = \frac{x(N+1)}{100}$$

متناسب با سری بارش اهواز دهک ششم و صدک هفتم آن را حساب کنید؟

مراحل :

برای داده های طبقه بندی نشده :

با توجه به مرتب کردن سری :

$$D_x = \frac{6(20 + 1)}{10} = 12.6$$

لذا

$$D_6 = \frac{228.2 + 234.8}{2} = 231.5$$

برای داده های طبقه بندی شده:

$$D_x = L + \frac{\frac{xN}{10} - F}{f} \times C$$

برای دهک ششم سری اهواز:

$$D_6 = \frac{6 \times 20}{10} = 12$$

$$D_x = 162.9 + \frac{12 - 2}{11} \times 76.5 = 232.44$$

محاسبه صدک ششم بارش اهواز :

محاسبه براساس صدک ششم به روش سری طبقه بندی انجام می گیرد. مراحل :

با توجه به مرتب کردن سری :

$$p_{60} = \frac{60(20 + 1)}{100} = 12.6$$

$$p_{60} = \frac{228.2 + 234.8}{2} = 231.5$$

محاسبه براساس سری طبقه بندی شده :

$$p_x = L + \frac{\frac{xN}{100} - F}{f} \times C$$

$$p_{60} = \frac{60 \times 20}{100} = 12$$

$$p_{60} = 162.9 + \frac{12 - 2}{11} \times 76.5 = 232.44$$

محاسبه دهک های بارش اهواز با نرم افزار SPSS :

Statistics

prec

N	Valid	20
	Missing	0
Percentiles	10	139.4900
	20	174.0400
	30	197.3100
	40	221.3800
	50	227.5500
	60	232.1600
	70	265.5300
	80	299.9800
	90	402.0100

محاسبه صدک های بارش اهواز با استفاده از نرم افزار SPSS :

Statistics

Precipitation of Ahvaz

N	Valid	20
	Missing	0
Percentiles	25	185.7750
	75	285.3750



Statistics

Precipitation of Ahvaz

N	Valid	20
	Missing	0
Percentile	25	185.7750
s	75	285.3750
	95	465.8750

محاسبه صدک های بارش اهواز با استفاده از نرم افزار NCSS :

NTS AND SETTINGS\JAVARIMY DOCUMENTS\ESF.S0

Percentile Section of Precipitation of the Ahvaz

Percentile	Value	95% LCL	95% UCL	Exact Conf. Level
99.5	468.8			
95.0	465.875			
90.0	402.01			
85.0	323.665	234.8	468.8	95.5319
80.0	299.98	234.8	468.8	95.6328
75.0	285.375	227.8	410.3	96.1823
70.0	265.53	227.3	410.3	97.5218
65.0	245.265	224.8	327.4	96.8303
60.0	232.16	219.1	302.5	96.3010
55.0	228.02	218.8	289.9	95.9722
50.0	227.55	188.1	271.8	95.8611
45.0	225.925	185	250.9	95.9722
40.0	221.38	171.3	234.8	96.3010
35.0	218.905	170.9	228.2	96.8303
30.0	197.31	136	227.8	97.5218
25.0	185.775	86.4	224.8	95.5904
20.0	174.04	86.4	219.1	95.6328
15.0	170.96	86.4	219.1	95.5319
10.0	139.49			

محاسبه دامنه میان چارکی سری های اقلیمی :

دامنهٔ میان چارکی<sup>1</sup> دامنهٔ تغییرات 50 درصدی سری ها می باشد . در تعریف دامنهٔ میان چارکی ، میانه را که 50 درصد داده است ، ملاک قرار داده و از پایین تا 25 درصد و از بالا تا 75 درصد گسترش می دهند . به عبارت دیگر دامنهٔ میان چارکی شامل چارک اول تا سوم می شود . حاصل تفاضل چارک سوم از چارک اول دامنهٔ میان چارکی بدست می آید .  
یعنی :

$$IQR = Q_3 - Q_1$$

در همین راستا مفهوم انحراف چارکی<sup>2</sup> هم مطرح می شود این شاخص میزان پراکندگی در اطراف مرکز توزیع متغیرهای اقلیمی را نشان می دهد. برای بدست آوردن انحراف چارکی باید ابتدا چارک اول و چارک سوم را محاسبه کرد . با میانگین تفاضل چارک اول و سوم انحراف چارکی بدست می آید . یعنی :

$$Q_D = \frac{Q_3 - Q_1}{2}$$

برای فهم بهتر دامنهٔ میان چارکی به موقعیت زیر توجه نمایید:

موقعیت شمارهٔ :

دامنهٔ میان چارکی بارش اهواز را محاسبه نمایید؟

محاسبهٔ دامنهٔ میان چارکی و انحراف چارکی بارش اهواز با استفاده از نرم افزار **Minitab** :

*Descriptive Statistics: rain*

Variable	Q1	Q3	IQR
rain	185.8	285.4	99.6

$$Q_D = \frac{285.4 - 185.8}{2} = 49.8$$

یا

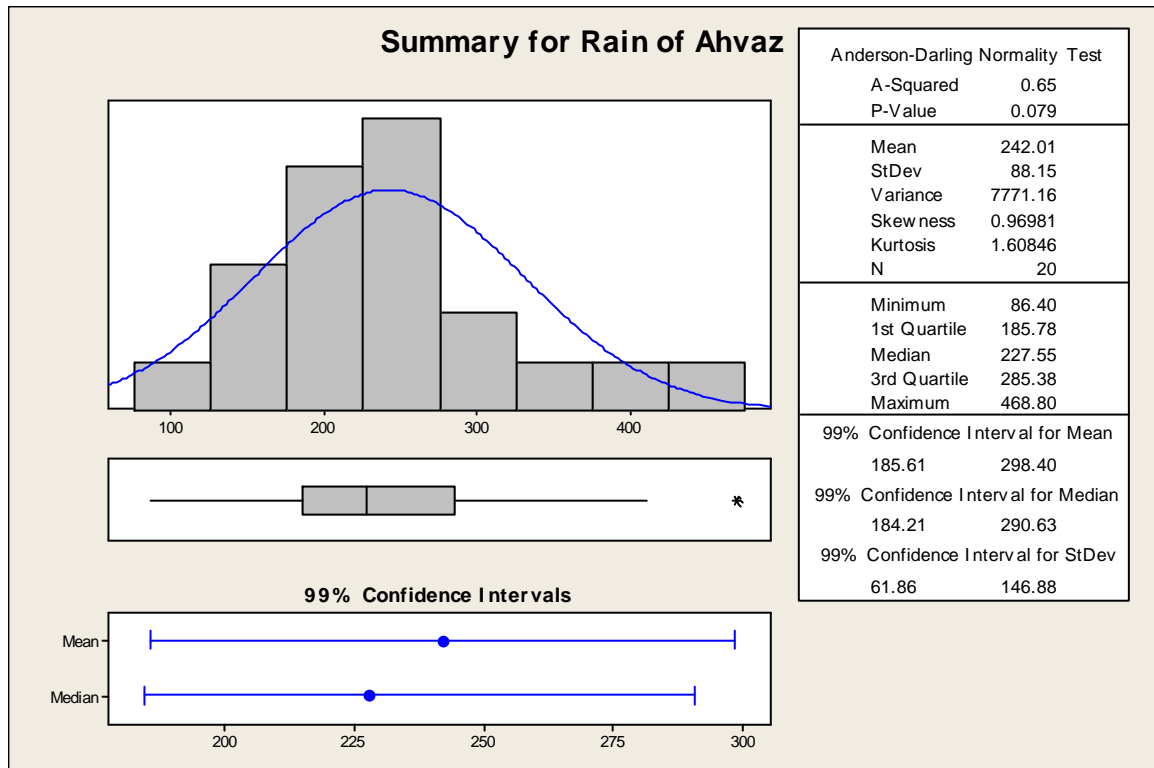
---

<sup>1</sup> - Interquartile range

<sup>2</sup> - Quartile Deviation

$$Q_D = \frac{IQR}{2} = \frac{99.6}{2} = 49.8$$

لذا بارش اهواز نشان می دهد که حدود **49/8** درصد از مرکز توزیع سری بارش اهواز انحراف دارد. یعنی :



محاسبه نمای متغیرهای اقلیمی :

نما یا مد<sup>1</sup> در یک توزیع آماری مقداری است که بیشترین تکرار یا فراوانی را در میان متغیرها یا سری داشته باشد. لذا این آماره برای نمایش داده هایی استفاده می شود که نسبت به دیگر داده ها ، بیشتر تکرار می شوند . به عبارت دیگر نما برای متغیرهای گسسته چنین تعریف می شود که بالاترین فراوانی را در نمونه یا جامعه دارد. اگر سری کوتاه و کم باشد داده هایی که بیشتر از سایر داده ها تکرار شده باشد ، مد است . اما در توزیع هایی که طبقه بندی شده باشد متغیری که فراوانی آن حداکثر باشد ، مد است. در بعضی سری ها ممکن است دو فراوانی حداکثری وجود داشته باشد در آن صورت توزیع دو مدی است . همچنین ممکن است توزیع بیشتر از دو مد داشته باشد که در این

صورت آن را توزیع چند نمایی می خوانند. نما آماره ای است که تعیین شاخص مرکزی به صورت تقریبی و سریع باشد. برای محاسبه نما به چند روش باید توجه نمود که عبارتند از :

- سری ساده
  - سری طبقه بندی شده دو نمایی
  - سری طبقه بندی شده
- برای فهم بهتر به موقعیت شماره توجه نمایید.  
موقعیت شماره :

اگر سری بارش 20 ساله (1986 تا 2005) در دست باشد ، نمای آن را حساب نمایید؟

	بارش	ردیف	بارش
1	297.90	11	176.70
2	112.10	12	236.20
3	138.90	13	88.90
4	115.50	14	228.50
5	81.30	15	<u>155.50</u>
6	294.40	16	191.30
7	153.80	17	114.60
8	177.60	18	<u>115.50</u>
9	227.00	19	208.00
10	147.00	20	192.00

مد در سری بارش ایستگاه آبادان 115.5 می باشد.

محاسبه نما با استفاده از نرم افزار Excel :

Column2 | Column1

	297.9
	112.1
	138.9
	115.5
	81.3
	294.4
	153.5
	177.6
	227
	147
	176.7
	236.2
	88.9
	228.5
	155.5
	191.3
	114.6
	115.5
	208
	192
<i>mode =</i>	115.5

محاسبه نما با سری های اقلیمی طبقه بندی شده:

در روش فراوانی های طبقه بندی شده به علت سری زمانی طولانی مورد تحلیل در اقلیم شناسی از اهمیت زیادی دارد. برای محاسبه سری طبقه بندی شده از رابطه شماره استفاده می شود.

$$MO = L + \frac{D_1}{D_1 + D_2} \times C$$

$D_1$  تفاضل فراوانی ماقبل طبقه نمادار از فراوانی طبقه نمادار و  $D_2$  تفاضل فراوانی طبقه مابعد طبقه نمادار از فراوانی طبقه نمادار می باشد.

برای تبیین بهتر مطلب لازم است به موقعیت شماره اشاره شود.

موقعیت شماره :

اگر سری بارش 20 ساله ایستگاه آبادان به صورت جدول زیر در دست باشد ، نمای آن را محاسبه نمایید؟

مراحل:

$$- \text{ تعیین دامنه تغییرات } (R = (297.9 - 81.3) = 216.6)$$

$$- \text{ تعیین تعداد طبقات } (k = \sqrt{N} = \sqrt{20} = 4.47 = 5)$$

$$- \text{ فاصله طبقات } (C = 216.6 \div 5 = 43.32)$$

طبقات	فراوانی
81/3 124/62	6
124/62 167/94	4
167/94 211/26	5
211/26 254/58	3
254/58 - 297/9	2
جمع	20

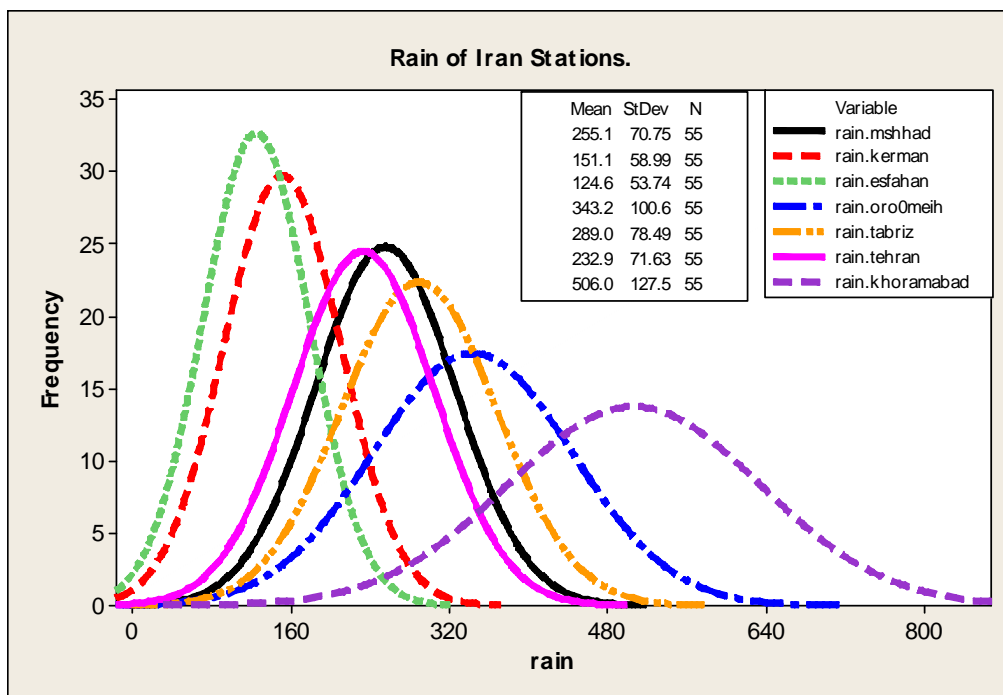
$$D_1 = 6 - 0 = 6$$

$$D_2 = 6 - 4 = 2$$

$$MO = 81.3 + \frac{6}{6+2} \times 43.32 = 113.79$$

به هر حال نما از شاخص های گرایش به مرکز است که برآورد آن سریع است و با تغییر در فراوانی ها مقدار آن تغییر خواهد کرد. رابطه بین

میانگین ، میانه ومد در شرایط بهنجاری سری منطبق بر هم می باشند. لذا میانگین < میانه < نما می باشد. به نمودار شماره توجه شود.



توزیع متفاوت بارش ایستگاه های ایران متناسب با اندازه گرایش به مرکز میانگین با میزان پراکنش ها در نمودار بالا قابل توجه است.

فصل چهارم: بررسی شاخص های پراکندگی متغیرهای اقلیمی

اندازه گیری شاخص های پراکندگی متغیرهای اقلیمی :

با توجه به پراکندگی و توزیع متغیرهای اقلیمی در مقیاس های زمانی و مکانی ، بررسی شاخص های پراکندگی در تحلیل های آماری اقلیم

از اهمیت زیادی برخوردار است . از مهمترین شاخص های پراکندگی می توان به آماره های زیر اشاره نمود:

- دامنه تغییرات
- انحراف از میانگین
- واریانس
- انحراف معیار
- نمره های معیار
- کجی یا چولگی
- کشیدگی
- گشتاورها
- ضریب تغییرات
- نمره های معیار

محاسبه دامنه تغییرات متغیرهای اقلیمی :

دامنه تغییرات<sup>1</sup> حاصل تفاضل بین بزرگترین و کوچکترین متغیرهای مشاهده شده می باشد . دامنه تغییرات را برای داده های کمی پیوسته و

گسسته را به ترتیب با فرمول های زیر اندازه گیری می نمایند:

$$R = \text{MAX}_d - \text{MIN}_d + 1$$

$$R = \text{MAX}_d - \text{MIN}_d$$

---

<sup>1</sup> - Range



هرچه مقدار دامنه تغییرات بیشتر باشد ، نشانگر پراکندگی بیشتر متغیرها می باشد و برعکس . نکته مهم در بررسی این شاخص این است که در تحلیل فقط از دو متغیر بزرگتر و کوچکتر استفاده می شود و تغییرات آن بستگی به حجم نمونه دارد . چون این شاخص قبلا مورد بررسی قرار گرفته است در این قسمت فقط بررسی متغیرها با استفاده از نرم افزار نشان داده می شود. برای فهم این شاخص به موقعیت شماره 1 توجه نمایید.

موقعیت شماره 1 :

اگر سری فشار 53 ساله (1951 – 2003) ایستگاه اصفهان در دست باشد ، دامنه تغییرات سری فشار آن ایستگاه را حساب کنید؟

سال	فشار	سال	فشار
1951	1008.70	1978	1005.20
1952	1009.30	1979	1006.40
1953	1008.70	1980	1005.20
1954	1006.00	1981	1005.60
1955	1005.10	1982	1005.40
1956	1006.70	1983	1005.60
1957	1008.70	1984	1005.80
1958	1004.90	1985	1004.90
1959	1003.60	1986	1006.40
1960	1004.90	1987	1006.30
1961	1004.80	1988	1004.40
1962	1006.50	1989	1005.80
1963	1003.00	1990	1004.80
1964	1005.50	1991	1005.00
1965	1005.70	1992	1006.00

1966	1005.70	1993	1006.40
1967	1004.80	1994	1004.68
1968	1005.40	1995	1010.50
1969	1005.60	1996	1011.00
1970	1004.40	1997	1009.40
1971	1004.90	1998	1010.20
1972	1006.90	1999	1009.50
1973	1005.30	2000	1009.40
1974	1005.70	2001	1009.20
1975	1005.50	2002	1009.30
1976	1005.30	2003	1009.30
1977	1005.10		

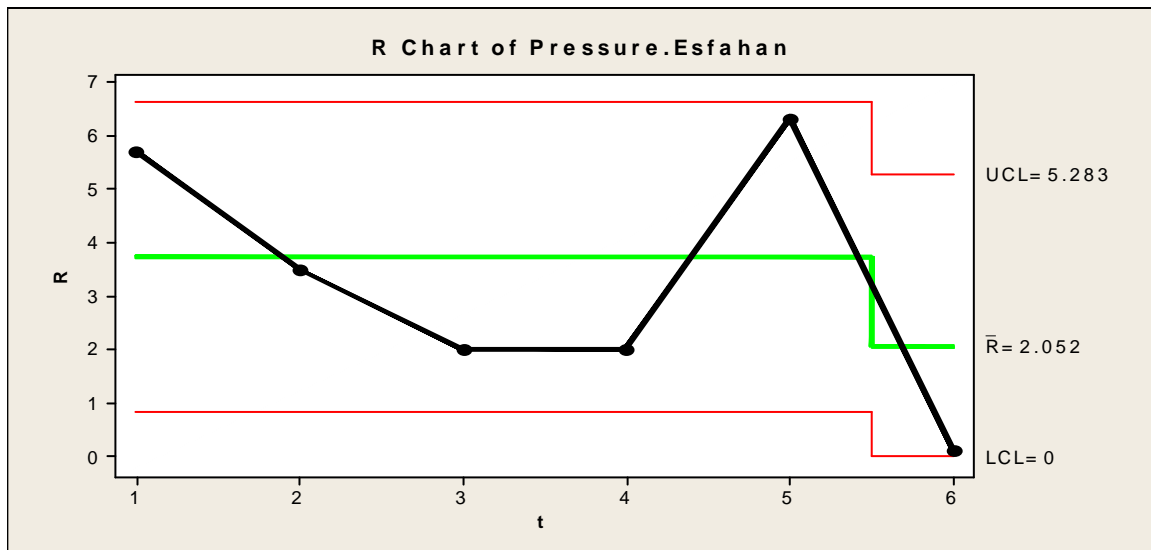
دامنه تغییرات سری با استفاده از نرم افزار SPSS :

#### Descriptive Statistics

	<i>N</i>	<i>Range</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>
<i>pressure</i>	53	8.00	1003.00	1011.00	1006.3845
<i>Valid N (listwise)</i>	53				

مطلب مهمی که در ارتباط با دامنه تغییرات متغیرهای اقلیمی باید به آن توجه نمود این است که حدود دامنه تغییرات در مطالعات و تحقیقات اقلیمی اهمیت زیادی دارد بر این اساس در یک دوره آماری محاسبه و در نظر گرفتن حدود اطمینان دامنه تغییرات ضروری است. با توجه به

موقعیت شماره  $I$  میانگین و حدود اطمینان دامنه تغییرات فشار ایستگاه اصفهان قابل توجه است. در طول دوره  $53$  ساله توزیع فشار ایستگاه مذکور میانگین دامنه تغییرات فشار برابر با  $2/05$  هکتوپاسکال می باشد. یعنی به طور متوسط در دوره مذکور حدود  $2/05$  هکتوپاسکال فشار ایستگاه اصفهان تغییر نموده است. از طرفی در طول آن دوره حدود اطمینان دامنه فشار تا  $5/283$  هکتوپاسکال امکان تغییرات را داشته است. نمودار شماره موضوع مذکور را مشخص می نماید. از نکاتی دیگری که در این راستا می توان به آن اشاره نمود آن است که میل کاهش دامنه تغییرات فشار ایستگاه اصفهان در پایان دوره بیشتر شده است. یعنی دامنه تغییرات فشار در ایستگاه مذکور به طرف پایان دوره کاهش یافته است.



لذا نمودار بالا دامنه فشار را در طول دوره سری نشان می دهد. علاوه بر نکاتی که اشاره شد نکته مهم دیگری که باید به آن اشاره نمود این است که نمودار دامنه تغییرات، توزیع دامنه فشار در طول دوره نشان می دهد. نکته دیگر این است که دامنه فشار در طول سری  $8$  هکتوپاسکال ( $1103 - 1014 = 8$ ) می باشد در حالی که توزیع دامنه فشار در طول دوره  $5/283$  هکتوپاسکال نشان داده می شود.

محاسبه انحراف از میانگین متغیرهای اقلیمی:

انحراف از میانگین<sup>1</sup> چگونگی توزیع متغیرها را نسبت به میانگین نشان می دهد. انحراف از میانگین دامنه مثبت و منفی دارد که در این شاخص قدرمطلق آن در نظر گرفته می شود. برای محاسبه انحراف از میانگین با توجه به فراوانی های طبقه بندی شده و طبقه بندی نشده آنها قابل محاسبه می باشد. لذا برای محاسبه سری متغیرهای اقلیمی طبقه بندی نشده می توان از رابطه زیر استفاده کرد:

<sup>1</sup> - Average Deviation

$$A_D = \frac{\sum |X_i - \bar{X}|}{N}$$

برای سری متغیرهای طبقه بندی شده :

$$A_D = \frac{\sum f_i |X_i - \bar{X}|}{N}$$

در مورد کاربرد انحراف از میانگین سری باید به قدرمطلق انحرافات توجه داشت . یعنی :

$$e_i = |X_i - \bar{X}|$$

برای فهم بهتر به موقعیت شماره 2 اشاره می شود.

موقعیت شماره 2 :

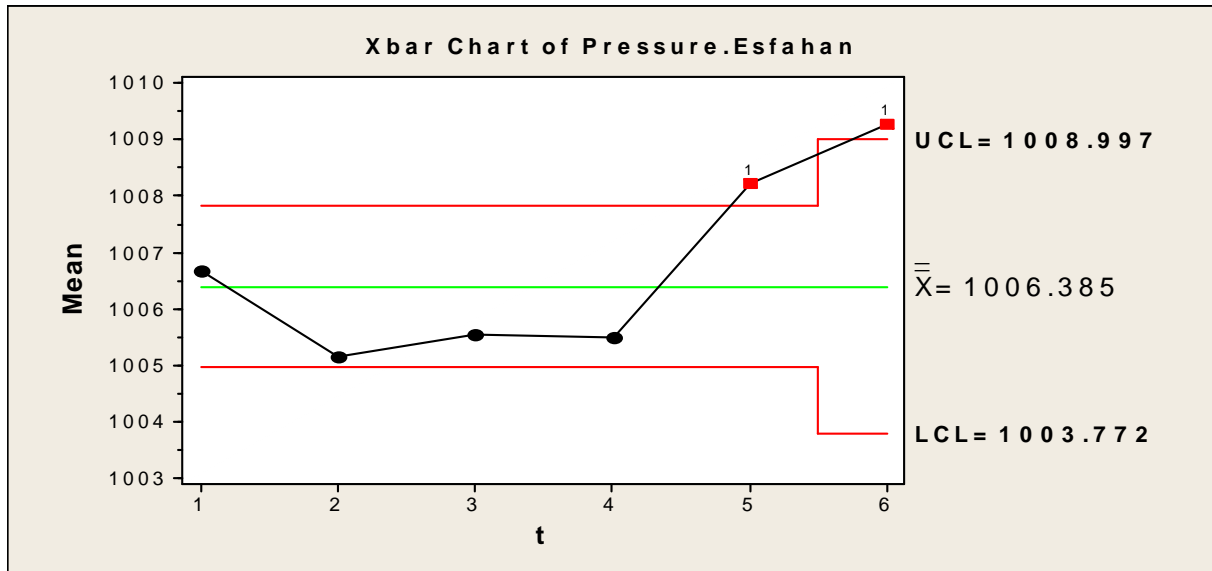
اگر سری فشار ایستگاه اصفهان در دست باشد ، انحراف از میانگین سری مذکور را به طرق مختلف محاسبه نمایید؟

سال	فشار	سال	فشار	قدرمطلق انحراف از میانگین	قدرمطلق انحراف از میانگین
1951	1008.70	1978	1005.20	2.316	1.184
1952	1009.30	1979	1006.40	2.916	0.016
1953	1008.70	1980	1005.20	2.316	1.184
1954	1006.00	1981	1005.60	0.384	0.784
1955	1005.10	1982	1005.40	1.284	0.984
1956	1006.70	1983	1005.60	0.316	0.784
1957	1008.70	1984	1005.80	2.316	0.584
1958	1004.90	1985	1004.90	1.484	1.484
1959	1003.60	1986	1006.40	2.784	0.016
1960	1004.90	1987	1006.30	1.484	0.084
1961	1004.80	1988	1004.40	1.584	1.984

1962	1006.50	1989	1005.80	0.116	0.584
1963	1003.00	1990	1004.80	3.384	1.584
1964	1005.50	1991	1005.00	0.884	1.384
1965	1005.70	1992	1006.00	0.684	0.384
1966	1005.70	1993	1006.40	0.684	0.016
1967	1004.80	1994	1004.68	1.584	1.704
1968	1005.40	1995	1010.50	0.984	4.116
1969	1005.60	1996	1011.00	0.784	4.616
1970	1004.40	1997	1009.40	1.984	3.016
1971	1004.90	1998	1010.20	1.484	3.816
1972	1006.90	1999	1009.50	0.516	3.116
1973	1005.30	2000	1009.40	1.084	3.016
1974	1005.70	2001	1009.20	0.684	2.816
1975	1005.50	2002	1009.30	0.884	2.916
1976	1005.30	2003	1009.30	1.084	2.916
1977	1005.10			1.284	
				جمع	82/38

$$A_D = \frac{\sum |82.38|}{53} = 1.55$$

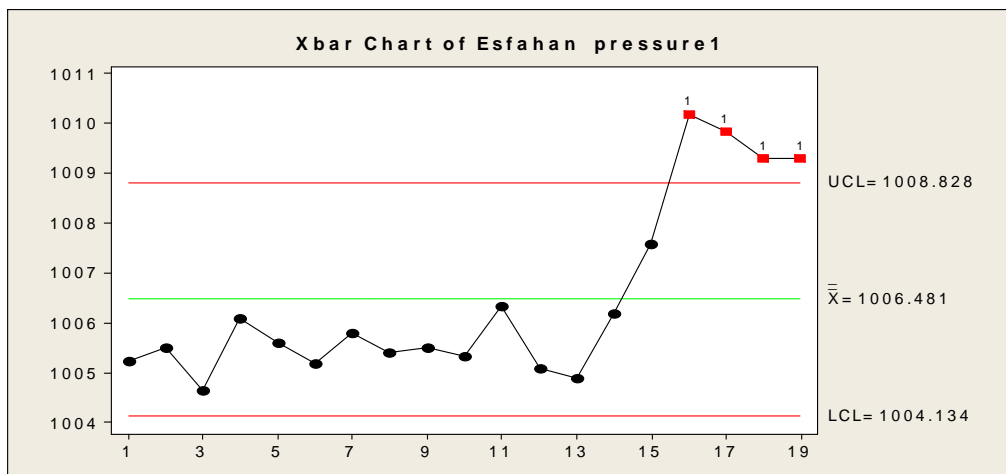
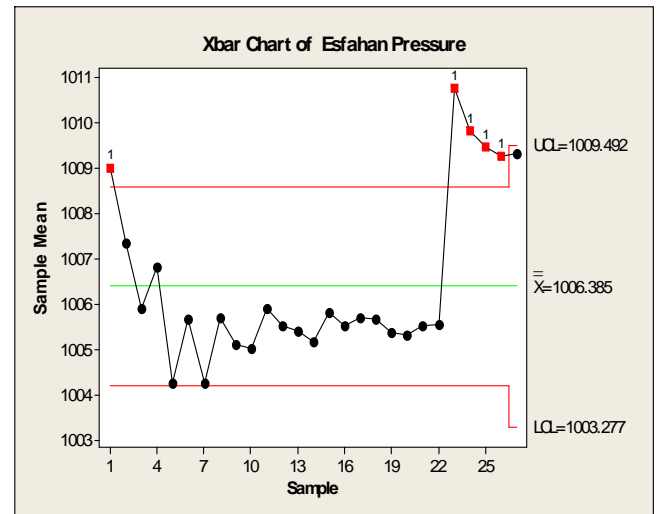
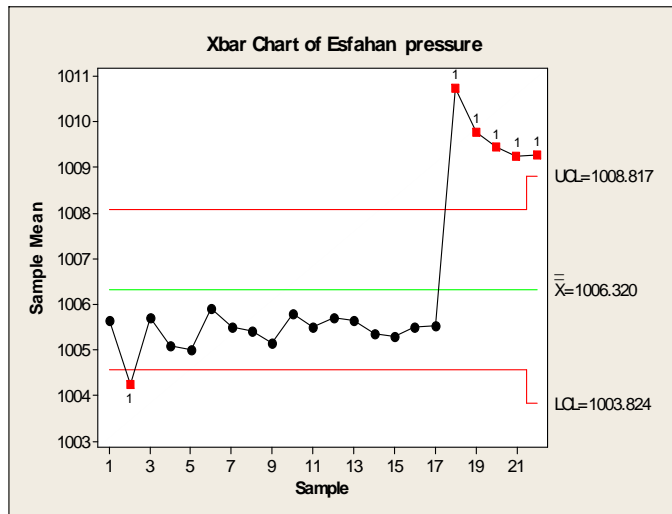
در تحقیقات اقلیمی بررسی میزان انحراف میانگین و حدود اطمینان آن اهمیت زیادی دارد . متناسب با موقعیت مورد نظر میزان انحراف از میانگین سری فشار ایستگاه اصفهان  $1/55$  هکتوپاسکال می باشد . از نکات مورد نظر در بررسی حدود اطمینان انحراف از میانگین سری مذکور میل افزایشی میزان انحراف از میانگین در پایان دوره می باشد .



نمودار بالا توزیع فشار را در طول دوره همراه با دامنه انحراف از میانگین نشان می دهد . نکته مهمی که باید در سری فشار ایستگاه اصفهان

به آن توجه نمود این است که علاوه بر میزان انحراف از میانگین سری که **1/55** هکتوپاسکال است میزان توزیع زمانی دامنه انحراف آن

سری بیش از **5/225** هکتوپاسکال را نشان می دهد. به نمودار شماره توجه شود:



نمودارهای بالا توزیع فشار ایستگاه اصفهان در دوره های مختلف نشان می دهند. نکته مهمی که در نمودارها می توان در قالب میزان دامنه حدود اطمینان انحراف از میانگین به آن اشاره نمود این است که با کاهش طول دوره فشار اصفهان میزان دامنه حدود اطمینان انحراف از میانگین کاهش یافته است. ولی آرایش تغییرات حدود اطمینان انحراف از میانگین تغییرات زیادی از خود نشان نمی دهد.

محاسبه انحراف از میانگین با استفاده از نرم افزار **NCSS**:

*Mean-Deviation Section of pressure*

Parameter	$ X-Mean $	$ X-Median $	$(X-Mean)^2$	$(X-Mean)^3$	$(X-Mean)^4$
Average	<u>1.554489</u>	1.413585	3.679919	5.995724	35.24923

محاسبه انحراف از میانگین با استفاده از نرم افزار **Excel**:

شاخص آماری	فشار
	1008.7
	1009.3
	1008.7
	1006
	1005.1
	1006.7
	1008.7
	1004.9
	1003.6
	1004.9
	1004.8
	1006.5
	1003
	1005.5
	1005.7
	1005.7
	1004.8
	1005.4
	1005.6
	1004.4
	1004.9
	1006.9
	1005.3
	1005.7
	1005.5
	1005.3
	1005.1
	1005.2
	1006.4
	1005.2
	1005.6
	1005.4
	1005.6
	1005.8
	1004.9
	1006.4



	1006.3
	1004.4
	1005.8
	1004.8
	1005
	1006
	1006.4
	1004.68
	1010.5
	1011
	1009.4
	1010.2
	1009.5
	1009.4
	1009.2
	1009.3
	1009.3
میانگین	1006.3845
انحراف از میانگین	1.554489

محاسبه انحراف از میانگین سری های طبقه بندی شده اقلیمی :

برای سری فشار اصفهان :

مراحل :

- تعیین دامنه تغییرات (  $R = (1011 - 1003) = 8$  )

- تعیین تعداد طبقات (  $k = \sqrt{N} = \sqrt{53} = 7.27 - 7$  )

- فاصله طبقات (  $C = 8 \div 7 = 1.143$  )

طبقات	فراوانی	میانه طبقات	قدرمطلق انحراف از میانگین	$f_i  X_i - \bar{X} $
1003-1004/143	2	1003/571	2/813	5/626

1004/143- 1005/286	14	1004/714	1/67	23/38
1005/628- 1006/429	21	1005/857	0/527	11/067
1006/942- 1007/572	3	1007	0/616	1/848
1007/572- 1008/715	3	1008/143	1/759	5/277
1008/715- 1009/858	7	1009/286	2/902	20/314
1009/858- 1011/001	3	1010/429	4/045	12/135
جمع	53			79/647

$$A_D = \frac{79.647}{53} = 1.5$$

در این شاخص بعلاوه این که قدر مطلق، سری ها در نظر گرفته می شود و اثر انحرافات بزرگ در مطالعات اقلیمی پیشرفته را مشخص نمی سازد، لذا کاربرد گسترده ای در مطالعات اقلیمی ندارد.

محاسبه انحراف از میانگین با استفاده از نرم افزار **NCSS**:

### Descriptive Statistics Report

#### Mean-Deviation Section of pressure

Parameter	X-Mean	X-Median	(X-Mean)^2	(X-Mean)^3	(X-Mean)^4
Average	1.55448914204434		1.41358490566038	3.67991911712353	
	5.99572357193339		35.2492317474454		
Std Error	0.160173054492299			0.639979761455859	
	1.42232614245681		8.4474672598398		

محاسبه واریانس متغیر های اقلیمی :

واریانس<sup>1</sup> یکی از مهمترین شاخص های اندازه گیری تغییرپذیری در مطالعات اقلیمی است که از تقسیم مجموع مجذور انحراف ها بر تعداد سری بدست می آید. به همین دلیل گاهی اوقات واریانس را میانگین مجذورها می نامند. واریانس میزان پراکندگی سری ها در اطراف میانگین را نشان می دهد. هر چه واریانس سری ها بیشتر باشد تشابهات سری ها کمتر است و سری ها از میانگین دورتر هستند. اگر سری داده ها با هم برابر باشند واریانس آنها صفر خواهد بود. روش های متعددی برای محاسبه واریانس وجود دارد که عبارتند از:

- محاسبه واریانس برای سری بدون فراوانی طبقه بندی شده
- محاسبه واریانس برای سری دارای فراوانی طبقه بندی نشده
- محاسبه واریانس برای سری دارای فراوانی طبقه بندی شده

محاسبه واریانس برای سری بدون فراوانی طبقه بندی شده :

برای محاسبه واریانس سری بدون فراوانی طبقه بندی شده می توان از رابطه زیر استفاده کرد:

$$V = \frac{\sum (X - \bar{X})^2}{N - 1} \quad \text{واریانس برای نمونه}$$

$$V = \frac{\sum (X - \bar{X})^2}{N} \quad \text{واریانس برای جامعه}$$

برای فهم بهتر مطلب به موقعیت شماره 3 اشاره می شود:

موقعیت شماره 3:

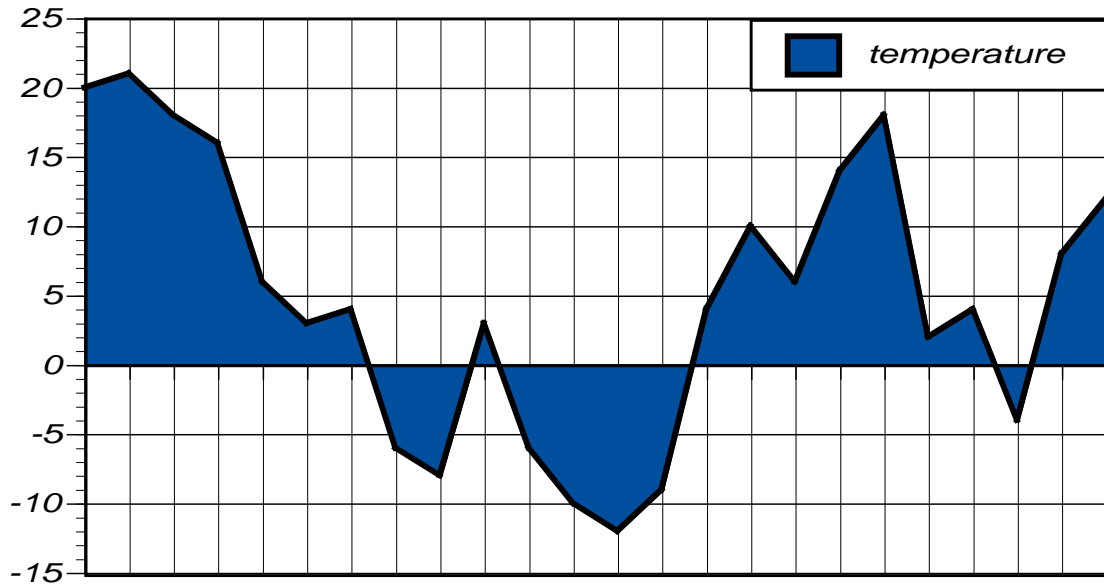
اگر سری میانگین 20 ساله (1986-2005) دمای ایستگاه ایلام در دست باشد واریانس آن سری را حساب کنید؟

سری دما	انحراف از میانگین سری	مجذور انحرافات
19.20	2.065	4.264
17.90	0.765	0.585
17.50	0.365	0.133

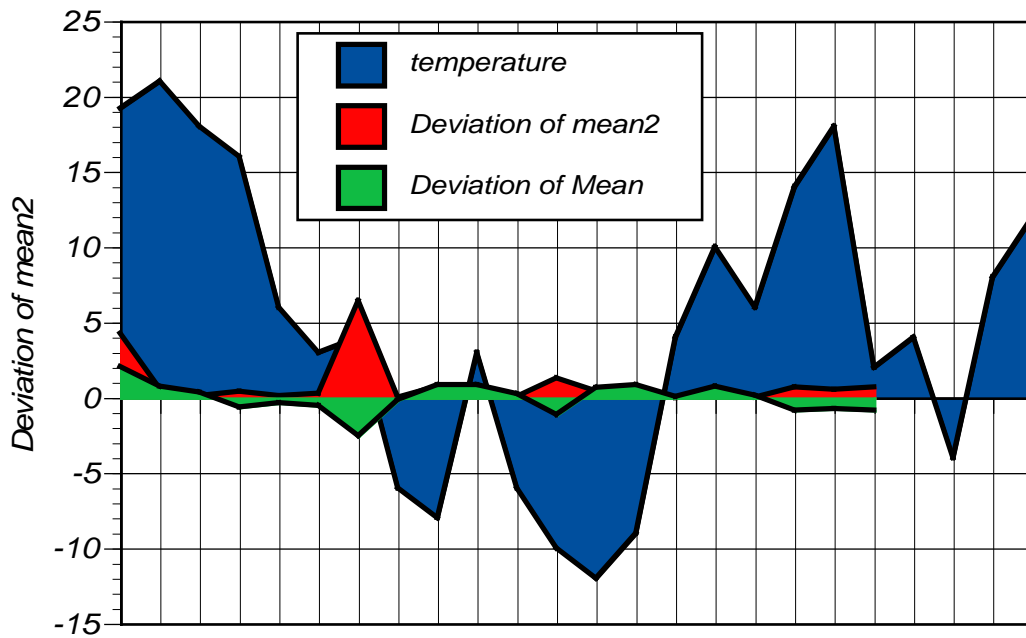
16.50	-0.635	0.403
16.80	-0.335	0.112
16.60	-0.535	0.286
14.60	-2.535	6.426
17.00	-0.135	0.018
18.00	0.865	0.748
18.00	0.865	0.748
17.40	0.265	0.070
16.00	-1.135	1.288
17.80	0.665	0.442
18.00	0.865	0.748
17.20	0.065	0.004
17.90	0.765	0.585
17.30	0.165	0.027
16.30	-0.835	0.697
16.40	-0.735	0.540
16.30	-0.835	0.697
342/7		18/825

$$V = \frac{18.825}{19} = 0.99$$

لذا :



اگر میانگین سری دمای ایلام با مجذور انحراف از میانگین سری مقایسه شود ، وضعیت واریانس سری دمای ایلام مشخص تر می شود. لذا:



محاسبه واریانس با استفاده از نرم افزار SPSS :

## Statistics

m.temp

N	Valid	20
	Missing	0
	Variance	<u>.991</u>

محاسبه واریانس برای سری دارای فراوانی طبقه بندی نشده :

برای محاسبه واریانس سری دارای فراوانی طبقه بندی نشده می توان از رابطه زیر استفاده کرد:

$$V = \frac{\sum f_i (X - \bar{X})^2}{N - 1} \quad \text{واریانس برای نمونه}$$

$$V = \frac{\sum f_i (X - \bar{X})^2}{N} \quad \text{واریانس برای جامعه}$$

برای بررسی روش بالا به موقعیت شماره 4 اشاره می شود.

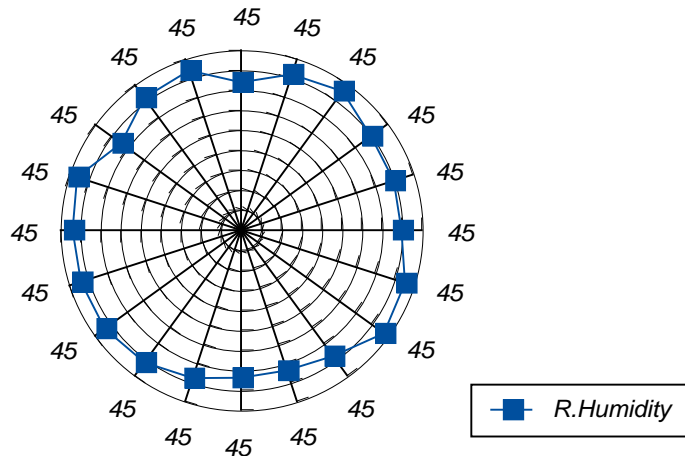
موقعیت شماره 4 :

اگر سری 20 ساله (2005 1986) رطوبت نسبی ایلام در دست باشد ، واریانس آن سری را حساب کنید؟

سری	فراوانی	$X.f_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$
37.00	4	148	- 3/5	12/25	49
39.00	2	78	- 1/5	2/25	4/5
40.00	3	120	- 0/5	0/25	0/75
41.00	3	123	0/5	0/25	0/75
42.00	4	168	1/5	2/25	9
43.00	3	129	2/5	6/25	18/75
44.00	1	44	3/5	12/25	12/25
جمع	20	810			95

$$\bar{X} = \frac{810}{20} = 40.5$$

$$V = \frac{95}{19} = 5$$



محاسبه واریانس با استفاده از نرم افزار SPSS :

*Statistics*

<i>N</i>	<i>Valid</i>	<i>20</i>
	<i>Missing</i>	<i>0</i>
	<i>Variance</i>	<u><i>5.000</i></u>

**Humidity of Ilam**

	<i>Frequenc y</i>	<i>Percent</i>	<i>Valid Percent</i>	<i>Cumulative Percent</i>
<b>37.00</b>	<b>4</b>	<b>20.0</b>	<b>20.0</b>	<b>20.0</b>
<b>39.00</b>	<b>2</b>	<b>10.0</b>	<b>10.0</b>	<b>30.0</b>

40.00	3	15.0	15.0	45.0
41.00	3	15.0	15.0	60.0
42.00	4	20.0	20.0	80.0
43.00	3	15.0	15.0	95.0
44.00	1	5.0	5.0	100.0
جمع	20	100.0	100.0	

محاسبه واریانس برای سری دارای فراوانی طبقه بندی شده :

برای محاسبه واریانس سری دارای فراوانی طبقه بندی شده می توان از رابطه زیر استفاده کرد:

$$V = \frac{\sum f_i (C_p - \bar{X})^2}{N - 1} \quad \text{واریانس نمونه}$$

$$V = \frac{\sum f_i (C_p - \bar{X})^2}{N} \quad \text{واریانس جامعه}$$

برای محاسبه واریانس به روش بالا به موقعیت شماره 5 اشاره می شود.

موقعیت شماره 5:

اگر سری 20 ساله (2005 - 1986) بارش ایستگاه اسلام آباد غرب در دست باشد ، واریانس آن را حساب کنید؟

مراحل:

- تعیین دامنه تغییرات (  $R = (761 - 293.2) = 467.8$  )

- تعیین تعداد طبقات (  $k = \sqrt{N} = \sqrt{20} = 4.47 = 4$  )

- فاصله طبقات (  $C = 467.8 \div 4 = 116.95$  )

طبقات	فراوانی	Cp	F.Cp	Cp - $\bar{X}$	$f_i (Cp - \bar{X})^2$
293/2 410/15	6	351/675	2110/05	-134/492	108528/59



410/15- 527/1	6	468/625	2811/75	17/725	1885/05
- 644/05 527/1	7	585/575	4099/025	99/408	69173/65
644/05-761	1	702/55	702/525	216/358	46810/78
جمع	20		9723/35		226398/07

$$\bar{X} = \frac{9723.35}{20} = 486.167$$

$$V = \frac{226398.07}{19} = 11915.69$$

محاسبه واریانس سری ها با نرم افزار SPSS :

#### Descriptive Statistics

	N	Mean	Variance
Precipitation of Ilam	1005	475.9663	11912.144
Valid N (listwise)	1005		

محاسبه انحراف معیار متغیرهای اقلیمی :

یکی دیگر از آماره های تغییرپذیری در تحلیل های آماری اقلیم انحراف معیار<sup>1</sup> می باشد. انحراف معیار ریشه دوم یا جذر واریانس می باشد. انحراف معیار همانند واریانس میزان پراکندگی متغیرهای اقلیمی پیرامون میانگین را نشان می دهد. بر این اساس هرچه انحراف معیار کوچکتر باشد نشانگر این واقعیت است که نمونه ها یا سری های مورد بررسی و سنجش متجانس تر و همگن تر هستند و تغییرپذیری آنها کمتر است. یعنی اختلاف بین سری های اقلیمی کمتر می باشد و میانگین پراکندگی متغیرها را بهتر نشان می دهد و برعکس. هرگاه به

<sup>1</sup> - Standard Deviation

اندازه یک انحراف معیار از میانگین کم یا به آن اضافه کنند، فاصله ای از  $X_1 = \bar{X} - S_x$  تا  $X_2 = \bar{X} + S_x$  بدست می آید. مساحتی از زیر منحنی مابین این فاصله و محور افقی **68/27** درصد از فراوانی کل را در بر دارد. هرگاه به میانگین به اندازه **2** برابر انحراف معیار کم یا اضافه شود، فاصله ای از  $X_1 = \bar{X} - 2S_x$  تا  $X_2 = \bar{X} + 2S_x$  بدست می آید. مساحتی که از سطح زیر منحنی مابین این فاصله و محور افقی **95/45** درصد از فراوانی کل را در بر دارد. هرگاه به میانگین به اندازه **3** برابر انحراف معیار کم یا اضافه شود، فاصله ای از  $X_1 = \bar{X} - 3S_x$  تا  $X_2 = \bar{X} + 3S_x$  بدست می آید. مساحتی از سطح زیر منحنی مابین این فاصله و محور افقی عموماً **99/73** درصد از فراوانی کل را در بر خواهد گرفت. هرگاه به میانگین به اندازه **4** برابر انحراف معیار کم یا اضافه شود، فاصله ای از  $X_1 = \bar{X} - 4S_x$  تا  $X_2 = \bar{X} + 4S_x$  بدست می آید، مساحتی از سطح زیر منحنی مابین فاصله و محور افقی **99/99** درصد از فراوانی کل را در بر خواهد گرفت. به عبارتی سطحی از یک منحنی توزیع نرمال (نمره معیار) را شامل می شود. برای تبیین این موضوع به موقعیت شماره **6** توجه نمایید.

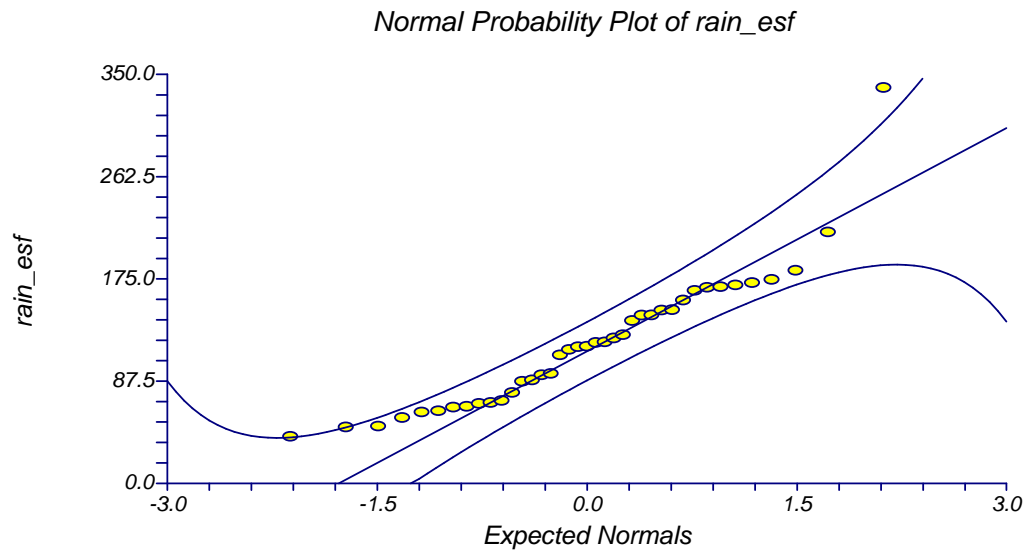
موقعیت شماره **6** :

دو سری بارش اصفهان و کاشان **39** ساله (**1951 - 1989**) را در نظر بگیرید، میزان پراکندگی متغیرها را مشخص نمایید /

#### Variation Section of rain\_esf

Parameter	Variance	Standard Deviation
Value	3330.511	<u>57.71058</u>

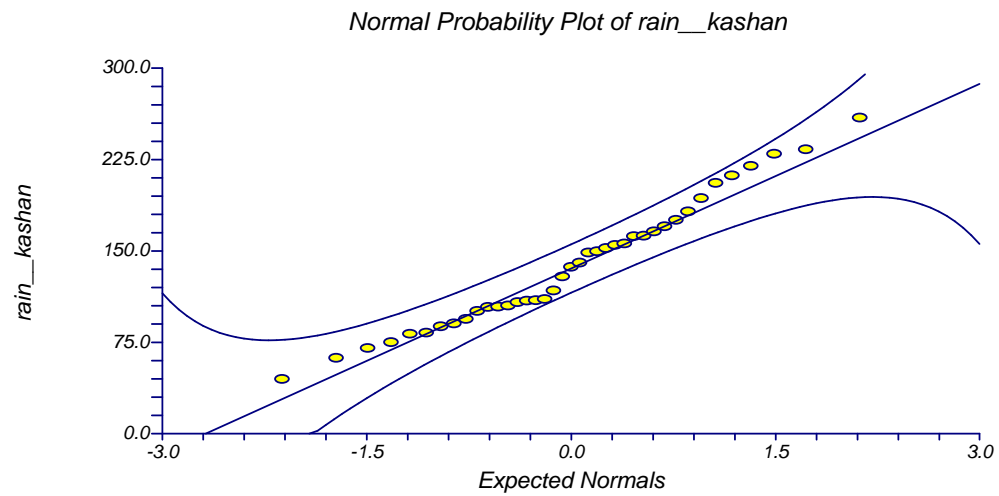
#### Plots Section of rain\_esf



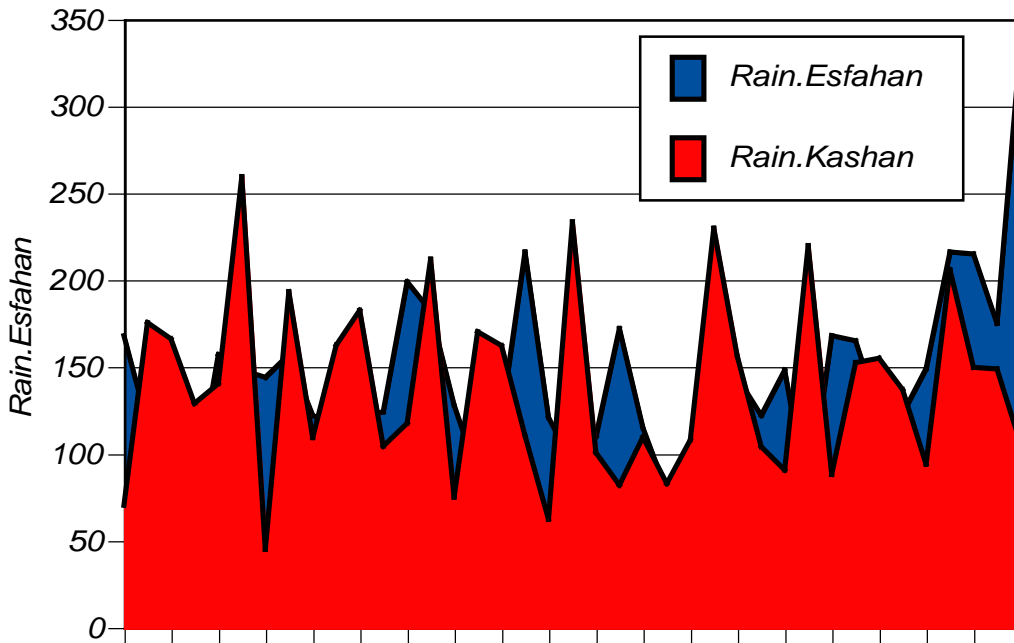
Variation Section of rain\_\_kashan

Parameter	Variance	Standard Deviation
Value	2733.705	<u>52.28484</u>

Plots Section of rain\_\_kashan



با توجه به نمودار های دو ایستگاه اصفهان و کاشان می توان نتیجه گرفت که هر چه انحراف معیار سری کمتر باشد پراکندگی سری کمتر و بهنجاری سری بیشتر خواهد بود. یعنی:



انحراف معیار به روش های مختلفی قابل محاسبه است آن روش ها عبارتند از :

- محاسبه انحراف معیار برای سری بدون فراوانی طبقه بندی شده.
- محاسبه انحراف معیار برای سری دارای فراوانی بدون طبقه بندی.
- محاسبه انحراف معیار برای سری دارای فراوانی طبقه بندی شده.

محاسبه انحراف معیار برای سری بدون فراوانی طبقه بندی شده.

برای محاسبه انحراف معیار سری بدون فراوانی طبقه بندی شده می توان از رابطه زیر استفاده کرد:

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$$

انحراف معیار نمونه

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

انحراف معیار جامعه

موقعیت شماره

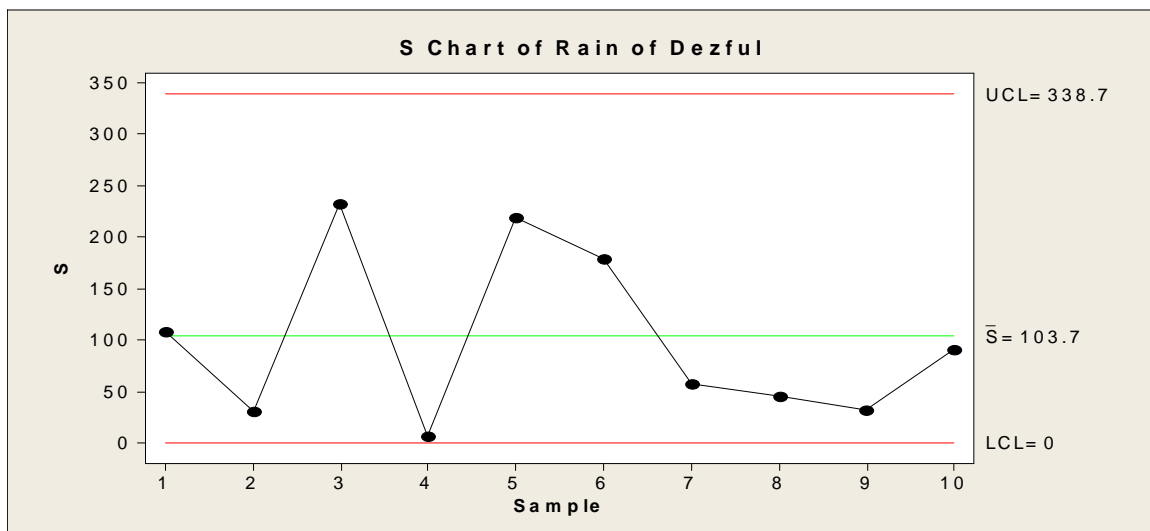
فرض کنید بارش 20 ساله (1986 - 2005) ایستگاه دزفول در دست باشد انحراف معیار آن را محاسبه کنید ؟

سری	انحراف از میانگین	$(X - \bar{X})^2$
510.30	86.665	7510.822
358.00	- 65.635	4307.953
405.30	- 18.335	336.172
361.80	- 61.835	3823.567
234.70	- 188.935	35696.434
562.20	138.565	19200.259
366.50	- 57.135	3264.408
375.60	- 48.035	2307.361
568.10	144.465	20870.136
258.70	- 164.935	27203.554
452.70	29.065	844.774
705.70	282.065	79560.664
448.60	24.965	623.251
530.10	106.465	11334.796
429.70	6.065	36.784
365.50	- 58.135	3379.678
390.00	- 33.635	1131.313
346.00	- 77.635	6027.193
465.60	41.965	1761.061

<b>337.60</b>	<b>- 86.035</b>	<b>7402.021</b>
<b>8472/7</b>		<b>236622/206</b>

$$S = \sqrt{\frac{236622.206}{19}} = 111.597 \quad \text{برای نمونه}$$

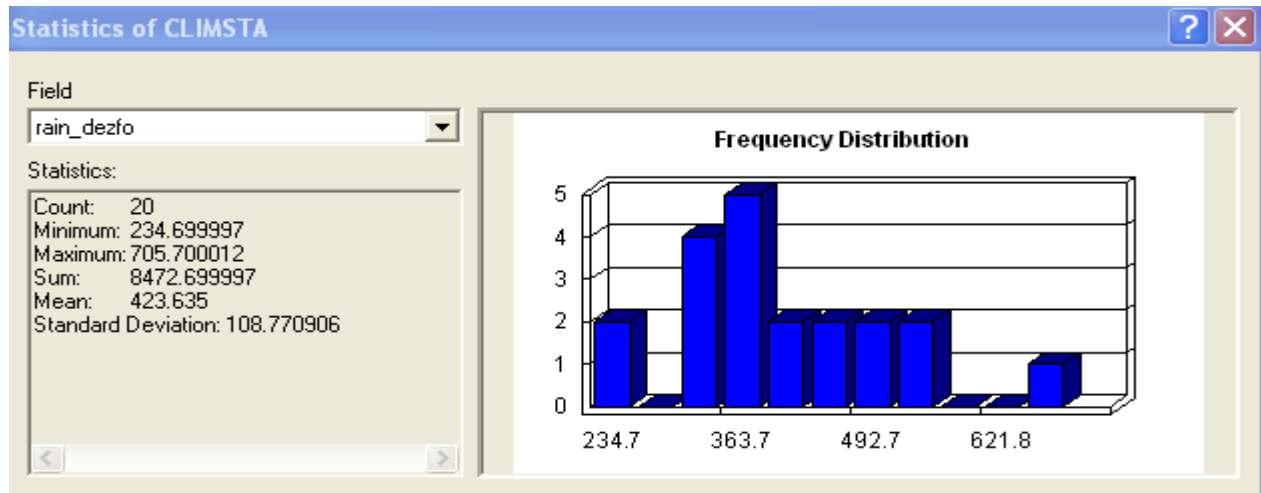
$$S = \sqrt{\frac{236622.206}{20}} = 108.771 \quad \text{برای جامعه}$$



با توجه به انحراف معیار سری **108/771** توزیع سری در قالب نمونه های میانگین انحراف معیار در طول دوره برابر با **103/7** میلی متر می

باشد و این توزیع انحراف معیار دامنه ای حدود **225** میلی متر را نشان می دهد .

محاسبه انحراف معیار با استفاده از نرم افزار **GIS** :



محاسبه انحراف معیار با استفاده از نرم افزار **Surfer** :

	X	Y	rain dezfoul
Mean:	4.065	3.695	423.635
Standard Deviation:	3.2233949494283	2.7993704649439	<u>108.77090730062</u>
Variance:	10.390275	7.836475	11831.110275

محاسبه انحراف معیار سری دارای فراوانی بدون طبقه بندی:

برای محاسبه انحراف معیار سری دارای فراوانی بدون طبقه بندی می توان از رابط زیر استفاده کرد:

$$S = \sqrt{\frac{\sum f_i (X - \bar{X})^2}{N - 1}}$$

انحراف معیار نمونه

$$S = \sqrt{\frac{\sum f_i (X - \bar{X})^2}{N}}$$

انحراف معیار جامعه

برای تبیین بیشتر به موقعیت شماره توجه نمایید.

موقعیت شماره :

اگر رطوبت نسبی 20 ساله (1986 - 2005) ایستگاه دزفول در دست باشد ، انحراف معیار آن را محاسبه کنید ؟

رطوبت نسبی $X_i$	فراوانی $F_i$	$X_i \times f_i$	$X_i - \bar{X}$	$f_i (X_i - \bar{X})^2$
37.00	1	5.0	-12/1	146/41
38.00	1	5.0	-11/1	123/21
47.00	3	15.0	-2/1	13/23
48.00	2	10.0	-1/1	2/42
49.00	3	15.0	-0/1	0/03
50.00	3	15.0	0/9	2/43
51.00	1	5.0	1/9	3/61



52.00	2	10.0	2/9	16/82
53.00	2	10.0	3/9	30/42
56.00	2	10.0	6/9	95/22
جمع	20	100.0		433/8

$$S = \sqrt{\frac{433.8}{19}} = 4.78 \quad \text{انحراف معیار برای نمونه :}$$

$$S = \sqrt{\frac{433.8}{20}} = 4.657 \quad \text{انحراف معیار برای جامعه :}$$

محاسبه انحراف معیار با استفاده از نرم افزار **Stata/SE10** :

```
. edit
(1 var, 20 obs pasted into editor)
- preserve
- preserve

. summarize var1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
var1	20	49.1	4.77824	37	56

محاسبه انحراف معیار سری دارای فراوانی طبقه بندی شده :

برای محاسبه انحراف معیار سری دارای فراوانی طبقه بندی شده از روابط زیر می توان استفاده نمود :

$$S = \sqrt{\frac{\sum f_i (C_p - \bar{X})^2}{N - 1}} \quad \text{انحراف معیار نمونه}$$

$$S = \sqrt{\frac{\sum f_i (C_p - \bar{X})^2}{N}} \quad \text{انحراف معیار جامعه}$$

برای فهم بهتر به موقعیت شماره توجه نمایید.

موقعیت شماره :

اگر سری بارش 20 ساله اسلام آباد غرب در دست باشد ، انحراف معیار آن را محاسبه نمایید؟

مراحل :

- تعیین دامنه تغییرات (  $R = (761 - 293.2) = 467.8$  )

- تعیین تعداد طبقات (  $k = \sqrt{N} = \sqrt{20} = 4.47 = 4$  )

- فاصله طبقات (  $C = 467.8 \div 4 = 116.95$  )

طبقات	فراوانی	Cp	F.Cp	C p - $\bar{X}$	$f_i (C p - \bar{X})^2$
293/2 410/15	6	351/675	2110/05	-134/492	108528/59
410/15- 527/1	6	468/625	2811/75	17/725	1885/05
- 644/05	7	585/575	4099/025	99/408	69173/65

527/1					
644/05-761	1	702/55	702/525	216/358	46810/78
جمع	20		9723/35		226398/07

$$S = \sqrt{\frac{226398.07}{19}} = 109.16 \quad \text{انحراف معیار نمونه}$$

$$S = \sqrt{\frac{226398.07}{20}} = 106.395 \quad \text{انحراف معیار جامعه}$$

محاسبه انحراف معیار سری بارش اسلام آباد غرب با استفاده از نرم افزار **Lisrel8.5**:

DATE: 03/12/2009

TIME: 03:49

P R E L I S 2.50

#### Univariate Summary Statistics for Continuous Variables

Variable	Mean	St. Dev.	T-Value	Skewness	Kurtosis	Minimum	Freq.
PREC	473.680	<u>111.350</u>	19.024	0.630	0.795	293.200	1
Maximum Freq.							
761.000	1						

محاسبه انحراف معیار بارش اسلام آباد غرب با نرم افزار **S-Plus**:

\*\*\* Summary Statistics for data in: DS7 \*\*\*

*rain*

**Min:** 293.2000  
**1st Qu.:** 390.2250  
**Mean:** 473.6800  
**Median:** 479.9000

3rd Qu.: 548.5500  
 Max: 761.0000  
 Total N: 20.0000  
 NA's : 0.0000  
 Variance: 12398.7501  
 Std Dev: 111.3497

محاسبه انحراف معیار بارش اسلام آباد غرب با نرم افزار EQS6.1:

**DESCRIPTIVE STATISTICS**

1 Variables are selected from file

Number of cases in data file are ..... 100

Variable ID	NAME	CASES	MEAN	SUM	SUM of SQUARE
1	V1	20	473.680	9473.600	235576.252
ID	Standard Deviation	Median	One Quartile	Three Quartile	Minimum
1	<u>111.350</u>	479.900	390.225	548.550	293.200
Maximum	761.000				

محاسبه انحراف معیار بارش اسلام آباد غرب با نرم افزار Eviews:

RESID	
473.6800	Mean
479.9000	Median
761.0000	Maximum
293.2000	Minimum
<u>111.3497</u>	Std. Dev.
20	Observations

با توجه به محاسبات می‌توان استنتاج نمود که انحراف معیار آماره ای است که مقدار آن از تمام متغیرها اثر می‌پذیرد. تغییر در مقدار هر متغیر موجب تغییر مقدار انحراف معیار می‌گردد. مقدار انحراف معیار ممکن است در اثر تعداد نسبتاً کمی از مقدار انتهایی شدیداً مخدوش شود. بنابراین انحراف معیار نیز مانند میانگین خود را در توزیع شدید چولگی از دست می‌دهد. از طرفی انحراف معیار را نمی‌توان در یک توزیع بی‌انتها، بدون داشتن اطلاعات اضافی محاسبه نمود. در استفاده از انحراف معیار به روش‌های مختلف مقدار بدست آمده از سری‌ها با هم متفاوت بودند، این تفاوت در اعداد بعنوان خطای طبقه‌بندی نامیده می‌شود. برای این که این خطا در انحراف معیار کاهش داده شود باید انحراف معیار محاسبه شده تصحیح شود. برای تصحیح انحراف معیار از فرمول تصحیح شپارد<sup>1</sup> استفاده می‌شود. فرمول شپارد عبارت است از:

$$S_c = \sqrt{s^2 - \frac{(C)^2}{12}}$$

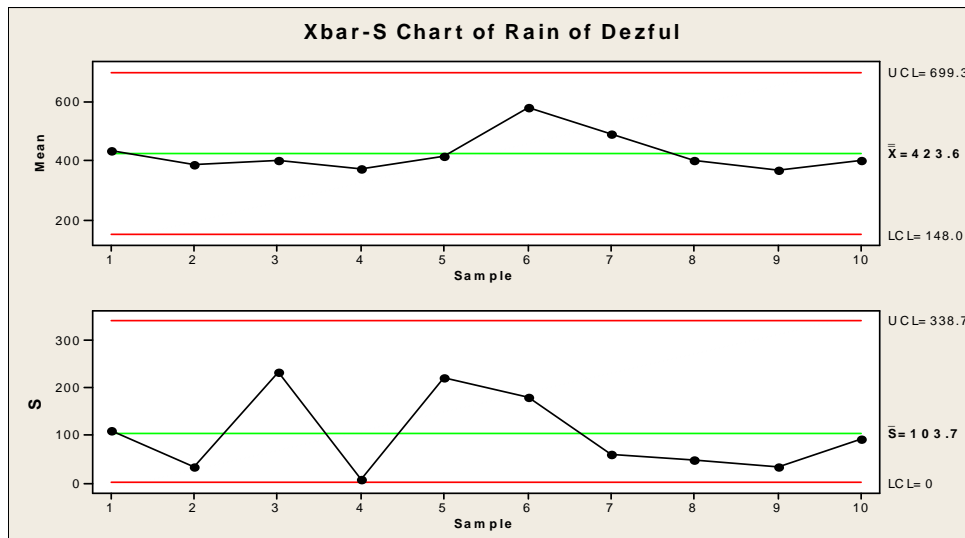
C فاصله طبقات،  $S_c$  انحراف معیار و  $SC$  انحراف معیار تصحیح شده است. به هر حال انحراف معیار هیچ وقت کمتر از صفر نمی‌باشد و با استفاده از انحراف معیار می‌توان مشخص نمود چند درصد متغیرها در فواصل مختلف نسبت به میانگین قرار دارند. مثلاً "توزیع بهنجار 68/26 درصد متغیرها به مقدار یک انحراف معیار بیشتر یا کمتر از میانگین قرار دارند. از نکات مهمی که باید در محاسبه انحراف معیار به آن توجه شود کنترل انحراف معیار می‌باشد. راه‌های متعددی برای کنترل انحراف معیار وجود دارد که نمونه ای از آن استفاده از رابطه بین دامنه تغییرات و انحراف معیار است. همانطوری که می‌دانید همیشه انحراف معیار از دامنه تغییرات کوچکتر است و اگر منحنی توزیع متقارن یا کمی چولگی داشته باشد و تعداد کل فراوانی‌ها زیاد باشد معمولاً انحراف معیار  $1/6$  ( $R / S = 6$ ) دامنه تغییرات است. هرچه تعداد فراوانی‌ها بیشتر می‌شود، این نسبت بیشتر می‌شود. برای مثال اگر تعداد فراوانی‌های جامعه ای 20 باشد تقریباً نسبت دامنه

---

Shppard,s Correction - <sup>1</sup>

تغییرات به انحراف معیار حدود 3/7 است و اگر فراوانی های جامعه ای 1000 باشد آن نسبت به 6/5 می رسد ، یعنی :  $\frac{R}{S} = 6.5$  .

مثلا" برای بارش دزفول می توان با استفاده از میانگین و انحراف معیار توزیع سری را نشان داد:



نمودار بالا نشان می دهد که تقریبا" آهنگ تغییرات میانگین و انحراف معیار سری باهم ارتباط دارند.

محاسبه نمره های معیار متغیرهای اقلیمی :

نمره های استاندارد یا معیار<sup>1</sup> نمراتی هستند که در یک توزیع موقعیت نسبی متغیرها را نسبت به میانگین نشان می دهند . به عبارت دیگر نمره های معیار فاصله متغیرها را تا میانگین بر حسب انحراف معیار مشخص می کنند و مشخص می سازند یک متغیر چند انحراف معیار بالاتر یا پایین تر از میانگین قرار دارد . نمره های معیار شامل نمره معیار  $Z$  ، نمره معیار  $T$  ، نمره هال و... هستند . برای محاسبه نمره معیار می توان از روابط زیر استفاده کرد:

$$Z = \frac{X_i - \bar{X}}{S_x} \quad \text{نمره معیار در نمونه}$$

$$Z = \frac{X - \mu_X}{\sigma_X} \quad \text{نمره معیار در جامعه}$$

بر این اساس نمره معیار، نسبت انحراف متغیرها از میانگین به انحراف معیار آنها می باشد. برای مثال اگر نمره معیار برابر با 2 باشد به این معنا است که مقدار نمونه 2 انحراف معیار بالاتر از میانگین قرار دارد. اگر نمره معیار صفر باشد به این معنی است که مقدار آن نمونه برابر با میانگین است. لذا هر چه مقدار نمره معیار بزرگتر باشد میزان انحراف از میانگین بیشتر است. با محاسبه نمره معیار می توان از جدول معین مساحت زیر منحنی را مشخص نمود. شکل توزیع متغیرهای معیار Z با شکل توزیع نمرات اصلی هیچ فرقی ندارد. دامنه نمرات معیار به ندرت از مثبت و منفی 3 خارج شود. میانگین نمره های معیار Z، صفر و انحراف معیار است. واحد اندازه گیری نمره های معیار Z، انحراف معیار است. متغیرهایی که بالاتر از میانگین قرار دارند، دارای Z مثبت هستند و برعکس. از طرفی سطح زیر منحنی بعنوان منحنی بهنجار و نرمال قلمداد می شود که محور عمودی آن متقارن است و داده ها حول میانگین متمرکز هستند و میانگین، میانه و نما باهم برابرند. توزیع نرمال معیار هم یکی از انواع توزیع نرمال است، به نحوی که میانگین آن صفر و انحراف معیار آن برابر یک برآورد می شود. مساحت زیر منحنی معادل یک یا صد درصد است. توزیع بهنجار متغیر تصادفی پیوسته X با میانگین  $\mu_X$  و انحراف معیار  $\sigma_X$  که دارای توزیع نرمال است تابع چگالی آن است. یعنی:

$$f_{(x)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$$

e پایه لگاریتم طبیعی (برابر با 2/7183)،  $\pi$  عدد ثابت 3/1416 است. همانطوری که مشخص است با افزایش میانگین منحنی توزیع به سمت راست گرایش می یابد، اما با افزایش انحراف معیار منحنی توزیع پهن تر می شود. توزیع نرمال با میانگین صفر و انحراف معیار یک را توزیع نرمال معیار می نامند. برای فهم بهتر به موقعیت زیر توجه شود.

موقعیت :

اگر سری 20 ساله (2005 – 1986) بارش ایستگاه روانسر در دست می باشد نمره معیار سری را همراه با سطح زیر منحنی سری را بر

حساب درصد محاسبه کنید؟

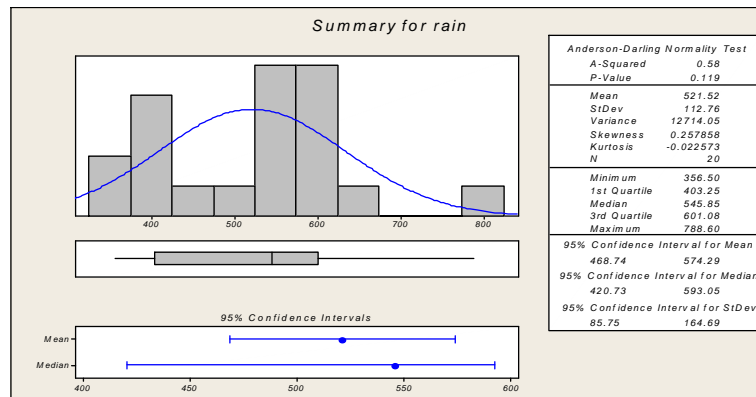
بارش	$z = \frac{x_i - \bar{x}}{s_x}$	مساحت بین	درصد مربوط	درصد مربوط	درصد احتمال
ایستگاه		میانگین و	به مقادیر کمتر	به مقادیر مساوی	وقوع سری

		مقدار Z	از مقدار Z	یا بیشتر از مقدار Z	بارش
440.00	- 0.72293	0/2642	0/7642	0/2358	40/9
583.80	0.55238	0/2088	0/7088	0/2912	38/59
393.60	- 1.13443	0/3708	0/8708	0/1292	45/22
563.20	0.36969	0/1406	0/6406	0/3594	36/32
386.60	- 1.19652	0/4750	0/975	0/025	49/20
556.80	0.31293	0/1217	0/6217	0/3783	35/57
609.40	0.77942	0/2794	0/7794	0/2206	41/29
611.90	0.80159	0/2881	0/7881	0/2119	41/68
788.60	2.36869	0/4909	0/9909	0/0091	49/6
414.80	- 0.94642	0/3264	0/8264	0/1736	43/25
629.80	0.96034	0/3315	0/8315	0/1685	43/64
602.80	0.72089	0/2642	0/7642	0/2358	40/90
534.90	0.11871	0/0438	0/5438	0/4562	32/64
356.50	- 1.46346	0/4279	0/9279	0/0721	47/21
399.40	- 1.08300	0/3599	0/8599	0/1401	44/43
360.00	- 1.43242	0/4236	0/9236	0/0764	47/21
569.50	0.42556	0/1628	0/6628	0/3372	37/07

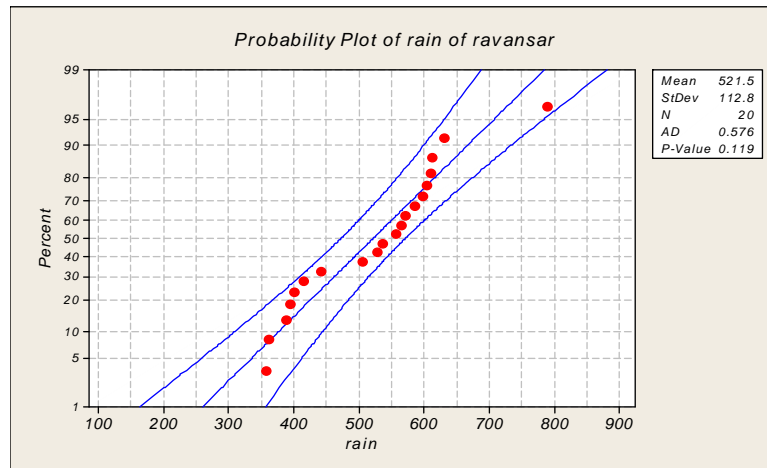


595.90	0.65970	0/2422	0/7422	0/2578	40/13
505.20	-0.14469	0/0557	0/5557	0/4443	33
527.60	0.05397	0/0199	0/5199	0/4801	31/56
میانگین 521/515					
انحراف معیار 112/757					

برای بدست آوردن مقدار مساحت زیر منحنی نرمال می توان از جدول مساحت زیر منحنی (جدول Z) استفاده کرد. با توجه به جدول بالا می توان نتیجه گرفت که 76/42 درصد بارش سالانه ایستگاه روانسر کمتر از 440 میلی متر است و حدود 23/58 درصد سری بارش روانسر بیشتر از 440 میلی متر می باشد و الی آخر.

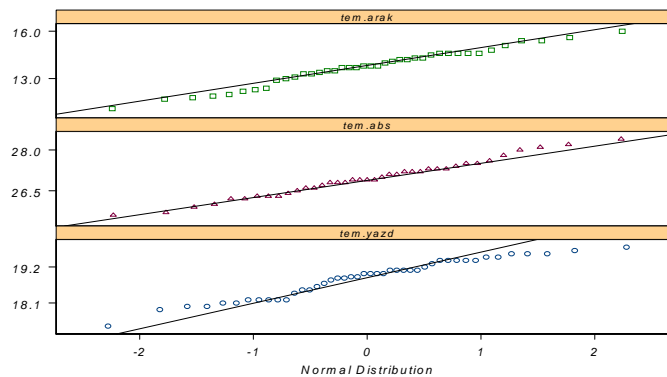


نمودار احتمال وقوع بارش روانسر :



موقعیت شماره :

اگر سری دمای 44 ساله (1952 - 1995) ایستگاه های یزد ، اراک و بندر عباس در دست باشد، توزیع نمره های معیار آنها چگونه می باشد؟



محاسبه نمره استاندارد  $T$  در مطالعات اقلیمی:

هنگامی که از نمره معیار  $Z$  استفاده می شود در برخی موارد نمره های بدست آمده منفی و اعشاری می شود . بنابراین برای از بین بردن علامت منفی نمرات آن با عدد ثابتی جمع می شود و برای از بین بردن اعشار نیز نمره  $Z$  در یک عدد ثابت ضرب می شود. اگر نمره معیار  $Z$  در عدد 10 ضرب شود و عدد حاصل با عدد 50 جمع شود ، نمره معیار  $T$  بدست می آید که دامنه آن بین 20 تا 80 است . یعنی :

$$T = 10Z + 50$$

پس نمره معیار  $T$  دارای میانگین 50 و انحراف معیار 10 است ولی نمره معیار  $Z$  دارای میانگین صفر و انحراف معیار یک است. نمره معیار مذکور منحنی توزیع را به 100 قسمت مساوی تقسیم می کند. برای فهم بهتر موضوع به موقعیت شماره اشاره می شود. موقعیت شماره :

اگر سری دمای 44 ساله (1952 - 1995) ایستگاه گرگان در دست باشد نمره معیار  $T$  سری مذکور را محاسبه کنید؟

ردیف	سری دما	$Z$	$T$	ردیف	سری دما	$Z$	$T$
1	17.6	-0.09571	49.04	23	17.1	-0.75992	42.40
2	17.67	-0.00272	49.97	24	18.3	0.83419	58.34
3	18	0.43566	54.36	25	17.3	-0.49424	45.06
4	18.5	1.09988	61.00	26	17.6	-0.09571	49.04
5	16.8	-1.15845	38.42	27	17.8	0.16998	51.70
6	17.1	-0.75992	42.40	28	18.6	1.23272	62.33
7	18.5	1.09988	61.00	29	18.1	0.56851	55.69
8	17.4	-0.36139	46.39	30	19.1	1.89693	68.97
9	18.5	1.09988	61.00	31	17.3	-0.49424	45.06
10	18.9	1.63125	66.31	32	18.6	1.23272	62.33
11	19	1.76409	67.64	33	16.7	-1.29129	37.09
12	18.3	0.83419	58.34	34	17.5	-0.22855	47.71
13	16.3	-1.82266	31.77	35	17.4	-0.36139	46.39
14	17.4	-0.36139	46.39	36	17.2	-0.62708	43.73
15	18.9	1.63125	66.31	37	17.6	-0.09571	49.04
16	17.6	-0.09571	49.04	38	17.8	0.16998	51.70
17	17.8	0.16998	51.70	39	17.5	-0.22855	47.71
18	16	-2.22119	27.79	40	17.4	-0.36139	46.39

19	18	0.43566	54.36	41	16.4	-1.68982	33.10
20	18.3	0.83419	58.34	42	16.6	-1.42413	35.76
21	17	-0.89276	41.07	43	16.8	-1.15845	38.42
22	17.3	-0.49424	45.06	44	18	0.43566	54.36

برای سال اول 49/04 نشان می دهد که متغیر (دما) مورد بررسی 0/0957 انحراف معیار پایین تر از میانگین واقع شده است ، والی آخره . در مطالعات آماری نمره های دیگری تحت عنوان نمره های نه گانه  $(Z_o = 2Z + 5)$ ، نمره های هال  $(Z_h = 14.28Z + 50)$ ، نمره های CEEB  $(Z_{CEEB} = 100Z + 500)^1$  و نمره های AGCT  $(Z_{AGCT} = 20Z + 100)^2$  وجود دارد که می توان از آنها استفاده کرد.

محاسبه چولگی یا کجی متغیرهای اقلیمی :

به توزیعی که میانگین آن به طرف چپ یا راست کشیده شده باشد ، توزیع دارای چولگی یا کجی <sup>3</sup> می گویند. ضریب چولگی هم آماره ای است که نحوه توزیع متغیرها اقلیمی (درجه نزدیکی توزیع به توزیع نرمال) و مقدار کجی را برای متغیر مورد بررسی نشان می دهد. مقدار ضریب چولگی مثبت و منفی <sup>3</sup> می باشد . برای محاسبه ضریب چولگی می توان از روابط زیر استفاده کرد:

$$Sk = \frac{\bar{X} - MO}{S}$$

یا

$$Sk = \frac{M_3}{S^3}$$

$M_3$  گشتاور مرتبه سوم ،  $MO$  نما است.

<sup>1</sup> - College entrance examination bord  
<sup>2</sup> - Army general classification test  
<sup>3</sup> - Skewness

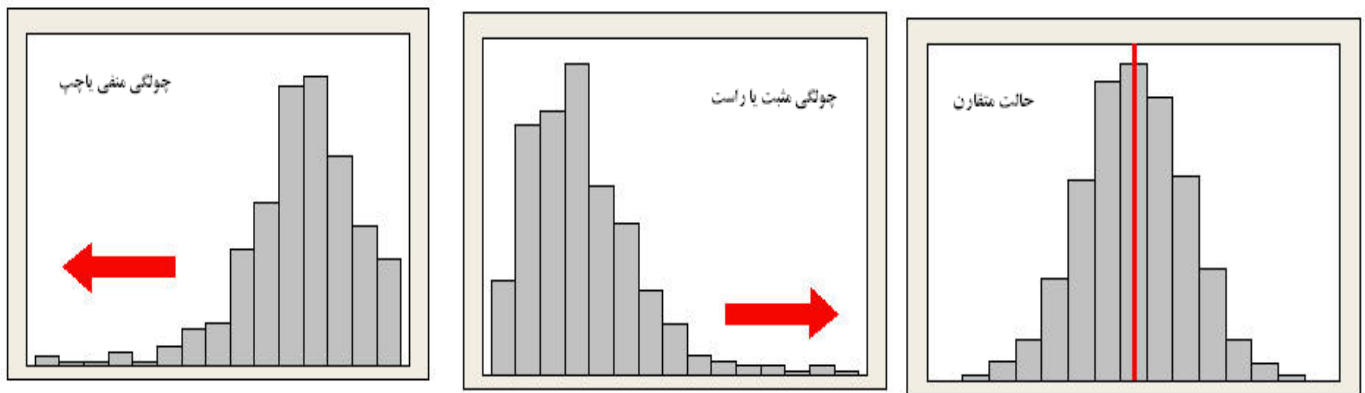
چولگی را می توان به انواع زیر مورد توجه قرار داد:

- چولگی مثبت

- چولگی منفی

- حالت متقارن

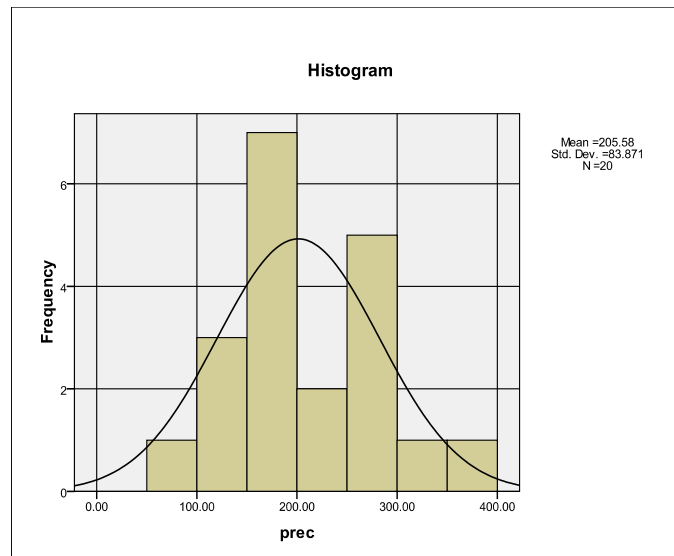
چولگی مثبت<sup>1</sup> یا چولگی به سمت راست زمانی است که میانگین بیشترین مقدار و میانه در بین نما و میانگین قرار داشته باشد ( $\bar{X} > M_d > MO$ ). در چولگی مثبت فاصله چارک های دوم و سوم بزرگتر از فاصله چارک اول و دوم است، یعنی  $(Q_2 - Q_1) < (Q_3 - Q_2)$  در چولگی مثبت اعداد اکثرا<sup>2</sup> در سمت چپ منحنی انباشته شده اند. به موقعیت شماره توجه شود. یعنی :



موقعیت شماره :

اگر سری بارش 20 ساله (1986 - 2005) ایستگاه بستان در دست باشد نمود چولگی سری را نشان دهید؟

<sup>1</sup> - Positive Skewness

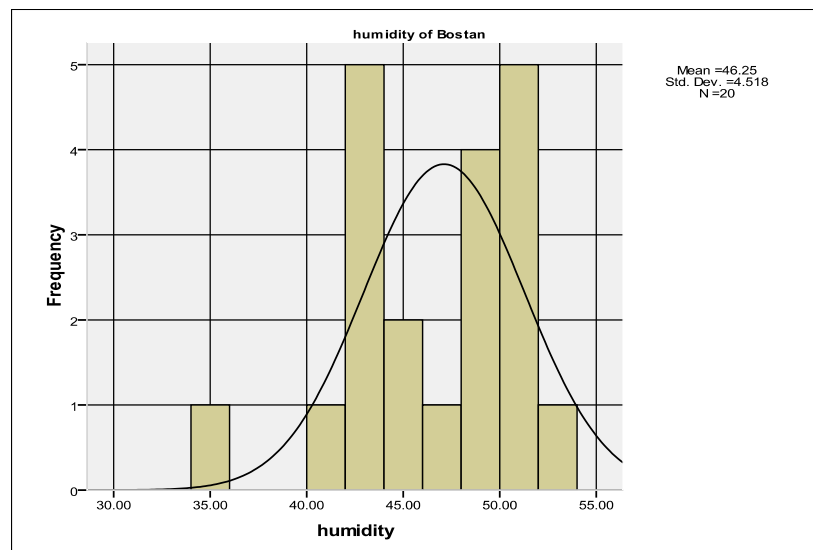


چولگی منفی :

چولگی منفی<sup>1</sup> زمانی است که نما دارای بالاترین مقدار ، میانگین دارای کمترین مقدار و میانه در بین نما و میانگین قرار داشته باشد، یعنی  $(\bar{X} < M_d < MO)$  . در چولگی منفی فاصله چارک های اول و دوم بزرگتر از فاصله چارک های دوم و سوم است ، یعنی  $(Q_2 - Q_1) > (Q_3 - Q_2)$  . در چولگی منفی متغیرها بیشتر در سمت راست منحنی انباشته شده اند. برای فهم بهتر به موقعیت شماره<sup>۱</sup> توجه شود.

موقعیت شماره<sup>۱</sup> :

رطوبت نسبی ایستگاه بستان برای نمود چولگی آن مشخص شود؟

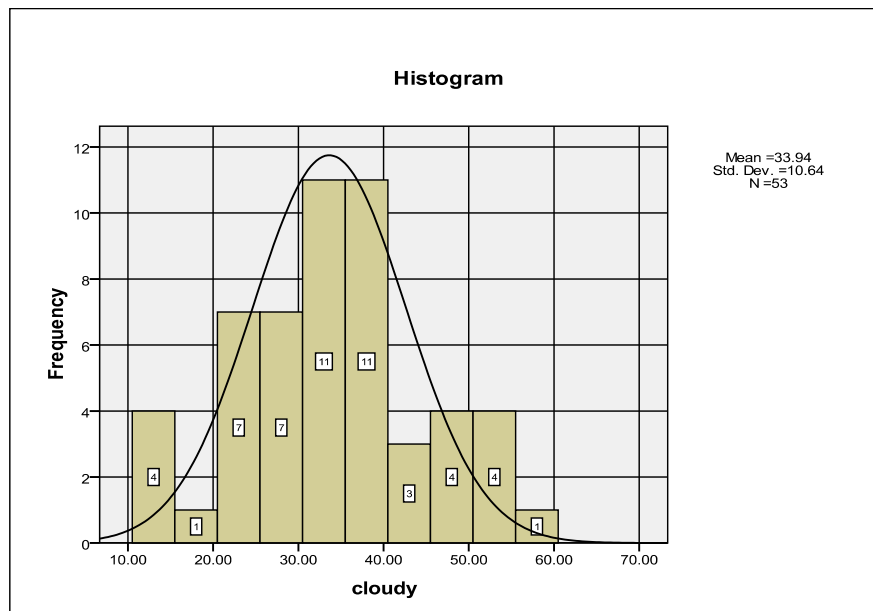


حالت متقارن :

هنگامی که نما ، میانگین و میانه با هم برابر باشند ، نمودار توزیع فراوانی حالت متقارن دارد . در این حالت فاصله چارک های اول و دوم برابر با فاصله چارک های دوم و سوم است ، یعنی  $(Q_2 - Q_1) = (Q_3 - Q_2)$  . در حالت متقارن مقدار چولگی برابر با صفر است . این شرایط در اقلیم بسیار نادر است . با این وجود بعضی از موقع ها چولگی بسیار ناچیز است . برای تبیین این مطلب به موقعیت زیر توجه نمایید .

موقعیت شماره :

اگر سری ابرناکی 53 ساله (1951 - 2003) اصفهان در دست باشد ، وضعیت بهنجاری سری چگونه است ؟



محاسبه ضریب چولگی متغیرهای اقلیمی :

برای محاسبه ضریب چولگی هم می توان از روابط زیر استفاده کرد:

$$Sk = \frac{\bar{X} - MO}{S}$$

یا

$$Sk = \frac{M_3}{S^3}$$

یا

$$Sk = \frac{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^3}{S^3}$$

اگر سری دارای توزیع فراوانی باشد :

$$Sk = \frac{\frac{1}{N} \sum_{k=1}^{k=k} f_i (X_i - \bar{X})^3}{S^3}$$

به هر حال اگر مقدار چولگی کوچکتر از صفر باشد، شکل توزیع فراوانی چولگی منفی یا چولگی به چپ دارد و اگر ضریب چولگی بزرگتر از صفر باشد، شکل توزیع مثبت یا چولگی به طرف راست دارد و اگر چولگی صفر باشد توزیع فراوانی متقارن است. طبق نتایج بررسی های پیرسون در هر توزیع فراوانی با چولگی خفیف و تک نمایی، میانه همیشه بین نما و میانگین یا به ترتیب با فاصله تقریبی دو به یک واقع می باشد، یعنی به نسبت دو از نما و یک از میانگین. به عبارت دیگر قدرمطلق میانگین و نما برابر است با سه برابر قدر مطلق تفاضل میانگین و میانه، یعنی  $|\bar{X} - M_d| = 3 |\bar{X} - MO|$ . برای فهم بهتر لازم است به موقعیت شماره اشاره شود.

موقعیت شماره :

اگر سری دمای سالانه 20 ساله (2005 - 1986) ایستگاه مسجد سلیمان در دست باشد، چولگی آن را حساب کنید؟

برای سری طبقه بندی شده:

مراحل :

- تعیین دامنه تغییرات (  $R = (27.1 - 23.7) = 3.4$  )

- تعیین تعداد طبقات (  $k = \sqrt{N} = \sqrt{20} = 4.47 = 4$  )

- فاصله طبقات (  $C = 3.4 \div 4 = .85$  )



طبقات	فراوانی	میانۀ طبقات	$C_p - \bar{X}$	$(C_p - \bar{X})^3$	$f_i (C_p - \bar{X})^3$
23/7 - 24/55	2	24/125	-1/445	-3/017	-6/034
24/55 - 25/4	7	24/975	-0/595	-0/21	-1/47
25/4 - 26/25	6	25/825	0/255	0/016	0/096
26/25 - 27/1	5	26/675	1/105	1/349	6/745
جمع	<b>20</b>				<b>-0/663</b>

$$Sk = \frac{\frac{-.663}{20}}{\frac{.831^3}{1}} = \frac{-.033}{.573} = -.0576$$

محاسبه چولگی سری دمای مسجد سلیمان با استفاده از نرم افزار **NCSS** :

#### Skewness and Kurtosis Section of C7

Parameter Skewness  
Value -0.3073463

محاسبه چولگی سری دمای مسجد سلیمان با استفاده از نرم افزار **SPSS** :

#### Descriptive Statistics

	N	Range	Mean	Std. Deviation	Variance	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error

<i>m.temp</i>	20	3.40	25.6350	.83115	.691	<u>-.333</u>	.512
Valid N (listwise)	20						

محاسبه ضریب چولگی با استفاده از ضرائب پیرسون:

برای محاسبه ضریب چولگی با استفاده از ضرائب پیرسون<sup>1</sup> می توان از روابط زیر استفاده کرد:

$$SK_1 = \frac{\bar{X} - M_d}{S}$$

$$SK_2 = \frac{3(\bar{X} - M_e)}{S}$$

برای سری دمای مسجد سلیمان:

#### Statistics

*m.temp*

N	Valid	20
	Missing	0
	Mean	25.6350
	Median	25.5500
	Mode	25.30 <sup>a</sup>
	Std. Deviation	.83115
	Skewness	-.333
	Std. Error of Skewness	.512

a. Multiple modes exist. The smallest value is shown

$$SK_1 = \frac{25.635 - 25.5}{.831} = .162$$

$$SK_2 = \frac{3(25.635 - 25.55)}{.831} = .306$$

چون مقدار چولگی ها مثبت هستند لذا توزیع فراوانی ها چولگی به طرف راست دارند. از نکات مهم در این خصوص وجود ضریب چولگی چارکی است. ضریب چولگی چارکی را به صورت زیر می توان تعریف نمود:

$$SKQ = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

هرگاه  $SKQ$  بزرگتر از صفر باشد توزیع فراوانی هم چولگی راست دارد و برعکس هرگاه  $SKQ$  کوچکتر از صفر باشد توزیع فراوانی چولگی چپ است. برای مثال چولگی چارکی باشد اصفهان چگونه می باشد؟

$$SKQ = \frac{(148.45 - 121) - (121 - 77.6)}{148.45 - 77.6} = -.225$$

بر این اساس چولگی توزیع فراوانی ها به چپ است. یعنی میانگین از میانه و میانه از نما کوچکتر است.

#### محاسبه کشیدگی متغیرهای اقلیمی :

کشیدگی<sup>1</sup> عبارت است از ارتفاع توزیع فراوانی در مقایسه با توزیع نرمال یا بهنجاری سری. کشیدگی نسبت میانگین توان چهارم انحراف از انحراف معیار به توان چهار است. اگر کشیدگی مساوی 3 باشد توزیع فراوانی تپه ای<sup>2</sup> و اگر کشیدگی کوچکتر از 3 باشد توزیع فراوانی فلاتی<sup>3</sup> و اگر کشیدگی بزرگتر از 3 باشد توزیع فراوانی قله ای<sup>4</sup> شکل خواهد بود. هنگامی که کشیدگی برابر با صفر باشد توزیع متغیرها توزیع نرمال یا بهنجار دارد در صورتی که کشیدگی مثبت باشد برآمدگی منحنی توزیع متغیرها در نقطه اوج قرار خواهد گرفت و در نهایت وقتی کشیدگی منفی باشد، منحنی توزیع متغیرها در نقطه اوج خوابیده خواهد بود. برای محاسبه کشیدگی از روابط زیر می توان استفاده نمود:

---

<sup>1</sup> - Kurtosis  
<sup>2</sup> - Mesokurtic  
<sup>3</sup> - Platykurtic  
<sup>4</sup> - Leptokurtic

$$k u = \frac{\sum (X_i - \bar{X})^4}{N} \div \left( \frac{\sqrt{\sum (X_i - \bar{X})^2}}{N} \right)^4$$

برای سری بدون فراوانی

$$k u = \frac{\sum f_i (X_i - \bar{X})^4}{N} \div \left( \frac{\sqrt{\sum f_i (X_i - \bar{X})^2}}{N} \right)^4$$

برای سری با فراوانی

اگر سری دمای 44 ساله (1952-1995) ایستگاه زاهدان در دست باشد مقدار کشیدگی آن را حساب کنید؟

محاسبه مقدار کشیدگی با استفاده از نرم افزار SPSS:

### Descriptive Statistics

	N	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
دمای زاهدان	44	0.6925806	0.480	0.033	0.357	<u>-0.052</u>	0.702

هر توزیع متغیری دارای چهار مشخصه است که بعنوان گشتاورهای<sup>1</sup> توزیع نامیده می شود که چولگی گشتاور سوم و کشیدگی گشتاور چهارم توزیع می باشد. یعنی:

$$m_1 = \frac{\sum (X_i - \bar{X})}{N} = 0$$

$$m_2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

$$m_3 = \frac{\sum (x_i - \bar{x})^3}{N}$$

$$m_4 = \frac{\sum (x_i - \bar{x})^4}{N}$$

گشتاور رتبه دوم واریانس نمونه ، گشتاور رتبه سوم چولگی و گشتاور رتبه چهارم کشیدگی منحنی است . به طور کلی گشتاور رتبه  $n$ م پیرامون مطابق رابطه زیر قابل محاسبه است:

$$m_n = \frac{\sum (x_i - \bar{x})^n}{N}$$

برای تبیین بهتر به موقعیت شماره توجه شود.

موقعیت شماره :

اگر سری دمای سالانه 20 ساله (2005 – 1986) ایستگاه مسجد سلیمان در دست باشد ، کشیدگی آن را حساب کنید؟

برای سری طبقه بندی شده:

مراحل :

- تعیین دامنه تغییرات (  $R = (27.1 - 23.7) = 3.4$  )

- تعیین تعداد طبقات (  $k = \sqrt{N} = \sqrt{20} = 4.47 = 4$  )

- فاصله طبقات (  $C = 3.4 \div 4 = .85$  )

سری میانگین دما	فراوانی	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$	$(X_i - \bar{X})^4$	$f_i (X_i - \bar{X})^4$
23.70	1	- 1/935	3/744	3/744	14/02	14/02
24.50	1	- 1/135	1/288	1/288	1/659	1/659
24.80	1	- 0/835	0/697	0/697	0/486	0/486
24.90	1	- 0/735	0/54	0/54	0/292	0/292
25.10	1	- 0/535	0/286	0/286	0/082	0/082
25.30	2	- 0/335	0/112	0/224	0/0126	0/0252
25.40	2	- 0/235	0/056	0/112	0/003	0/006
25.50	1	- 0/135	0/018	0/018	0/0003	0/0003
25.60	2	- 0/035	0/001	0/002	0/000001	0/000002
26.00	1	0/365	0/133	0/133	0/0178	0/0178
26.10	1	0/465	0/216	0/216	0/0467	0/0467
26.20	1	0/565	0/319	0/319	0/102	0/102
26.40	1	0/765	0/585	0/585	0/342	0/342
26.50	1	0/865	0/748	0/748	0/56	0/56
26.60	1	0/965	0/931	0/931	0/867	0/867
26.70	1	1/065	1/13	1/13	1/286	1/286
27.10	1	1/465	2/146	2/146	4/6	4/6

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 سالانه مسجد سلیمان نشان می دهد که توزیع متغیرهای دما در نقطه اوج خوابیده می باشند.

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

#### Descriptive Statistics

	Skewness		Kurtosis	
	Statistic	Std. Error	Statistic	Std. Error
میانگین دمای سالانه مسجد سلیمان	<b>-.333</b>	<b>0.512</b>	<b><u>0.154</u></b>	<b>0.992</b>

محاسبه 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<sup>1</sup> یا پراکندگی نسبی واریانس<sup>2</sup> می باشد. این شاخص برای مقایسه 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 مقایسه پراکندگی سریها با واحدهای مختلف استفاده نمود. ضریب

<sup>1</sup> - Coefficient of Variation

<sup>2</sup> - Relative Variance

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

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

موقعیت شماره :

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

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

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

با توجه به ضرائب تغییرات، سری دمای یزد میزان پراکندگی آن نسبت به دمای اصفهان بیشتر است. محاسبه 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
Value	-0.429703378021337	-0.633722857249912	2.30270244951713	-
of Variation	0.445020526745701	2.48564593301435E-02	<u>3.09190345628066E-02</u>	
Std Error	0.261801115787813		0.455219910457506	
	2.71691161399019E-03			

*Skewness and Kurtosis Section of temperature\_esfahan*

Coefficient



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

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

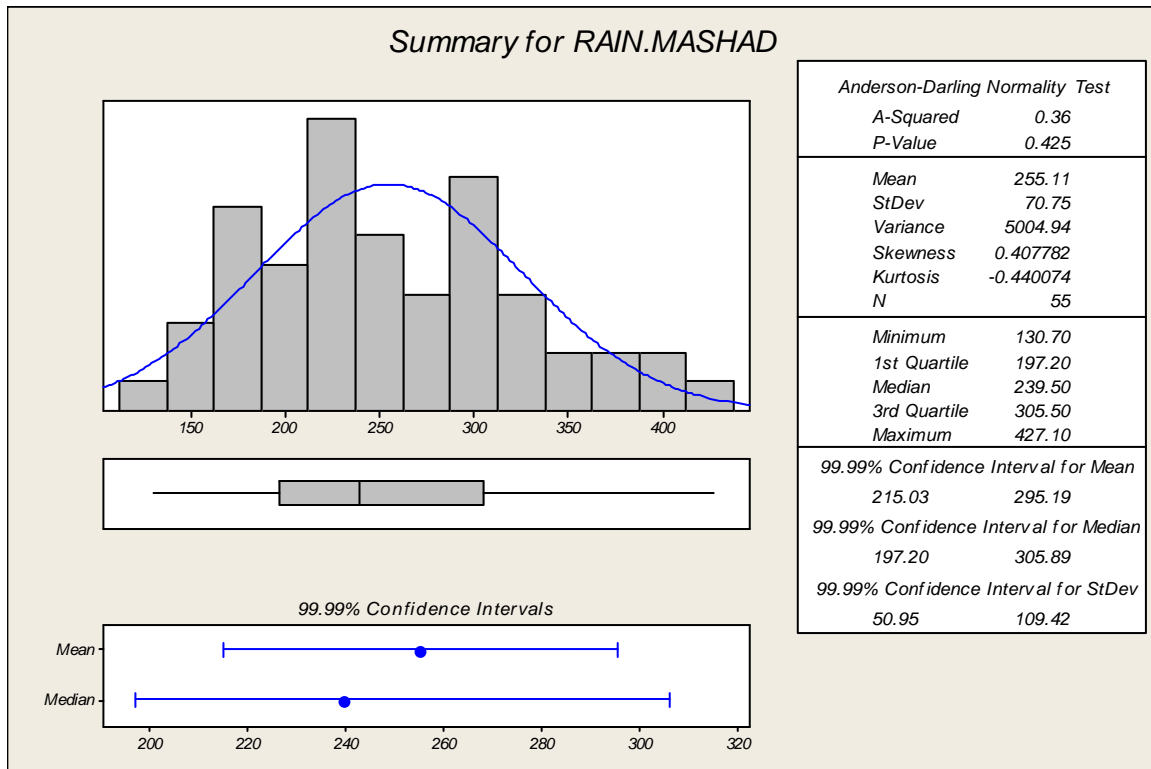
*Descriptive Statistics: yazd; esfahan*

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

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

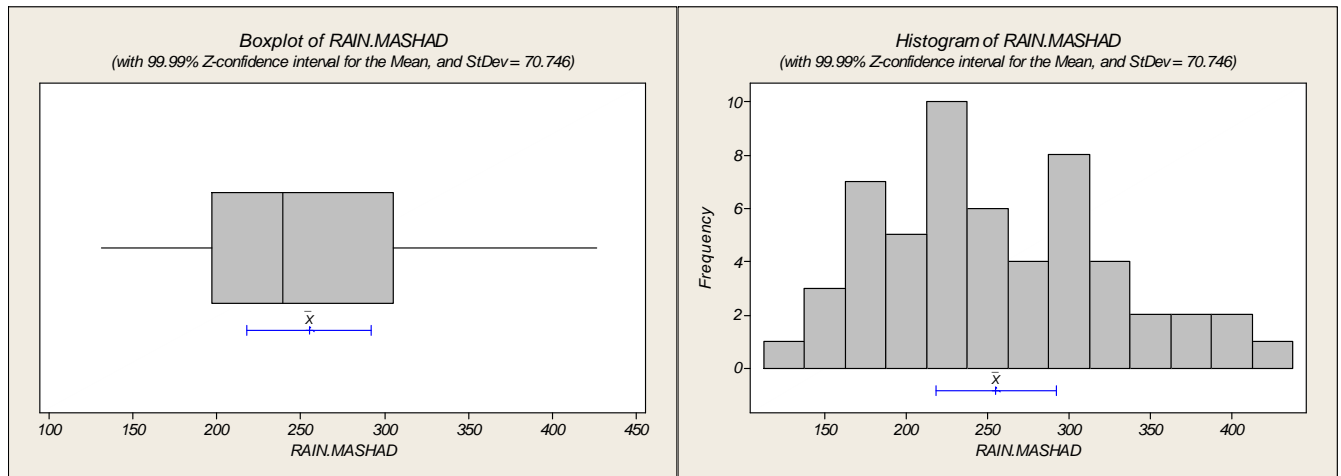
برای کاربرد و ارائه گزاره های آماره های توصیفی ، از سری بارش ایستگاه های مختلف ایران استفاده شده است . بر این اساس بارش 56

ساله (2005 1954) ایستگاه مشهد در قالب گزاره 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 ساندرسون ( $0/425$ ) می توان استدلال نمود که فرض تحقیق ما در خصوص بهنجاری سری مشهد پذیرفته می شود. (مقدار ضریب از مقدار  $0/05$  بیشتر است).

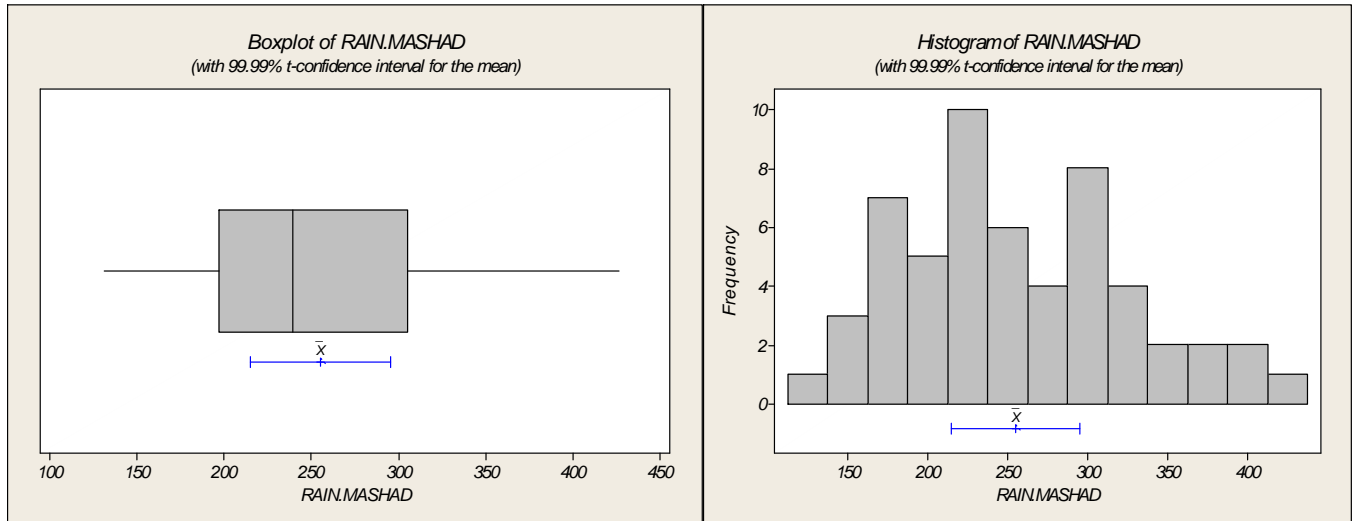
ضریب تعیین مقدار  $0/36$  را نشان می دهد. میانگین بارش  $255/11$  میلی متر وانحراف معیار  $70/75$  می باشد. میزان چولگی  $0/408$  نشانگر کشیدگی توزیع می باشد و مقدار عدم تقارن را نشان می دهد و مقدار مثبت آن میل به چولگی به راست را نشان می دهد. همچنین مقدار  $0/44$  کشیدگی پهن شدگی توزیع را نشان می دهد و مقادیر چارک اول ( $197/2$ ) و چارک سوم ( $305/5$ ) و میانه آن ( $239/5$ ) می باشد. از نکات دیگر تحلیل آماره ها مقادیر فاصله اطمینان سری بارش مشهد است. بطوری که فاصله 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 فاصله اطمینان برای میانه با احتمال  $99/99$  بین  $192/2$  تا  $305/19$  میلی متر می باشد و فاصله 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 فاصله اطمینان نمره معیار سری بارش مشهد با احتمال **99/99** درصد با در نظر گرفتن میزان انحراف معیار سری ، بارش سری بین **217/992** تا **292/219** میلی متر می باشد. همچنین میزان میانگین سری براساس نمره معیار **255/105** میلی متر می باشد و انحراف از

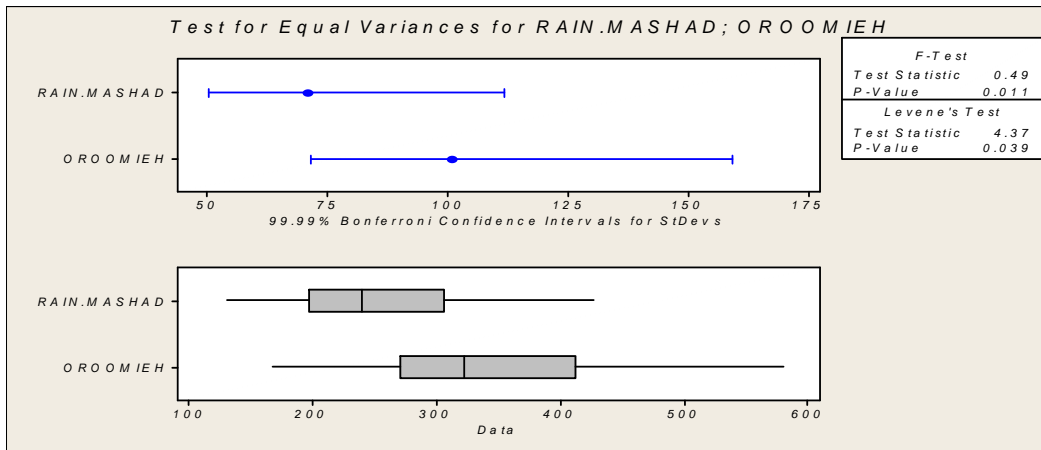
میانگین **9/539** میلی متر می باشد. یعنی  $SE.Mean = \frac{S}{\sqrt{N}} = \frac{70.746}{\sqrt{55}} = 9.539$  . همچنین در این رابطه می توان به تحلیل آماره دیگری

پرداخت .

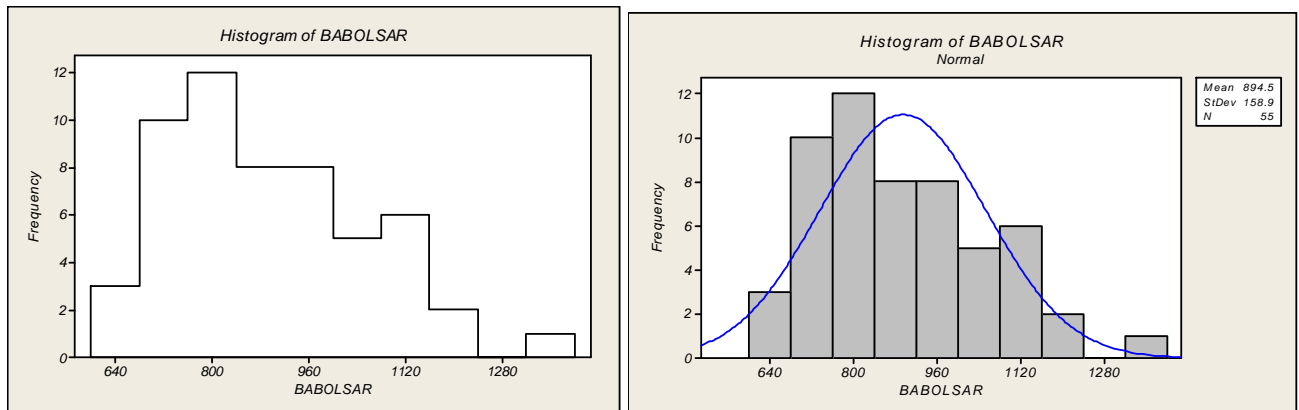


با توجه به فاصله اطمینان آزمون **t** سری بارش مشهد با احتمال **99/99** درصد با در نظر گرفتن میزان میانگین سری ، بارش سری بین **295/186** تا **215/025** میلی متر می باشد. همچنین میزان میانگین سری **255/105** میلی متر می باشد و انحراف از میانگین **9/539** میلی متر می باشد. از نکاتی که باید در تحلیل آماره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

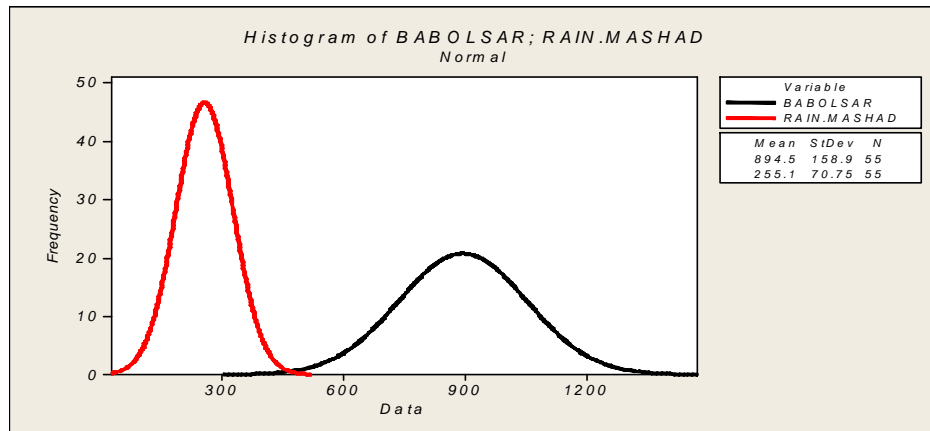
باشد. در این مورد از آزمون‌های **Levene** و **Bartlett** می‌توان استفاده کرد. آزمون اول نسبت به آزمون دوم حساسیت کمتری نسبت به سنجش بهنجاری سری‌ها دارد. بر این اساس برای سنجش سری‌ها از بارش ایستگاه‌های مشهد و ارومیه استفاده شد.



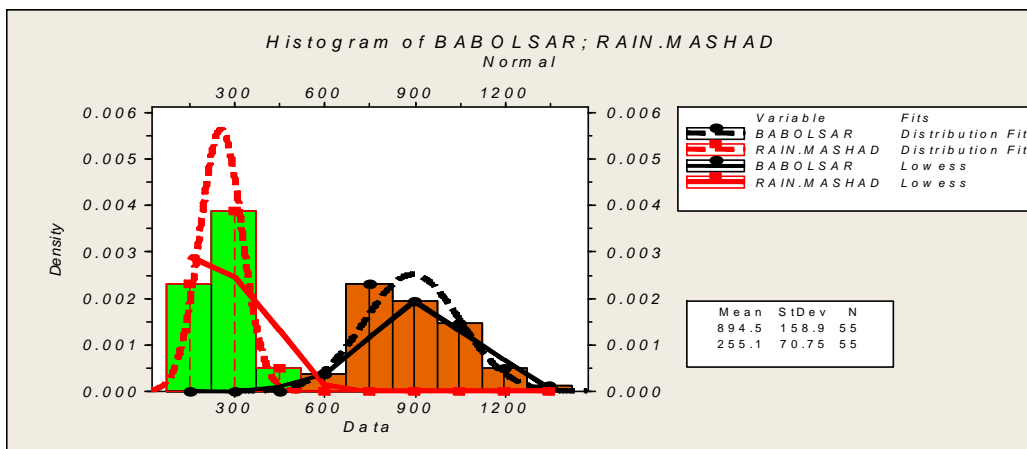
با توجه به آزمون **F** نشان می‌دهد سری بارش ایستگاه‌ها نرمال یا بهنجار می‌باشند. آزمون لیون غیر نرمال بودن سری را مورد سنجش قرار می‌دهد. به کاربرد آزمون را در قسمت‌های بعدی کتاب اشاره می‌شود. یکی دیگر از رویکردهای تحلیل آماره‌های توصیفی استفاده از نمودار هستوگرام است که شکل، مکان، تمایل مرکزی و گستردگی توزیع را نشان می‌دهد. برای کاربرد آن از سری 55 ساله (2005-1951) بابلسر استفاده شده است.



نمودارها فراوانی سری بارش و وضعیت برازش سری بارش بابلسر را نشان می‌دهند. نمودار دیگری که در برازش سری اهمیت دارد، نمونه‌ای از سری هستوگرام است. یعنی:



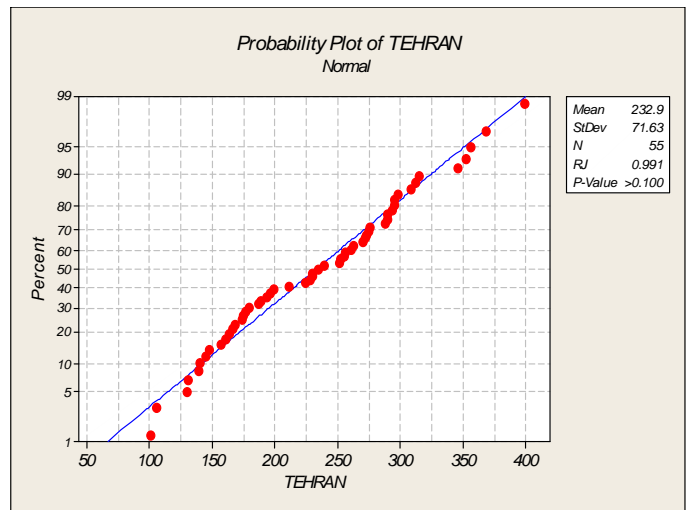
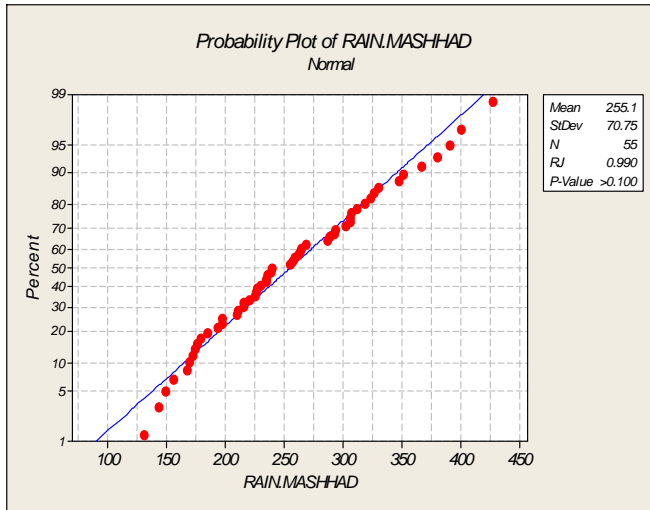
نمودار بالا میزان کشیدگی سری های بارش مشهد و بابلسر را نشان می دهد. بطوری که میزان کشیدگی بارش مشهد بیشتر از سری بابلسر است. هر چند که هر دو سری کشیدگی منفی را نشان می دهند. یعنی سری های بارش مشهد و بابلسر از اوج وضعیت نرمال فاصله دارند. یعنی:



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

بررسی بهنجاری سری های اقلیمی:

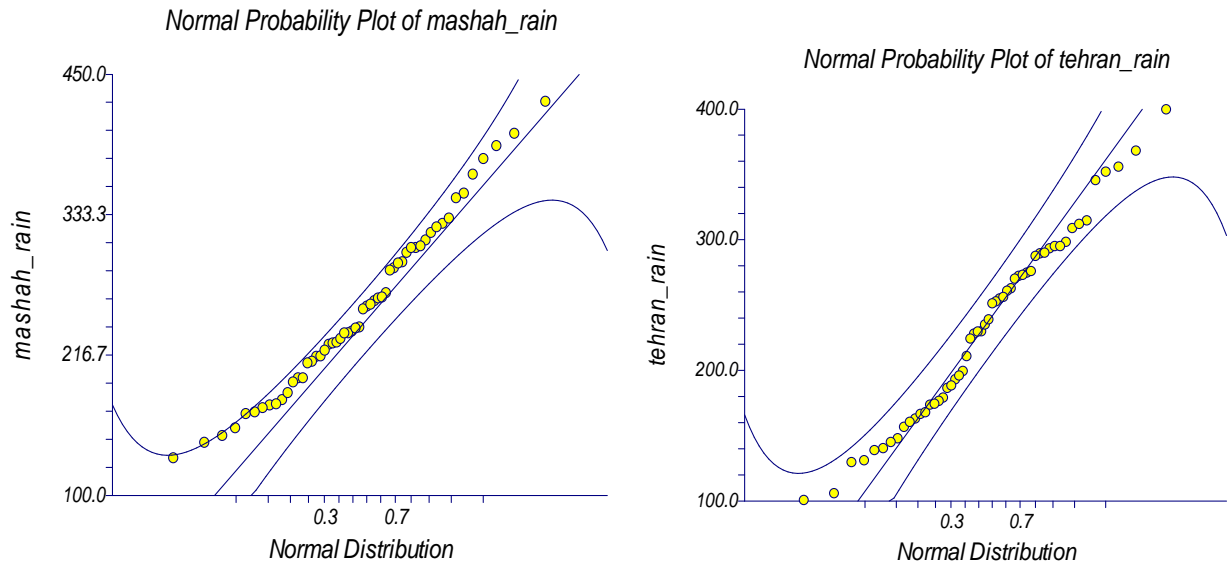
برای بررسی بهنجاری سری های اقلیمی از آزمون های مختلفی می توان استفاده کرد. نمونه ای از روش ها، استفاده از نمودار بهنجاری است. قابلیت های مختلف نرم افزاری فراوانی وجود دارد که می تواند به این فرآیند توجه نمود. بر این اساس برای بررسی از نرم افزار های minitab و NCSS استفاده می شود. برای انجام این فرآیند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



با توجه به سری های 55 ساله (1951-2005) بارش مشهد و تهران از آزمون آریانا- جونیر<sup>1</sup> استفاده شده است. این 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 سالانه بارش مشهد و تهران به شرایط بهنجاری نزدیک هستند یا به عبارتی سری ها بهنجار می باشند. آزمون ریانا- جونیر را می توان از رابطه زیر استفاده کرد:

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

$x_i$  مشاهده 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:



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.

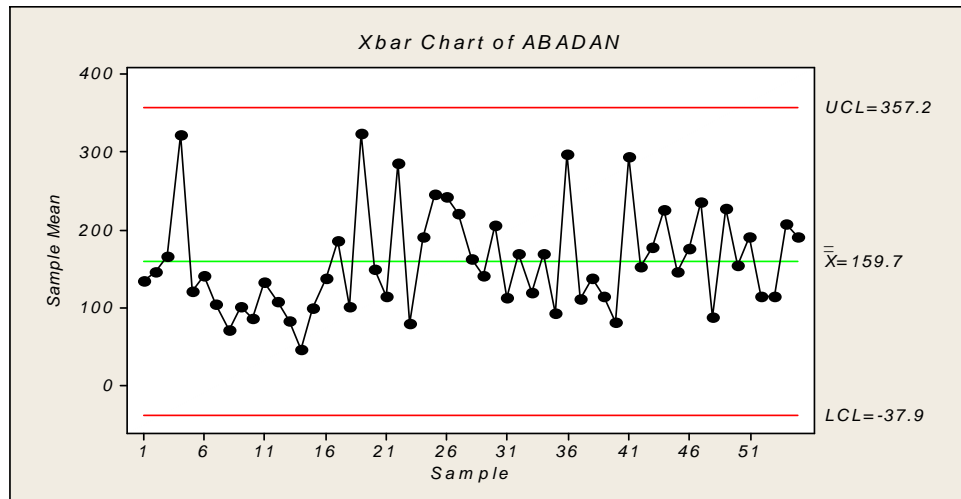
#### کنترل آماره های توصیفی در تحلیل متغیرهای اقلیمی :

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<sup>1</sup> (CL)، حدود کنترل<sup>2</sup> (UCL و LCL) و نقاط ترسیمی هستند. برای تبیین کنترل آماره های توصیفی به موقعیت های زیر اشاره شده است.

موقعیت شماره :

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

<sup>1</sup> Center Line -  
<sup>2</sup> Upper Control Limit & Lower Control Limit -



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

$$P-S = \frac{S_p}{C_4(d)}$$

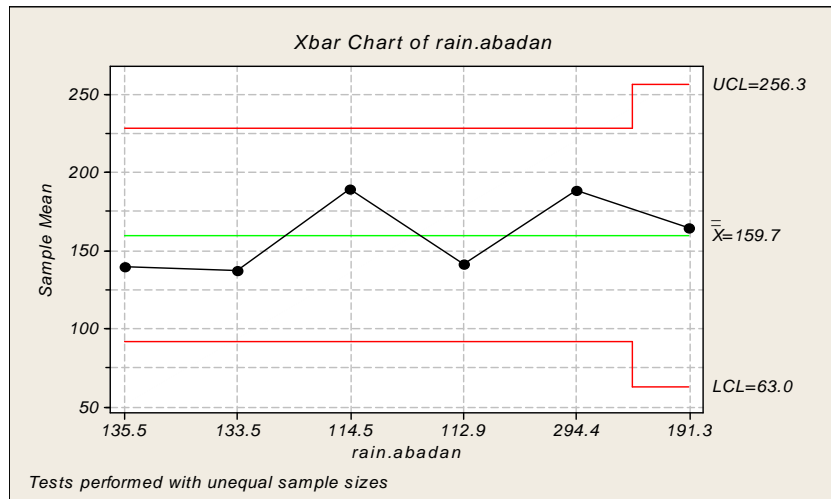
$$S_p = \sqrt{\frac{\sum_i \sum_j (X_{ij} - \bar{X}_i)^2}{\sum_i (n_i - 1)}}$$

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

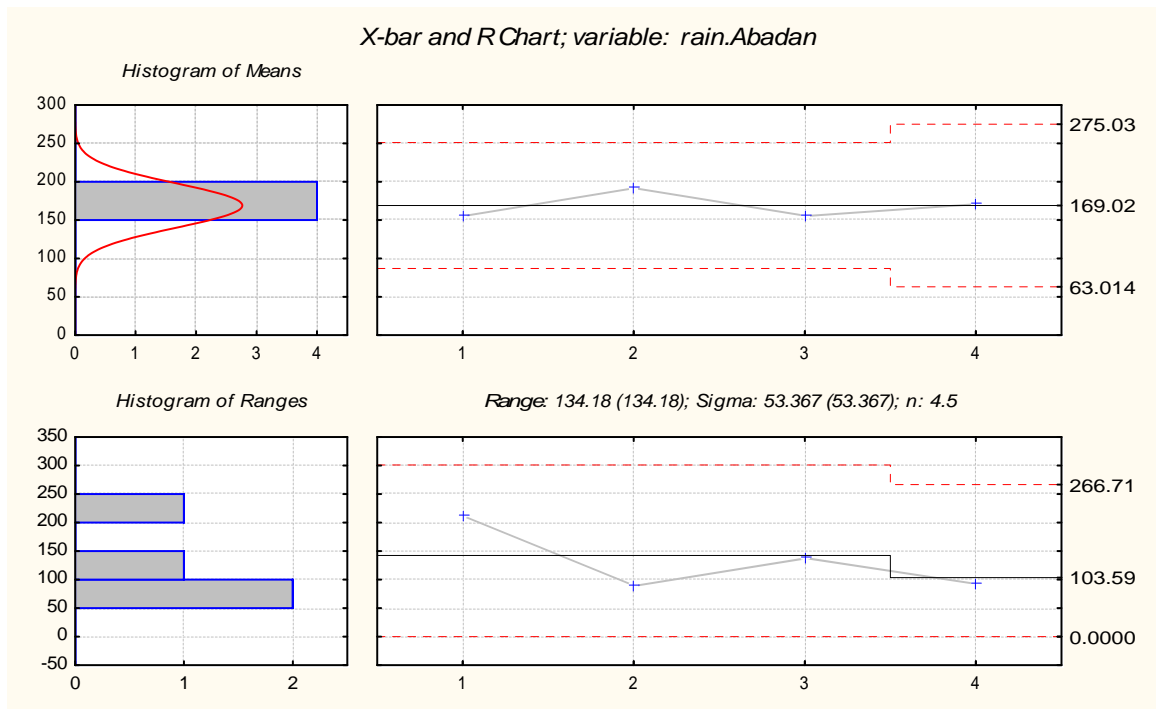
$$C_4(d + 1) = \sqrt{\frac{2}{d}} \times \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}$$

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





کنترل بارش آبادان با استفاده از نرم افزار **Statistica7** :



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 شماره نشان داده شده است. برای محاسبه میانگین انحراف معیار می توان از رابطه زیر استفاده کرد:

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

اگر سری مورد نظر توزیع بهنجار باشد رابطه میانگین و انحراف معیار سری را می توان به صورت رابطه 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<sub>4</sub>** مقدار ثابت است که فقط به اندازه نمونه **n** بستگی دارد و مقادیر آن به ازای **n** های متفاوت از جدول معین قابل استخراج است . خط

مرکز نمودار **S** برابر با  $\bar{S}$  است . برای محاسبه حدود اطمینان کنترل نمودار ترسیمی لازم است تخمین دقیقی از انحراف معیار داشت. ثابت

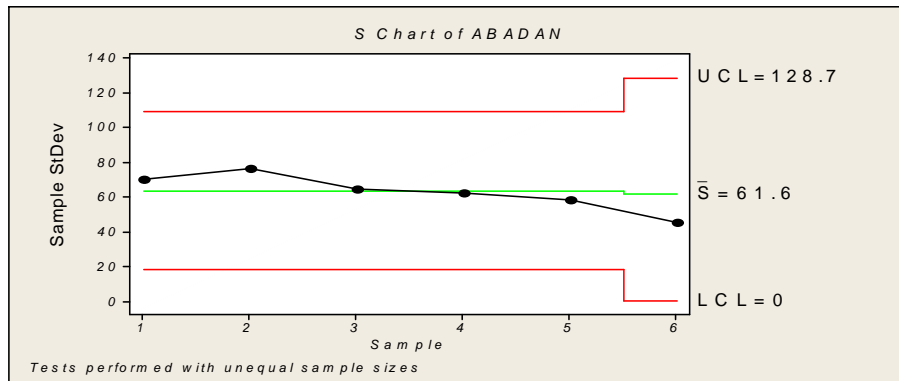
می شود انحراف معیار سری برابر با  $\bar{X} \sqrt{1 - C_4^2}$  است . لذا حدود اطمینان کنترل نمودار انحراف معیار برابر خواهد بود:

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

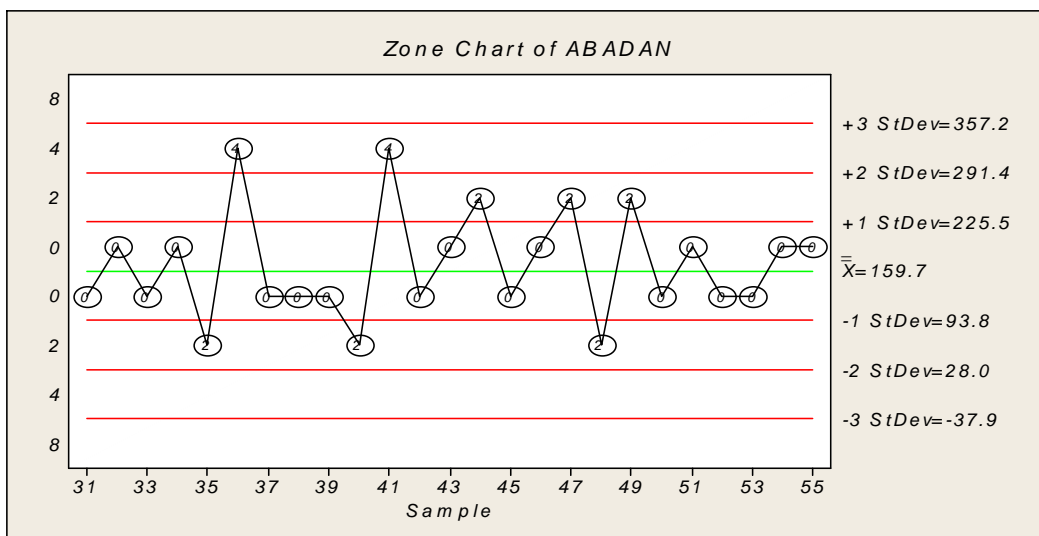
$$UCL = \bar{S} + 3 \bar{X} \sqrt{1 - C_4^2}$$

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 دامنه متحرک (I-MR) :

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

$$MR_i = |X_i - X_{i-1}|$$

اگر توزیع شرایط بهنجاری داشته باشد رابطه بین میانگین و دامنه متحرک برابر است با:

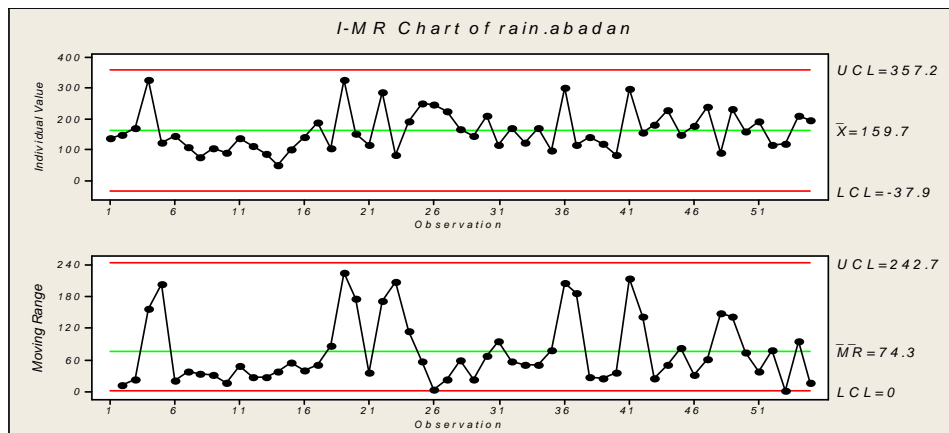
$$\hat{\sigma} = \frac{\overline{MR}}{D_2}$$

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

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

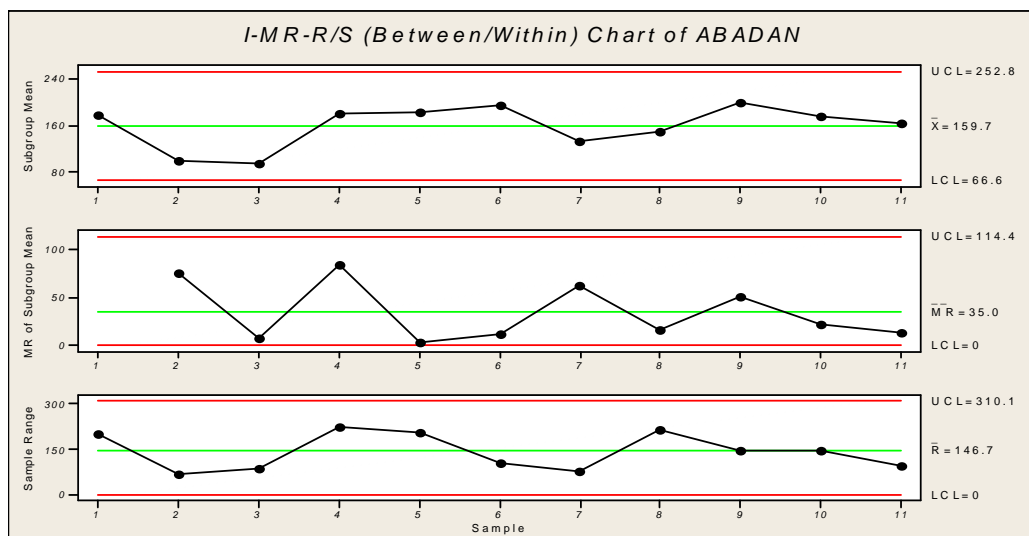
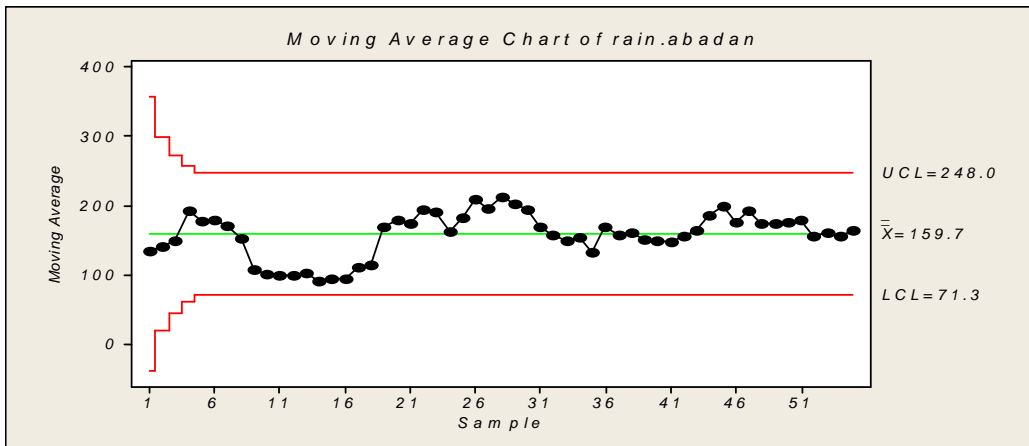
$$LCL = \bar{X} - 3 \frac{\overline{MR}}{D_2}$$

سری بارش آبادان با نمودار مذکور مورد بررسی قرار گرفت :



کنترل سری بارش آبادان به صورت دامنه 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 بین  $71/3$  تا  $248$  در نوسان را نشان می دهد. از طرفی هرچه رتبه میانگین متحرک افزایش یابد میل سری به میانگین نزدیک تر می شود.



سری بارش آبادان در برابر میانگین، میانگین متحرک و دامنه 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 دامنه متحرک، میانه دامنه 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:

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

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

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

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

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

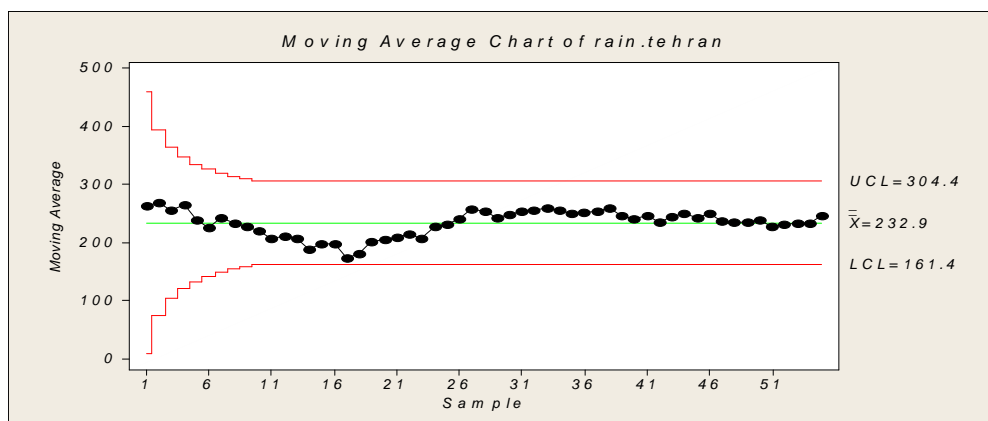
$$UCL = \bar{X} + K(S/n)\sqrt{1/n_1 + \dots + 1/n_1}$$

حدود اطمینان سری اگر  $n$  بزرگتر از  $w$  باشد:

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

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

بر این اساس سری بارش 55 ساله (2005-1951) 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<sup>1</sup> (EWMA) یکی از روش های مهم به منظور شناسایی تغییرات جزئی در هر مجموعه ای از متغیرهای

اقلیمی است. آماره این نمودار را می توان به صورت زیر نشان داد:

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

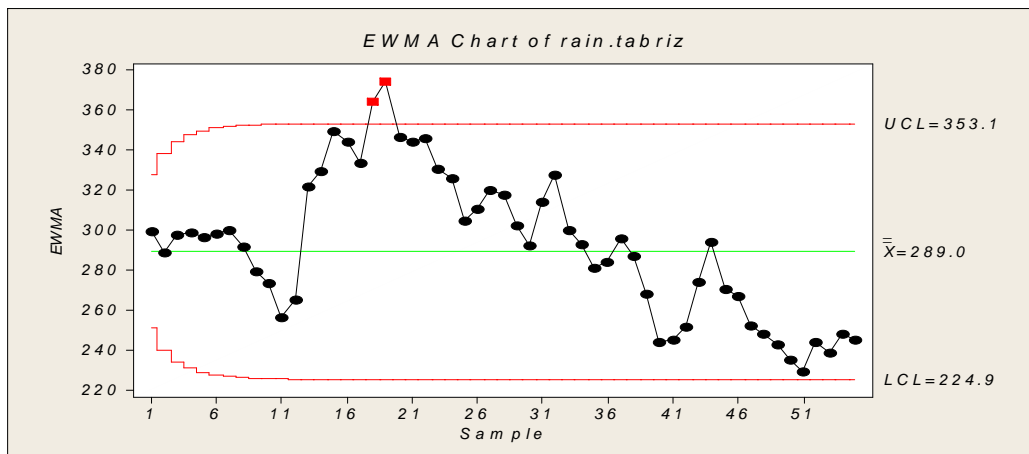
$\bar{X}_t$  میانگین متحرک وزنی در زمان  $t$  و  $\lambda$  ضریب وزنی برای متغیرها بطوری که  $0 < \lambda < 1$  است. حدود اطمینان سری را می توان برای

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 ساله (1954-2005) ایستگاه تبریز مورد بررسی قرار گرفت :



کنترل جمع تجمعی متغیرهای اقلیمی :

<sup>1</sup> Exponentially Weighed Moving Average

یکی از نمودار های مهمی که می توان متغیرهای در حال کنترل را بررسی نمود نمودار جمع تجمعی<sup>1</sup> (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 رابطه زیر بدست می آید:

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

شیب ماسک از فرمول زیر بدست می آید:

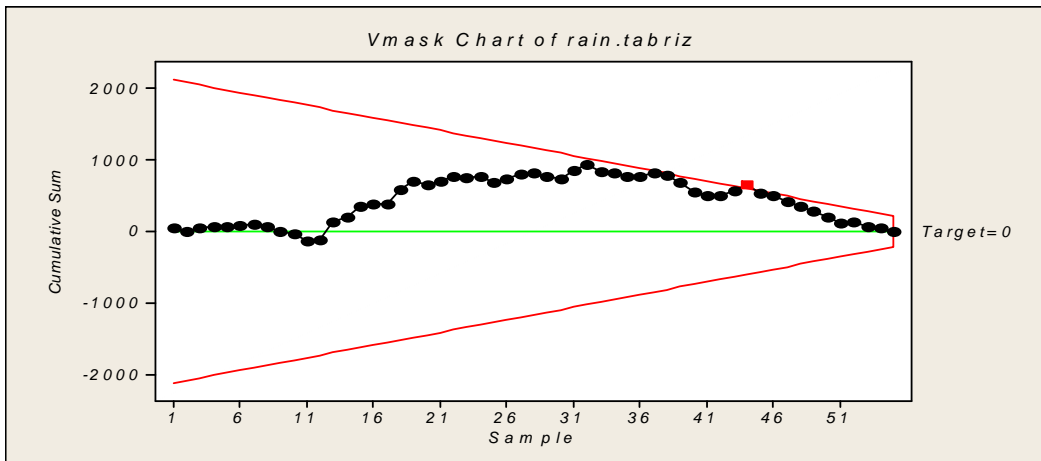
$$k \frac{S}{\sqrt{N}}$$

عرض ماسک از فرمول زیر بدست می آید:

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

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





کنترل متغیرهای اقلیمی با استفاده از کنترل چند متغیره:

اکثراً در تحلیل های اقلیمی مناطق مختلف معمولاً از چند متغیر استفاده می شود. یکی از روش های کنترل، استفاده از شیوه های کنترل چند

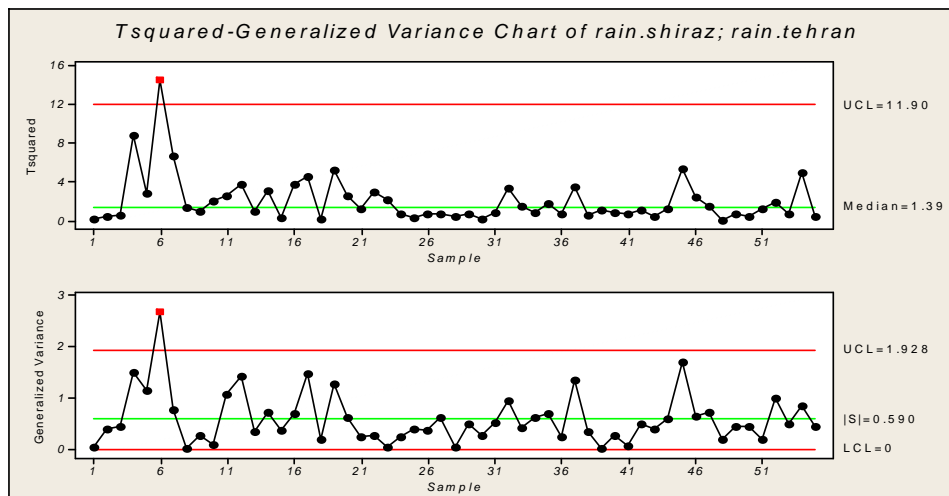
متغیره می باشد. در استفاده از کنترل چند متغیره آماری از آماره  $T^2$  می توان استفاده نمود. در این آماره از میانگین نمونه ها یا سری ها،

واریانس نمونه ها، کواریانس نمونه ها استفاده می شود. مقدار آماره  $T^2$  را می توان از رابطه زیر به دست آورد:

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

برای تبیین بیشتر از سری بارش 55 ساله (1951-2005) ایستگاه های شیراز و تهران استفاده شده است. نمودار زیر کنترل چند متغیره

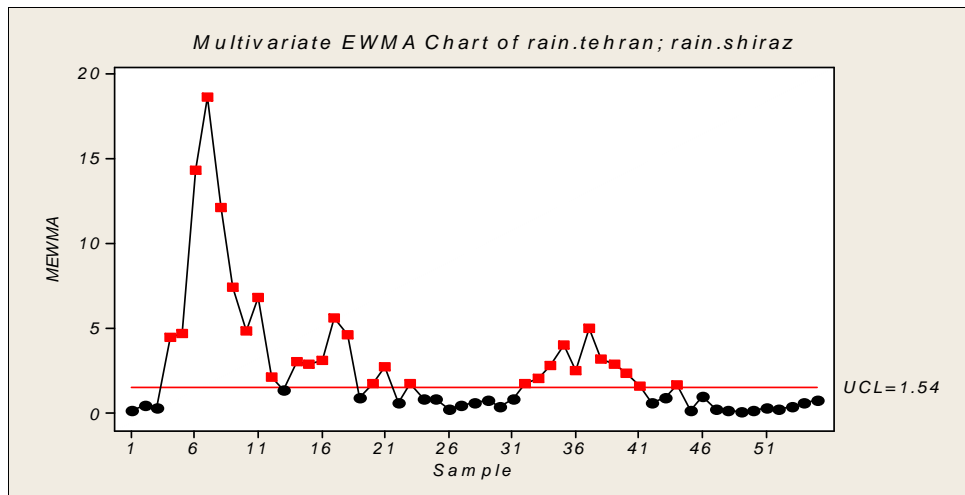
بارش شیراز و تهران را نشان می دهد:



نمودار دیگری که در کنترل چند متغیرها بکار می رود نمودار میانگین متحرک وزنی نمایی<sup>1</sup> (EWMA) می باشد. در این آماره کنترل متوسط طول دوره<sup>2</sup> (ARL) اهمیت زیادی دارد که در واقع متوسط تعداد دوره هایی است که بعد از آن تغییرات در فرآیند را به دست آورده می شود. مقدار آماره آن را می توان از رابطه 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:

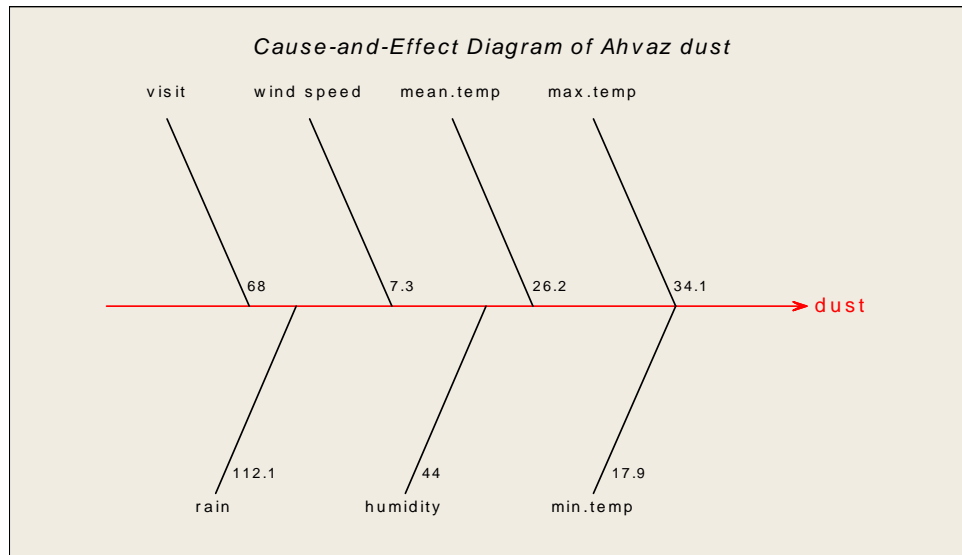


تحلیل علی متغیرهای اقلیمی با استفاده از نمودار علت و معلول:

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 در یک ساختار درختی مانند به نمایش در می آیند. شاخه های درختی اغلب به وسیله 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 باره علت های یک مشکل می باشد. نمودارهای علت و معلول هم بطور ساده و هم به

<sup>1</sup> - Exponentially weighted moving averages  
<sup>2</sup> - Average Run Length

صورت مرکب ترسیم می شوند. بر این اساس برای فهم بهتر موضوع از متغیرهای متعددی (سری 20 ساله 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 قاعده 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 قاعده 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 می‌دهد. مرحله اول سنجش سری و مرحله دوم سنجش سری به صورت مدل است. یعنی سنجش به صورت داده‌ای و مدلی است. سطوحی که در تحقیقات مورد استفاده قرار می‌گیرند به عنوان مقیاس‌های سطوح سنجش مطرح می‌باشند. مقیاس‌های سطوح سنجش مقیاس‌های اسمی، رتبه‌ای، فاصله‌ای و نسبتی می‌باشند. متغیرها در قالب سطوح سنجش مورد استفاده قرار می‌گیرند. مقیاس‌های اسمی به متغیرهایی تعلق می‌گیرد که تفکیک‌پذیر باشند و بتوان آنها را در طبقه‌ها یا دسته‌های جداگانه قرار داد تا تفاوت کمیت‌های مربوط مشخص شود. در این مقیاس ملاک طبقه‌بندی مبتنی بر ویژگی‌های مشترک افراد یا رویدادها است و ملاک‌های طبقه‌بندی جنبه ریاضی - آماری ندارد. وقتی متغیر مستقل از چند سطح تشکیل شده باشد، متغیر مستقل اسمی نامیده می‌شود. برای مثال اگر هدف سطح بندی منطقه‌ای 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 می‌شود. مقیاس رتبه‌ای یا ترتیبی، نشان‌دهنده وضعیت و جایگاه یک فرد یا یک گروه یا یک چیز در داخل مجموعه‌ای از افراد می‌باشد که از طریق دادن کمیت و مقدار به تک‌تک آنها انجام داده می‌شود. در این مقیاس به تعداد افراد یا پدیده‌ها رتبه وجود دارد که مشخص‌کننده 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 رده‌ها و مقولات امکان تنظیم داده‌ها را با تعیین اولویت‌ها و ترتیب‌ها فراهم می‌کند. مقیاس فاصله‌ای صفت معینی را در خصوص افراد یا اشیاء یک گروه با رعایت تفاوت بین فاصله آنها، مشخص می‌کند. مقیاس فاصله‌ای نه تنها به ترتیب اشیاء توجه دارد بلکه فاصله بین آنها را نیز مشخص می‌سازد. علاوه بر آن در این مقیاس، مبدأ صفر وجود ندارد. مقیاس نسبتی صفت مشخصی را در خصوص افراد یا پدیده‌های یک گروه و رعایت تساوی بین فاصله آنها با یکدیگر مشخص می‌کند. 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 می‌شود. مقیاس نسبتی کامل‌ترین و دقیق‌ترین مقیاس اندازه‌گیری است. داده‌های مورد استفاده در تحقیقات اقلیمی قالب فاصله‌ای و نسبتی دارند. سطوحی که در تحقیقات مورد استفاده قرار می‌گیرند بعنوان مقیاس‌های سطوح سنجش مطرح می‌باشند. مقیاس‌های سطوح سنجش مقیاس‌های اسمی، رتبه‌ای، فاصله‌ای و نسبتی می‌باشند. متغیرهایی که در قالب سطوح سنجش مورد استفاده قرار می‌گیرند از یک نظر متغیرهای کمی و کیفی هستند. متغیرهای کمی در قالب مقیاس‌های فاصله‌ای و نسبتی مورد سنجش قرار می‌گیرند و متغیرهای کیفی در قالب مقیاس‌های اسمی و رتبه‌ای مورد سنجش قرار می‌گیرند. متغیرهای کیفی که متغیرهای مقوله‌ای نیز نام دارند، متغیرهایی هستند که مقادیر عددی به خود نمی‌

گیرند یا کمیت پذیر نیستند. متغیرهای کیفی شامل ویژگی‌هایی از قبیل جنسیت، مذهب، نژاد، زبان، سواد، دین و... هستند. متغیرهای کمی متغیرهایی هستند که برای اندازه‌گیری آنها می‌توان اعداد را به وضعیت‌های آزمودنی و قابل‌سنجش برطبق قاعده‌ای، معین نمود. این متغیرها را متغیرهای کمیت پذیر هم می‌گویند. مانند: بارش، دما، فشار، سرعت باد، رطوبت و... متناسب با متغیرها، مقیاس‌ها در قالب سطوح سنجش مورد توجه قرار می‌گیرند. مقیاس اسمی به متغیرهایی تعلق می‌گیرد که تفکیک پذیر باشند و بتوان آنها را در طبقه‌ها یا دسته‌های جداگانه قرار داد تا تفاوت کمیت‌های مربوط مشخص شود. در این مقیاس ملاک طبقه‌بندی مبتنی بر ویژگی‌های مشترک افراد یا رویدادها است و ملاک‌های طبقه‌بندی جنبه ریاضی - آماری ندارد. وقتی متغیر مستقل از چند سطح تشکیل شده باشد، متغیر مستقل اسمی نامیده می‌شود. برای مثال اگر هدف سطح بندی منطقه‌ای 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 می‌شود. مقیاس رتبه‌ای یا ترتیبی، نشان‌دهنده وضعیت و جایگاه یک فرد یا یک گروه یا یک چیز در داخل مجموعه‌ای از افراد می‌باشد که از طریق دادن کمیت و مقدار به تک‌تک آنها انجام داده می‌شود. در این مقیاس به تعداد افراد یا پدیده‌ها رتبه وجود دارد که مشخص‌کننده سلسله‌مراتب آنهاست. مثلاً "طبقه بندی ایران به چند قلمرو از نظر داشتن یا نداشتن رطوبت". بر این اساس در این مقیاس اعداد منسوب به رده‌ها و مقولات امکان تنظیم داده‌ها را با تعیین اولویت‌ها و ترتیب‌ها فراهم می‌کند. مقیاس فاصله‌ای صفت معینی را در خصوص افراد یا اشیاء یک گروه با رعایت تفاوت بین فاصله آنها، مشخص می‌کند. مقیاس فاصله‌ای نه تنها به ترتیب اشیاء توجه دارد بلکه فاصله بین آنها را نیز مشخص می‌سازد. علاوه بر آن در این مقیاس، مبدأ صفر وجود ندارد. مقیاس نسبتی صفت مشخصی را در خصوص افراد یا پدیده‌های یک گروه و رعایت تساوی بین فاصله آنها با یکدیگر مشخص می‌کند. در این مقیاس نه تنها مقدار صفت مورد نظر کمی و زیادی آن نسبت به افراد دیگر در داخل یک گروه، بلکه مقدار تفاوت آنها با یکدیگر هم مشخص می‌شود. مقیاس نسبتی کامل ترین و دقیق ترین مقیاس اندازه‌گیری است. اکثر داده‌های مورد استفاده در تحقیقات اقلیمی قالب فاصله‌ای و نسبتی دارند.

سنجش داده اقلیمی متناسب با سطوح مقیاس سنجش :

یکی از متداول‌ترین تقسیم‌بندی‌های آماری سطوح سنجش داده‌ها بر اساس عاملی و ناعاملی بودن آنها می‌باشد. بر این اساس زمانی می‌توان داده‌ها را مورد سنجش قرار داد که ماهیت عاملی یا پارامتریک و ناعاملی یا ناپارامتریک آنها مشخص شود. لذا پیش‌فرض‌های مهمی برای کاربرد آزمون‌های عاملی و ناعاملی وجود دارد که باید اقلیم‌شناس در کاربرد به آنها توجه نماید. چنانچه متغیر مورد مطالعه در تحلیل‌های کمی دارای توزیع بهنجار باشد و آزمودنی‌ها بوسیله نمونه‌گیری تصادفی بدست آمده باشند، برای تحلیل می‌توان آزمون‌های عاملی یا پارامتریک را بکار گرفت. چنانچه متغیر مورد نظر توزیع بهنجاری نداشته باشد برای تحلیل باید از آزمون‌های ناعاملی یا ناپارامتریک استفاده نمود. آزمونهای ناپارامتری معمولاً<sup>1</sup> برای متغیرهایی که در آن مقیاس‌های اسمی و رتبه‌ای اندازه‌گیری شده‌اند، بکار می‌رود. در مواردی که تعداد نمونه کوچک باشد برای متغیرهای اندازه‌گیری شده در مقیاس فاصله‌ای نیز از آزمونهای ناپارامتریک می‌توان استفاده نمود. در حالیکه آزمون‌های پارامتریک را می‌توان برای متغیرهایی که حداقل در مقیاس‌های فاصله‌ای اندازه‌گیری شده‌اند، استفاده کرد. در تحلیل‌های پارامتریک برای توزیع متغیر مورد اندازه‌گیری در جامعه‌ای که نمونه‌ای از آن بدست آمده است، پیش‌فرض‌هایی صورت می‌گیرد. اعتبار نتایج آزمون پارامتری به صادق بودن این پیش‌فرض‌ها بستگی دارد. از جمله این پیش‌فرض‌ها بهنجار بودن، استقلال مشاهدات و برابری واریانس می‌باشد. آزمونهایی که هیچ‌گونه فرضی در باره توزیع آنها وجود ندارد، بصورت تحلیل‌های آزاد توزیع مطرح می‌شوند که اغلب ناپارامتریک هستند. در این مورد آزمون‌های فرضیه‌ای که استفاده می‌شوند می‌توانند یک دامنه و یا دو دامنه باشند. برای بررسی رابطه سطوح مذکور می‌توان از ضرائب زیر استفاده کرد:

چنانچه متغیرها، اسمی (مانند جنسیت، مذهب و ...) باشند برای تعیین میزان همبستگی بین دو متغیر از ضریب لاندا ( $\lambda$ ) و فی ( $\phi$ ) استفاده می‌شود<sup>1</sup>.

ضریب لاندا ( $\lambda$ ):

ضریب همبستگی لاندا برای محاسبه حدود رابطه بین متغیرهای اسمی بکار می‌رود. این ضریب یک طرفه است و رابطه بین دو متغیر را فقط در یک جهت نشان می‌دهد. دامنه ضریب، بین صفر و یک متغیر است. صفر نشان‌دهنده عدم رابطه بین دو متغیر است و یک نشان می‌دهد که بین دو متغیر رابطه کامل وجود دارد. برای محاسبه ضریب لاندا از معادله زیر استفاده می‌شود:

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

<sup>1</sup> - جاوری، مجید: مقدمه‌ای بر روش‌های تحقیق در علوم انسانی با تأکید بر جغرافیا، انتشارات طلور، 1384.

$f_i$  بیشترین فراوانی هر طبقه متغیر مستقل ،  $F_d$  بیشترین 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 ، رابطه مذکور چگونه است ؟

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

$$\sum f_i = 50 + 50 + 60 = 160$$

$$\sum F_d = 110$$

$$\lambda = \frac{160 - 110}{360 - 110} = \frac{50}{250} = .2$$

ضریب فی  $(\phi)$  :

برای تعیین همبستگی متغیرهای اسمی که هر کدام دو طبقه داشته باشند از ضریب فی می توان استفاده نمود . دامنه 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 محاسبه ضریب فی از معادله زیر استفاده می شود :

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

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

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

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

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

سنجش درجه قوت رابطه بین متغیرهای ترتیبی :

برای سنجش رابطه بین متغیرهای 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 رابطه 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 رابطه بین متغیرهای فاصله ای و نسبتی، می توان از ضریب همبستگی متغیرها مانند ضریب همبستگی پیرسون استفاده کرد . نحوه 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 بیشمار 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 فرضیه تحقیق اقلیم شناس این باشد که با افزایش ارتفاع کوههای البرز بارش کم می شود، این فرضیه توصیفی است و چنانچه فرضیه اقلیم شناس این باشد که بارش از نوع برف نسبت به باران با افزایش ارتفاع البرز کاهش می یابد، فرضیه تفاوتی و چنانچه فرضیه اقلیم شناس میزان ارتباط بارش در ارتفاعات البرز با سرعت توده هوا باشد، فرضیه مورد استفاده فرضیه رابطه ای است. چنانچه فرضیه اقلیم شناس بررسی رابطه 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. فرضیه های تحقیق در مقام آزمون بصورت مخالف یا صفر نوشته می شوند و زمانی که استنباط آماری موجب رد فرضیه صفر بشود، فرض تحقیق پذیرفته می شود. اگر فرضیه تحقیق مبین تفاوت بین دو متغیر باشد، در اینصورت رابطه دو میانگین فوق چنین است  $(\mu_1 \neq \mu_2)$  ولی اگر فرضیه تحقیق، مبین جهت تفاوت فوق نیز باشد، یعنی اگر فرض، مبین بیشتر یا کمتر بودن یکی نسبت به دیگری باشد، در این حالت رابطه بصورت  $(\mu_1 > \mu_2 \quad \mu_1 < \mu_2)$  خواهد بود. اگر احتمال رابطه فوق که به صورت یک مقدار عددی از آزمون بدست آمده است، تحت فرضیه صفر برابر یا کمتر از  $\alpha$  یعنی سطح معنی داری که غالباً به شکل درصد  $(\alpha = 1\%, \alpha = 5\%)$  مورد استفاده است، باشد فرض صفر رد شده و فرض تحقیق قبول می شود و برعکس. سطح معنی داری در تحلیل های آزمون ها، مبین پذیرش یا عدم پذیرش فرض صفر در سطح معین (99% یا 95% یا 90%) یا (0/01 یا 0/05 یا 0/1) می باشد. همچنین درجه آزادی در این قالب مورد استفاده قرار می گیرد. درجه آزادی به میزان مستقل بودن مشاهدات از یکدیگر اشاره می کند. با توجه به مطالب جدول شماره 1 سطح سنجش و آماره های مربوط را نشان می دهد.

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

### 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 مؤلفه مهمی از فرآیند های تصادفی را تشکیل می دهند . مشخصه های مدل ساده 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 شماره 2 مشاهده کرد.

جدول شماره 2: آزمون های عاملی و ناعاملی سنجش سری تصادفی:

نام آزمون	آزمون	قاعده آزمون	قاعده تصمیم گیری	نتیجه آزمون
توالی	$Z = \frac{ R - \mu_R }{\sigma_R}$	سری در سطح معین	$Re\ ject : H_0 \text{ if }  Z  > Z_{\alpha/2}$	اگر مقدار نمره آزمون کمتر از مقدار نمره

<p>جدول باشد فرض صفر رد می شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری تصادفی می باشد .</p>		<p>تصادفی است <math>H_0 =</math>  سری در سطح معین تصادفی نیست <math>H_a =</math></p>		
<p>اگر نمره آزمون 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 .</p>	<p><math>Rejct : H_0 \text{ if }  Z  &gt; Z_{\alpha/2}</math></p>	<p>سری تصادفی است <math>H_0 :</math>  سری تصادفی نیست : <math>H_a</math></p>	<p><math>Z = \left  \frac{U - \mu_U}{\sigma_U} \right </math></p>	<p>نقاط گردش</p>
<p>اگر مقدارنمره کمتر از مقدار نمره جدول باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری تصادفی است .</p>	<p><math>Rejct : H_0 \text{ if }  Z  &gt; Z_{\alpha/2}</math></p>	<p>سری بدون روند است <math>H_0 :</math>  سری روند دارد <math>H_\alpha :</math></p>	<p><math>Z = \frac{V - \mu_V}{\sigma_V}</math></p>	<p>نشانه</p>
<p>اگر نمره آزمون از نمره جدول کمتر باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری بدون روند است.</p>	<p><math>Rejct : H_0 \text{ if }  Z  &gt; Z_{\alpha/2}</math></p>	<p>سری روند <math>H_0 :</math> ندارد  سری دارای روند است : <math>H_\alpha</math></p>	<p><math>Z = \frac{r_s - \mu_{rs}}{\sigma_{rs}}</math></p>	<p>دنیلس</p>

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

<p>اگر نمره آزمون از نمره جدول کمتر باشد ، فرض صفر رد نمی شود و نتیجه گرفته می شود که با احتمال 95 درصد اطمینان سری تصادفی است.</p>	$\text{Reject} : H_0 \text{ if } Q_m > X_\alpha^2(m)$	$H_0 : \rho_k = 0$ $H_a : \rho_k \neq 0$	$Q_m = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-k}$	پرایس-بوکس Q
---	---	--	---	--------------

مطالعه 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 نمره آزمون از نمره 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 شماره 3 مشخص شده است .

جدول شماره 3: آزمون سنجش سری گرایش دار:

نام آزمون	آزمون	قاعده آزمون	قاعده تصمیم گیری	نتیجه آزمون
$t$	$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	$H_0 : \rho = 0$ $H_a : \rho \neq 0$	$Reject : H_0 \text{ if }  t_r  > t_{\alpha/2}$	اگر مقدار نمره آزمون کمتر از مقدار نمره 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 معمولاً "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 دوره 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 محاسبه آزمون مذکور می توان از معادله زیر استفاده کرد :

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

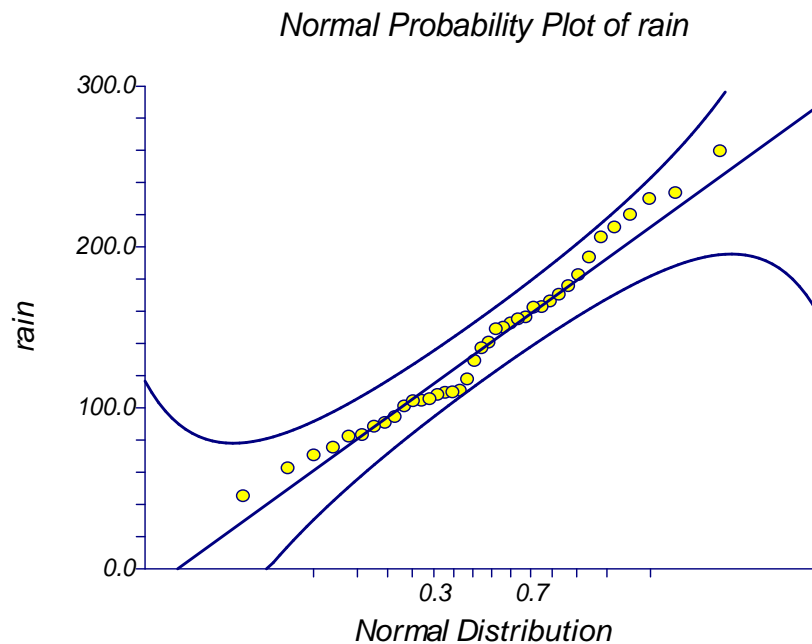
قاعده تصمیم گیری :

$$Reject : H_0 \text{ if } : H > X_{\alpha}^2(L-1)$$

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

تغییرات دوره ای را در یک سری زمانی که خیزهای بلند مدت (غیر از روند و نوسانات فصلی) دارند می‌توان 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 می‌باشد و معمولاً "طولانی تر از یک سال می‌باشد. برای سنجش تغییرات دوره ای لازم است از آزمون های آماری استفاده کرد. یکی از شیوه های سنجش تغییرات دوره ای استفاده از آزمون 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 شماره 8 که هدف آن سنجش بهنجاری سری بارش ایستگاه کاشان است توجه نمود.

موقعیت شماره 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<sup>1</sup>:

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 شماره 1:

وقتی هدف تحقیق در اقلیم شناسی بررسی رابطه دو متغیر اسمی باشد می توان رابطه متغیر ها را در قالب جدول توافقی متناسب با رابطه 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 محقق 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 :

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

<sup>1</sup> - به کتاب شیوه های تجزیه و تحلیل کمی در اقلیم شناسی (با تاکید بر مدل های روند ) مؤلف مراجعه شود.



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

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

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

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

درجه 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 و فرض صفر رد

نمی شود و نتیجه گرفته می شود که رابطه متغیرها براساس شانس است و واقعی نیست.

مقدار ضریب خی دو نشانگر معنی داری سری از نظر آماری است و رابطه متغیرها را مشخص نمی سازد. لذا برای مشخص کردن میزان همبستگی یا رابطه بین متغیرها 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<sup>1</sup> استفاده کرد. فرمول تصحیح یتس برابر است با:

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

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<sup>2</sup> بعنوان یک آزمون ناعاملی (مشخصه های کاربرد آزمونها 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 محاسبه آزمون استفاده کرد. نمونه های مورد نظر در محاسبه آزمون را می توان به انواع زیر تقسیم نمود<sup>3</sup>:

- 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 نمونه 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

<sup>1</sup> - Yates Correction

<sup>2</sup> - Mann-Whitney Test

<sup>3</sup> - 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 مرتبه  $(u)$  هر نمره را به طور جداگانه در جامعه حساب کنید .
- مقدار مرتبه محاسبه شده را با مقدار بحرانی جدول مقایسه نمود و نتیجه گیری نمایید . اگر مقدار هر مرتبه  $(u)$  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 نمره 12 در روش V از هیچ کدام نمره های روش G بزرگتر نیست ، نمره 13 روش V از هیچ نمره روش G بزرگتر نمی باشد ، نمره 15 روش V از دو نمره (13 و 14) روش G بزرگتر است . نمره 16 روش V از دو نمره (13 و 14) روش G بزرگتر است . نمره 18 روش V از سه نمره (13 ، 14 و 17) روش G بزرگتر است و الی آخر

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

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 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 نمونه متوسط :

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

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

$R_1$  مجموع رتبه های نمونه اول و  $n_1$  حجم نمونه اول می باشد . مرتبه های نمونه دوم را می توان از رابطه زیر بدست آورد:

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

$R_2$  مجموع رتبه های نمونه دوم و  $n_2$  حجم نمونه دوم می باشد.

مثال :

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

منطقه 1 $(n_1)$	منطقه 2 $(n_2)$
-----------------	-----------------

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

مراحل عملیات :

- رتبه نمرات باید محاسبه شود (پایین ترین نمره رتبه اول و الی آخر) .
- مجموع رتبه های هر نمونه بدست آورده شود.
- مقادیر مرتبه ها با استفاده از معادلات آزمون محاسبه شود .
- مقدار مرتبه کوچکتر محاسبه شده با مقدار جدول بحرانی مقایسه می شود . اگر مرتبه ( $u$ ) محاسبه شده بزرگتر از مقدار مرتبه 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 :

رتبه نمرات :

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

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 مرتبه جدول (23) است، لذا فرض صفر رد نمی شود و با احتمال 95 درصد نتیجه گرفته می شود بین نمرات خدمات شهری در دو منطقه 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 - وینتی به توزیع بهنجاری سری نزدیک می باشد. لذا بعد از محاسبه آزمون مقدار Z با عدد 1196 (در سطح 0105) 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 ناحیه آن تحت بررسی باشد. چنانچه تعداد افراد مهاجر در ناحیه اول 40 نفر و تعداد افراد مهاجر در ناحیه 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 مراحل محلی مقدار مرتبه (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 (7/52) از مقدار ضریب 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 - وینتی با استفاده از نرم افزار SPSS :

آزمون t برای تحلیل مقایسه میانگین نمونه آماری با جامعه آماری :

از آزمون t می توان به عنوان یک آزمون پارامتری یا عاملی برای مقایسه 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

می‌شود. زمانی که حجم نمونه از 30 بیشتر می‌شود تفاوت بین توزیع آزمون مذکور با توزیع بهنجار خیلی ناچیز می‌شود. لذا توزیع و پراکندگی آزمون  $t$  بیشتر از پراکندگی توزیع بهنجاری است و این موضوع با کاهش حجم نمونه بیشتر می‌شود. همچنین هر چه درجه آزادی و حجم نمونه بزرگ‌تر باشد مقدار  $t$  جدول برای رد فرض صفر کوچک‌تر می‌شود. 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 فاصله‌ای یا نسبی می‌توان از این آزمون استفاده کرد. برای کاربرد آزمون  $t$  در مطالعات جغرافیایی می‌توان در قالب انواع یک نمونه ای<sup>1</sup>، نمونه‌های وابسته<sup>2</sup> و دونمونه مستقل<sup>3</sup> استفاده کرد.

- کاربرد آزمون  $t$  یک نمونه ای در تحلیل‌های اقلیمی :

آزمون  $t$  استیودنت وقتی استفاده می‌شود که بخواهند ببینند آیا میانگین یک متغیر تفاوت معنی‌داری با یک مقدار آزمون دارد یا خیر. یا اگر بخواهند بررسی کنند یا میانگین و انحراف معیار مشخص، نمونه 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 نمونه‌ها، متغیر مستقل در سطح سنجش مقیاس اسمی و متغیر وابسته در سطح سنجش فاصله‌ای 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 منطقه‌ای که بعد از بدست آوردن  $t$  و تعیین درجه آزادی با مراجعه به جدول آماره  $t$  می‌توان به سؤال بالا جواب داد، اگر مقدار قدر مطلق  $t$  به دست آمده از مقدار  $t$  جدول بیشتر باشد، فرض صفر رد می‌شود، یعنی تفاوت بین دو میانگین معنی‌دار است و برعکس اگر مقدار  $t$  بدست آمده از مقدار بحرانی جدول کمتر باشد فرض صفر رد نمی‌شود و نتیجه گرفته می‌شود با احتمال معین بین متغیرها اختلاف معنی‌داری وجود ندارد. برای محاسبه آزمون مذکور می‌توان از فرمول زیر استفاده کرد :

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

$\bar{X}$  میانگین نمونه،  $\mu$  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، مقایسه یک گروه با یک عدد ثابت به حالت‌های مختلف قابل بررسی می‌باشد. برای تبیین بیشتر به مثال‌های زیر توجه نمایید:

- هر گاه انحراف معیار جامعه معلوم باشد:

مثال: اگر استدلال و فرض شود که میانگین بارش ایران 240 میلی‌متر در سال 1388 باشد، برای بررسی این استدلال به طور تصادفی از میان ایستگاه‌های اقلیم‌شناسی و سینوپتیکی ایران 100 ایستگاه انتخاب شده است نتایج نشان می‌دهد که میانگین بارش ایران 230 میلی‌متر با انحراف معیار 20 میلی‌متر است. حال مقدار آزمون  $t$  آن را حساب کنید :

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

$$df = 100 - 1 = 99$$

چون مقدار قدر مطلق  $t$  (5) محاسبه شده با درجه آزادی 99 و سطح معنی‌داری 0/05 از مقدار بحرانی جدول (1198) بزرگتر است، فرض صفر رد می‌شود و نتیجه گرفته می‌شود با احتمال 95 درصد اطمینان اختلاف معنی‌داری 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 نمونه بزرگ (بیش از 30) بدست آمده باشد.

مثال:

میزان بارش در 40 ایستگاه اندازه‌گیری 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؟

$$H_0: \mu = 220$$

$$H_a: \mu \neq 220$$

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

با توجه به قدر مطلق  $t$  محاسبه شده (8/43) با درجه آزادی 39 از مقدار ضریب بحرانی جدول (2/021) در سطح 0/05 بزرگتر است لذا فرض صفر رد می‌شود و نتیجه گرفته می‌شود که میانگین بارش ایستگاه مورد نظر 220 میلی متر نمی‌باشد.

- 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 می‌باشد. 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

$$H_0: \mu = 150$$

$$H_a: \mu \neq 150$$

$$\bar{X} = 142$$

$$S = 38.887$$

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

چون مقدار  $t$  محاسبه شده (0/65) از مقدار ضریب بحرانی جدول (2/262) کوچکتر است فرض صفر رد نمی‌شود لذا با احتمال 95 درصد اطمینان نتیجه گرفته می‌شود 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					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
VAR00002	<u>-6.51</u>	9	.532	-8.00000	-35.8183	19.8183

- کاربرد آزمون  $t$  با نمونه‌های وابسته در تحلیل‌های استنباطی:

از این آزمون در تحلیل‌های جغرافیایی استفاده می‌شود که هر نمونه در دو وضعیت متفاوت مورد سنجش قرار گیرد. یعنی میانگین نمونه در دو مرحله متفاوت معنی دار مورد سنجش و بررسی قرار داده می‌شود. در این آزمون سنجیده می‌شود 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 نمونه‌ها، متغیرهای مستقل (مقیاس اسمی) و متغیر وابسته (فاصله ای یا نسبی) بودن و در هر نمونه آماری دو اندازه‌گیری وجود داشته باشد، توجه نمود. برای استفاده از این آزمون می‌توان از رابطه زیر استفاده کرد<sup>1</sup>:

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

$d$  تفاضل نمره‌ها و  $N$  حجم نمونه می‌باشد. برای تبیین بیشتر موضوع به مثال زیر توجه شود:

مثال:

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

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

<sup>1</sup> - 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 مقایسه 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 نمونه 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 رابطه زیر استفاده کرد:

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

$$df = n_1 + n_2 - 2$$

به ترتیب میانگین نمونه دوم و اول  $n_1$  و  $n_2$  به ترتیب حجم نمونه 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		15 119	0	39/553

$$\bar{X}_1 = \frac{80.25}{6} = 13.375$$

$$\bar{X}_2 = \frac{119.5}{9} = 13.278$$

$$S_1^2 = \frac{24.721}{6} = 4.12$$

$$S_2^2 = \frac{39.553}{9} = 4.39$$

تقریباً واریانس نمونه ها ( $S^2$  و  $S^1$ ) 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 مقایسه 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 اثرگذاری 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 مقایسه 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. قاعده آزمون به صورت زیر مورد نظر است:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_\alpha : \mu_1 \neq \mu_2 \neq \dots \neq \mu_k$$

آزمون:

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

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

مراحل محاسبه آزمون :

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

انجام مراحل :

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

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

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

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

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

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 .

محاسبه 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 تحلیل‌های اقلیمی :

کاربرد همبستگی‌ها در تحلیل‌های اقلیمی<sup>1</sup> :

در مطالعات اقلیمی بررسی رابطه دو متغیر از طریق توزیع نقاط از اهمیت زیادی برخوردار است. شاخص آماری که رابطه دو متغیر را در توصیف می‌کند، ضریب همبستگی می‌باشد. این ضریب در مطالعات اندازه ارتباط بین دو متغیر را نشان می‌دهد. در تحقیقات اقلیم‌شناسی این شاخص نشان می‌دهد 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 داده‌ها کاربرد آنها متفاوت می‌باشد. در مطالعات اقلیمی وقتی ضریب همبستگی جهت توصیف رابطه خطی استفاده می‌شود ضریب همبستگی مقادیری از صفر تا یک خواهد داشت. برای مثال رابطه خطی بین میزان بارش با مقدار ارتفاع هدف تحقیق می‌باشد. رابطه بین این دو متغیر به صورت مثبت (افزایشی دو متغیر) منفی (افزایش یک متغیر به ازای کاهش متغیر دیگر) مد نظر قرار می‌گیرد. به هر حال هرگاه هدف بررسی ارتباط بین دو متغیر باشد، باید از همبستگی سری‌ها استفاده کرد. دامنه ضریب همبستگی  $\pm$  یک است. به طوری که :

- ضریب همبستگی منفی به معنی رابطه معکوس میان دو متغیر است.
- ضریب همبستگی مثبت به معنی رابطه مستقیم بین دو متغیر است.
- ضریب همبستگی صفر به معنی عدم رابطه بین دو متغیر می‌باشد.

با توجه به موضوع تحقیق، تفسیر ضریب همبستگی در تحقیقات متفاوت است. برای محاسبه ضریب همبستگی متناسب با فرض تحقیق و متغیرهای مورد استفاده باید از نوع خاصی ضریب همبستگی استفاده کرد. جدول شماره 4-1 روش‌های مختلف محاسبه همبستگی را در بین متغیرهای مختلف نشان می‌دهد.

جدول شماره 4-1: روش‌های مختلف محاسبه همبستگی بین متغیرهای مختلف:

روش	نماد	متغیر اول	متغیر دوم	ملاحظات
همبستگی پیرسون	$r$	پیوسته	پیوسته	دقیق‌ترین روش با کمترین خطای معیار
رتبه‌ای	$Rho$	رتبه‌ای	رتبه‌ای	غالباً برای تعداد آزمودنی‌های کمتر از 30 نفر استفاده می‌شود.
کندال	$\tau$	رتبه‌ای	رتبه‌ای	وقتی تعداد آزمودنی‌ها کمتر از 10 نفر است استفاده می‌شود.
دو رشته‌ای	$r_{bis}$	دو	ارزشی	مقدار آن گاهی از یک بیشتر می‌شود

<sup>1</sup> - برای اطلاع بیشتر به کتاب مقدمه‌ای بر روش‌های 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 محاسبه نمره های آزمودنی 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 غالباً کمتر از ضریب دورشته ای است
تتراکوریک	$r_t$	دو ارزشی ساختگی	دو ارزشی ساختگی	زمانی بکار می رود که بتوان دو متغیر را در یک نقطه معین نصف کرد.
فی	$\phi$	دو ارزشی واقعی	دو ارزشی واقعی	برای محاسبه 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	$\eta$	پیوسته	پیوسته	به منظور تعیین همبستگی غیر خطی بکار می رود.

برای محاسبه 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 محاسبه رابطه بین متغیرها از

ضریب همبستگی پیرسون که بعنوان ضریب همبستگی از نوع پارامتری یا عاملی قلمداد می شود می توان استفاده کرد . برای محاسبه ضریب همبستگی پیرسون (داده های خام ) از معادله ذیل استفاده می شود :

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

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

$$r_{xy} = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}} = \frac{6 \times 56 - (15)(21)}{\sqrt{(330 - 225)(546 - 441)}} = 0.2$$

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 :

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

$(Cov_{xy})^2$  واریانس مشترک بین متغیرهای  $x$  و  $y$  و  $V_x^2 V_y^2$  کل واریانس متغیرها می باشد. چنانچه نمرات استاندارد شده را داشته باشیم ، محاسبه ضریب همبستگی آن از طریق معادله ذیل بدست می آید :

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

$Z$  ها نمرات استاندارد شده  $x$  و  $y$  هستند . برای استاندارد کردن نمرات همان گونه که قبلاً اشاره شد می توان از رابطه 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 معادله :

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

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

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

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

سپس:

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

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

<i>x</i>	<i>y</i>	<i>z<sub>x</sub></i>	<i>z<sub>y</sub></i>	<i>z<sub>x</sub>·z<sub>y</sub></i>
<b>1</b>	<b>3</b>	<b>-0/88</b>	<b>-0/29</b>	<b>0/255</b>
<b>3</b>	<b>5</b>	<b>0/29</b>	<b>0/88</b>	<b>0/2552</b>
<b>2</b>	<b>2</b>	<b>-0/29</b>	<b>-0/88</b>	<b>0/2552</b>
<b>0</b>	<b>4</b>	<b>-1/46</b>	<b>0/29</b>	<b>-0/4234</b>
<b>4</b>	<b>1</b>	<b>0/88</b>	<b>-1/46</b>	<b>-0/2848</b>
<b>5</b>	<b>6</b>	<b>1/46</b>	<b>1/46</b>	<b>2/136</b>
جمع				<b>1/189</b>

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

ضریب همبستگی رتبه ای اسپیرمن:

وقتی مقیاس اندازه گیری داده ها رتبه ای باشد، از ضریب همبستگی اسپیرمن برای محاسبه رابطه متغیرها استفاده می شود. برای

محاسبه ضریب همبستگی اسپیرمن از معادله 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 نتیجه رتبه دانشجویان در دو درس است، داشته باشیم، ضریب همبستگی آنها را محاسبه کنید؟

شماره افراد	رتبه در درس اقلیم	رتبه در درس آمار
<b>1</b>	<b>1</b>	<b>2</b>
<b>2</b>	<b>2</b>	<b>3</b>
<b>3</b>	<b>3</b>	<b>1</b>
<b>4</b>	<b>4</b>	<b>5</b>
<b>5</b>	<b>5</b>	<b>4</b>

متناسب با فرمول:

شماره افراد	رتبه در درس اقلیم	رتبه در درس آمار	تفاضل رتبه ها (D)	D <sup>2</sup>
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 متغیر  $y$  است لازم است از ضریب تعیین استفاده شود. اگر بر حسب درصد بخواهیم ضریب تعیین را حساب کنیم می توان از رابطه ذیل استفاده کرد:

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

اگر مثال بالا را در نظر بگیریم،  $36 = (0.6)^2 \times 100 = D$  درصد پراکندگی بین دو متغیر مشترک است. قابل اشاره است که ضریب تعیین هیچگاه صفر نخواهد بود و دامنه 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:

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

نتیجه بدست آمده  $t$  با توجه به درجه 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 متغیر در سطح اطمینان  $(0/05)$  و  $(0/01)$  معنی دار است.

## کاربرد رگرسیون در تحلیل های اقلیمی<sup>1</sup>:

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 اقلیمی بعضی مواقع هدف پیش بینی یک متغیر بوسیله متغیر دیگر است. لذا برای پیش بینی یک متغیر توسط متغیر دیگر باید به یک رابطه منظم و نظامدار بین آن دو متغیر توجه نمود. بر این اساس در تحلیل

<sup>1</sup> - به مقاله تجزیه و تحلیل داده های اقلیمی با استفاده از تحلیل رگرسیونی، از مؤلف مراجعه شود.

رگرسیون‌ی توجه به تغییر در متغیرهای مستقل که مقادیر متغیرهای وابسته را چگونه تحت تأثیر قرار می‌دهد، حائز اهمیت است. وجه تمایز بین متغیرهای مستقل و وابسته همیشه کاملاً<sup>1</sup> روشن نمی‌باشد و آن بستگی به اهداف تحقیق دارد. واژه رگرسیون<sup>1</sup> به معنای برگشت است و نشان می‌دهد که مقدار یک متغیر به متغیر دیگری بر می‌گردد. این واژه اولین بار توسط فرانسیس گالتن<sup>2</sup> در سال 1877 به کار برده شده است. با توجه تعداد متغیرها، در مطالعات اقلیمی علاوه بر رگرسیون خطی ساده از رگرسیون چندگانه استفاده می‌شود. رگرسیون یا تحلیل کمترین مربعات یک شیوه آماری برای برآورد رابطه تابعی بین یک متغیر وابسته و یک یا چند متغیر مستقل یا پیش‌بینی‌کننده<sup>3</sup> است. در تحلیل رگرسیون متغیرهای توضیحی ممکن است غیر قابل اندازه‌گیری باشند که به آنها متغیرهای کیفی گفته می‌شود. یکی از راههای واردکردن متغیرهای کیفی در مدل رگرسیون استفاده از متغیرهای مجازی<sup>4</sup> است. لذا در تحلیل رگرسیون به تعیین روابط بین متغیرها ( رگرسیون توصیفی ) و تحلیل روابط بدست آمده ( رگرسیون استنباطی ) توجه می‌شود. شکل عمومی رگرسیون متغیر وابسته و متغیرهای مستقل ممکن است یک سری زمانی باشند و یا هم نباشند. در تحلیل این روش باید به فرض‌های زیر توجه داشت :

- 1- رابطه 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 می‌توان از نمودارهای احتمال بهنجار، نمودارهای پراکنش مانده‌ها در برابر سری پیش‌بینی شده، نمودار پراکنش مانده‌ها در برابر متغیرهای مستقل و آزمون‌های آماری استفاده نمود. مدل رگرسیون زمانی صحیح خواهد بود که مانده‌ها در اطراف صفر بطور تصادفی توزیع شده باشند. وقتی در یک مدل رگرسیون از یک متغیر مستقل استفاده می‌شود، مدل مذکور را مدل رگرسیون خطی ساده<sup>5</sup> گویند و وقتی بیش از یک متغیر مستقل در یک مدل استفاده می‌شود، به آن مدل، مدل رگرسیون چندگانه<sup>6</sup> گفته می‌شود. لازم است برای بررسی رگرسیون رابطه بین سری‌ها و طول سری‌های بررسی شود. برای تحلیل رگرسیون داده‌های اقلیمی باید به مراحل هم‌چون بیان مسئله، انتخاب متغیرها، جمع‌آوری سری یا داده‌ها، تشخیص و تعیین مدل، انتخاب روش برازش، مناسب بودن مدل و کاربرد مدل برای

---

*Regression* -<sup>1</sup>  
*Francis Galton* -<sup>2</sup>  
*Predictor Variable* -<sup>3</sup>  
*Dummy Variable* -<sup>4</sup>  
*Simple Linear Regression* -<sup>5</sup>

*Multiple Regression* -<sup>6</sup>

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

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

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

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

قواعد آزمون ها :

$$\text{Reject} : H_0 \text{ if } |t| > t_{\alpha/2}$$

$$\text{Reject} : H_0 \text{ if } |t_r| > t_{\alpha/2}$$

$$\text{Reject} : H_0 \text{ if } : F > F_\alpha$$

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

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

**2-4- کاربرد رگرسیون خطی ساده :**

در رگرسیون خطی ساده هدف این است که با محاسبهٔ عامل‌های  $\beta_0$  و  $\beta_1$ ، مدل خطی ساده‌ای را به داده‌های برازش اختصاص داده شود. روشی که برای برازش کردن داده‌ها بکار می‌رود روش کمترین مربعات یا  $LS$  می‌باشد. در این روش معادلهٔ خط برازش طوری شکل می‌گیرد که مجموع مربعات توان‌های دوم انحرافات عمودی از خط برازش شده به حداقل برسد. برآورد ضرایب به روش حداقل مربعات صورت می‌گیرد. برای محاسبهٔ ضرایب می‌توان از معادلات زیر استفاده کرد:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (5-4)$$

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

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

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

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

$\beta_0$  عرض از مبدأ،  $\beta_1$  شیب خط،  $Y$  متغیر وابسته و  $X$  متغیر مستقل می‌باشد. با توجه معادلات بالا مدل رگرسیون خطی ساده را می‌توان به صورت معادلهٔ زیر نشان داد:

$$Y = \beta_0 + \beta_1 x + \varepsilon \quad (10-4)$$

در معادلهٔ بالا  $\varepsilon$  عامل خطا می‌باشد. با توجه به مدل بالا معادلهٔ خط برازش شده را می‌توان بصورت معادلهٔ زیر نوشت:

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

مقدار خطا حاصل از تفاضل مقدار اولیه و مقدار پیش‌بینی شده می‌باشد. به طریقی از حاصل تفاضل مقادیر پیش‌بینی شده از مقدار اصلی سری مقدار خطاها یا مانده‌ها<sup>1</sup> بدست می‌آید.

---

<sup>1</sup> 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 -متغیرها مستقل از یکدیگرند .

جدول شماره 4-1 ضرایب معادله 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 شماره 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 معادله رگرسیون برابر است با :

$$y = 18.7 + 0.00856 t$$

حال اگر فرض شود که  $H_0 : \beta_0 = 0$  و  $H_0 : \beta_1 = 0$  انجام شود ، لذا :

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

- آزمون 1 :

برای این آزمون از آماره  $t$  استفاده می شود . مقدار آن برابر است با :

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

از طرفی ناحیه 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 معادله رگرسیونی صفر نیست . روش دیگر جهت این آزمون استفاده از آماره *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.

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

- آزمون 2 :

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 معادله رگرسیونی بدون شیب رگرسیونی است . کل نتیجه آزمون ها را می توان به صورت جدول شماره 4-2 نشان داد .

جدول شماره 4-2 : نتایج کلی آزمون های مورد استفاده برای دمای یزد :



<i>Regression Estimation Section</i>		
	<i>Intercept</i>	<i>Slope</i>
<i>Parameter</i>	<i>B(0)</i>	<i>B(1)</i>
<i>Regression Coefficients</i>	18.6677	0.0086
<i>Lower 95% Confidence Limit</i>	18.2853	-0.0081
<i>Upper 95% Confidence Limit</i>	19.0502	0.0252
<i>Standard Error</i>	0.1888	0.0082
<i>Standardized Coefficient</i>	0.0000	0.1687
<i>T Value</i>	98.8957	1.0410
<i>Prob Level (T Test)</i>	0.0000	0.3046
<i>Reject H0 (Alpha = 0.0500)</i>	Yes	No
<i>Power (Alpha = 0.0500)</i>	1.0000	0.1736
<i>Regression of Y on X</i>	18.6677	0.0086
<i>Inverse Regression from X on Y</i>	12.8210	0.3009
<i>Orthogonal Regression of Y and X</i>	18.6673	0.0086

#### 3-4- رگرسیون و تحلیل واریانس سری:

ایده 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  $Y$  را نتوان به تغییرات  $X$  نسبت داد آنگاه وجود رابطه 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 شماره 4-3 نشان داد.

جدول شماره 4-3: تحلیل واریانس سری دمای یزد :

<i>Analysis of Variance Section</i>						
<i>Source</i>	<i>DF</i>	<i>Sum of Squares</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>Probe Level</i>	<i>Power (5%)</i>
<i>Intercept</i>	1	13841.37	13841.37			
<i>Slope</i>	1	0.3622045	0.3622045	1.0838	0.3046	0.173
<i>Error</i>	37	12.36575	0.3342096			
<i>Adj. Total</i>	38	12.72796	0.3349463			
<i>Total</i>	39	13854.1				

$s = \text{Square Root}(0.3342096) = 0.5781086$

با توجه به جدول تحلیل واریانس (*ANOVA*) رابطه 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 :

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

- آماره آزمون :

$$F = 1/084$$

- ناحیه بحرانی :

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

بنابراین فرض  $H_0 : \beta_1 = 0$  رد نمی شود . به عبارت دیگر دلایل کافی دال بر رابطه خطی بین متغیرها وجود ندارد . با توجه به جدول شماره 3 مجموع مربعات مدل  $0/362^1$  ، مجموع مربعات خطا  $12/365^2$  و مجموع مربعات کل  $12/278^3$  می باشند. از نکات قابل توجه در جدول خلاصه تحلیل واریانس شاخص ضریب تعیین می باشد . جدول شماره 4 ، ضریب تعیین را نشان می دهد .

جدول شماره 4-4 : خلاصه تحلیل واریانس سری دمای یزد :

---

$SS_{Regression} - 1$

$SS_{Residual} - 2$

$SS_{Total} - 3$

<i>Summary Section</i>			
<i>Parameter</i>	<i>Value</i>	<i>Parameter</i>	<i>Value</i>
<i>Dependent Variable</i>	<i>C2</i>	<i>Rows Processed</i>	<i>40</i>
<i>Independent Variable</i>	<i>C1</i>	<i>Rows Used in Estimation</i>	<i>39</i>
<i>Frequency Variable</i>	<i>None</i>	<i>Rows with X Missing</i>	<i>1</i>
<i>Weight Variable</i>	<i>None</i>	<i>Rows with Freq Missing</i>	<i>0</i>
<i>Intercept</i>	<i>18.6677</i>	<i>Rows Prediction Only</i>	<i>0</i>
<i>Slope</i>	<i>0.0086</i>	<i>Sum of Frequencies</i>	<i>39</i>
<i>R-Squared</i>	<i>0.0285</i>	<i>Sum of Weights</i>	<i>39.0000</i>
<i>Correlation</i>	<i>0.1687</i>	<i>Coefficient of Variation</i>	<i>0.0307</i>
<i>Mean Square Error</i>	<i>0.3342096</i>	<i>Square Root of MSE</i>	<i>0.578108</i>

در تحلیل رگرسیون ضریب تعیین تعدیل شده  $R_{adj}^2 = 1 - \frac{SS_{res} / (n-p)}{SS_T / (n-1)}$  وجود دارد که در ارزیابی و مقایسه مدل های رگرسیون 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 بزند

را یا به طریقی معنی دار بودن مدل رگرسیون را مشخص می سازد ، را می توان استفاده کرد. دامنه ضریب تعیین  $0 \leq R^2 \leq 1$  است . این ضریب مناسب بودن مدل را مشخص نمی کند بلکه درصد تغییرات متغیر وابسته است که بوسیله متغیر مستقل تبیین می شود . اگر در تحلیل رگرسیون نقاط تکراری باشد ضریب تعیین یک نمی شود ، چون مدل تحلیل رگرسیون نمی تواند میزان تغییرات مربوط به مانده ها یا خطا سری را تعیین کند . وقتی که مدل رگرسیون صحیح باشد مقدار ضریب تعیین بستگی به دامنه تغییرات متغیر مستقل دارد ، یعنی هر چقدر مقدار پراکنش متغیر مستقل بیشتر باشد مقدار ضریب تعیین بیشتر خواهد بود و برعکس . از طرفی ضریب تعیین دلالت بر بالا بودن مقدار شیب خط رگرسیون نمی باشد . چون مقدار ضریب تعیین مقدار شیب خط رگرسیون را اندازه گیری نمی کند . از طرفی ضریب تعیین با ضریب همبستگی رابطه دارد یعنی توان دوم ضریب همبستگی ضریب تعیین است و مقدار آن با داشتن ضریب تعیین بین مقدار خطا و متغیر وابسته برابر با  $\sqrt{1-R^2}$  است . هرچه مقدار همبستگی بین متغیرها بیشتر باشد پراکنش متغیرها به خط رگرسیون نزدیکتر می شوند . اگر قدر مطلق همبستگی کامل باشد همه نقاط بر روی خط رگرسیون قرار می گیرند . با توجه به دامنه ضریب تعیین مقدار آن را بصورت درصد

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

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

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

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

$$E(Y / X = x^*) = \beta_0 + \beta_1 x^* \quad (12-4)$$

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

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

متناسب با رابطه بالا فاصله اطمینان به صورت زیر می باشد:

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

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

	<i>Fit</i>	<i>SE Fit</i>	<i>95% CI</i>	<i>95% PI</i>
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 رابطه متغیرها پرداخته شد معمولاً "باید محقق به این نتیجه برسد که سری ها قابلیت تحلیل رگرسیونی دارند. محقق باید به این نتیجه برسد که آیا رابطه ای خطی بین متغیرها وجود دارد و آیا توزیع متغیر وابسته بهنجار است و یا واریانس آن برای مقادیر مختلف متغیر مستقل ثابت است. بر این اساس بخش 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 محقق 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 محقق 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.

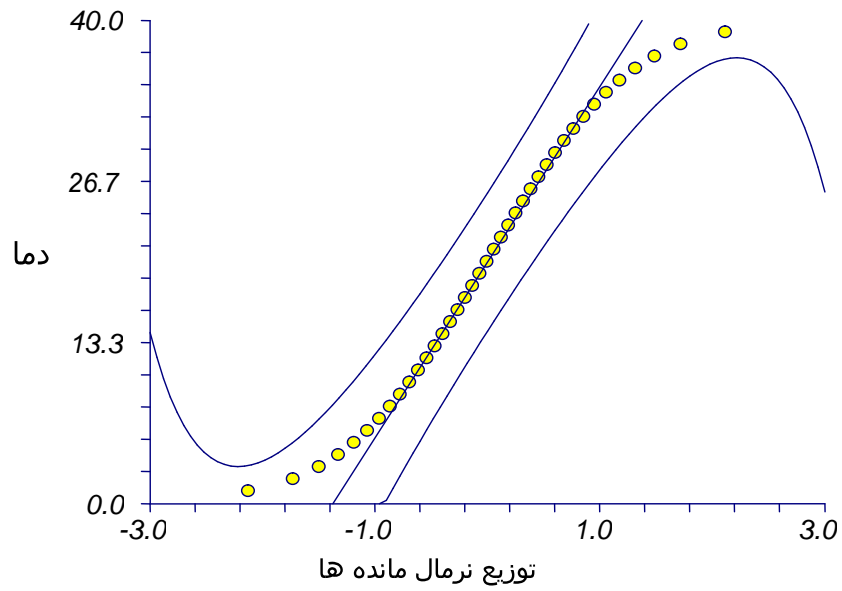
**3-** هنگام نمایش مانده ها در مقابل مقادیر پیش بینی شده نباید الگویی مشاهده شود.

**4-** مانده ها تقریباً "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 :

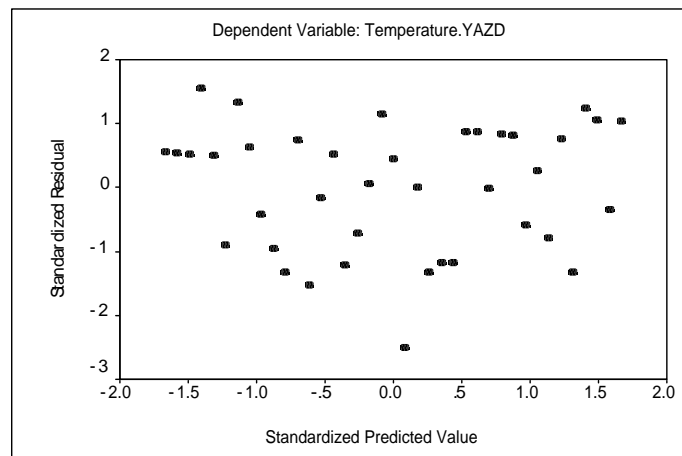
نمودار شماره 4-1 : توزیع نرمال سری دمای یزد



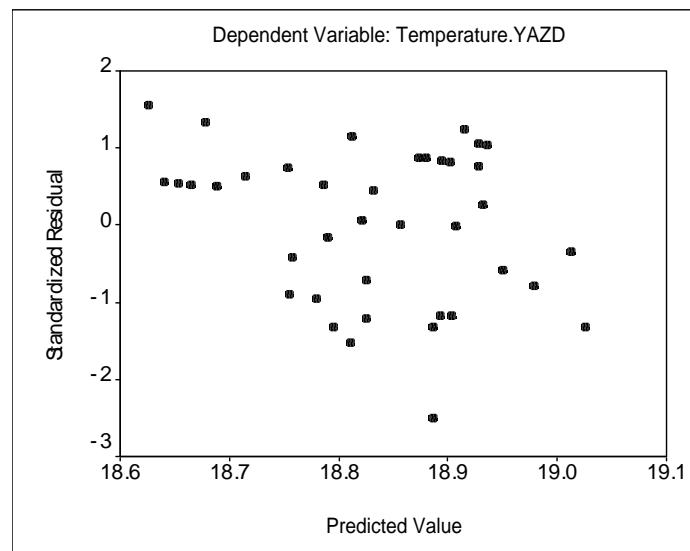
با توجه به نمودار بالا ، سری مانده های دمای یزد وضعیت بهنجاری را نشان می دهد .

- 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 شماره 4-2 : نمودار مانده های استاندارد شده دمای یزد :



نمودار شماره 4-3: نمودار مانده های استاندارد شده با مقادیر پیش بینی شده دمای یزد:



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

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



برای بررسی استقلال مانده ها از همدیگر می توان از آزمون دوربین \_ واتسون<sup>1</sup> استفاده کرد ، اگر مقدار بدست آمده بین  $1/5$  تا  $2/5$  باشد می توان تقریباً به استقلال مانده ها از همدیگر پی برد. با توجه به نتیجه کاربرد آزمون دوربین \_ واتسون می توان به مستقل بودن مانده ها پی برد. جدول شماره 5 نتیجه آزمون فوق را نشان می دهد.

جدول شماره 4- 5: آزمون دوربین \_ 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

#### 6-4- 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 رابطه 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<sup>2</sup> نامیده می شوند. ضرایب رگرسیون جزئی در نمونه برآوردی از ضرایب نامعلوم جامعه است. همانند رگرسیون خطی ساده ، در رگرسیون چندگانه باید به فرض های زیر باید مورد توجه قرار گیرد:

- 1- رابطه 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

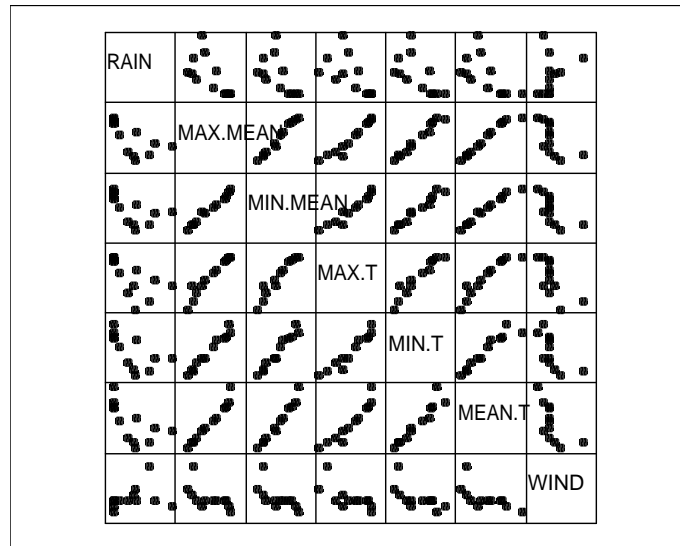
<sup>1</sup> - Durbin\_Watson

<sup>2</sup> - 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 رابطه متغیرها استفاده شده است. برای کاربرد مدل رگرسیون چندگانه می توان از معادله زیر استفاده کرد:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (18-4)$$

جدول شماره 4-6: ماتریس نمودارهای پراکنش متغیرهای ایجاد بارش خرم آباد:



با توجه به جدول ماتریس نمودارهای پراکنش متغیرها، رابطه خطی بین بعضی متغیرها وجود دارد. برای تبیین بهتر لازم است براساس معادله شماره (4-18) به تحلیل رگرسیون چندگانه پرداخت.

$$y = 203.73 - 16.76 x_1 + 37.38 x_2 - 0.46 x_3 - 4.74 x_4 - 7.726 x_5 + 7.70 x_6$$

4-7- بررسی مناسب بودن مدل :

وقتی یک مدل رگرسیون اجرا می‌شود 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 می‌دهند.

جدول شماره 4-7: خلاصه آماره مدل رگرسیون چندگانه:

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

جدول شماره 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 شماره 4-9: ضرایب رگرسیون چندگانه 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	-.4369	-2.694	.043
	MIN.MEAN	37.378	9.211	.6785	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 خلاصه آماره های تحلیل، ضریب تعیین تحلیل نشان می دهد که 87 درصد از پراکندگی مشاهده شده ایجاد بارش خرم آباد توسط شش متغیر مستقل توجیه می شود و به طریقی متغیرهای مستقل رابطه 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 شماره 8 آماره  $F = 5.59$  ( $F = \frac{MST}{MSE} = \frac{3111.775}{556.796}$ ) 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 رابطه خطی بین دو متغیر مستقل و وابسته وجود دارد. تفسیر این ضریب در رگرسیون چندگانه پیچیده تر است. اگر ضرایب رگرسیون جزئی 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 شماره 10 ضرایب رگرسیونی جزئی متغیرهای مؤثر در ایجاد بارش خرم آباد را نشان می دهد.

جدول شماره 4-10: ضرایب همبستگی جزئی متغیرهای ایجاد بارش خرم آباد:

		RAIN	MAX.MEAN	MIN.MEAN	MAX.T	MIN.T	MEAN.T	WIND	HUMIDITY
RAIN	Pearson Correlation	1.000							
	Sig. (1-tailed)								
MAX.MEAN	Pearson Correlation	-.643	1.000						
	Sig. (1-tailed)	.012							
MIN.MEAN	Pearson Correlation	-.554	.989	1.000					
	Sig. (1-tailed)	.031	.000						
MAX.T	Pearson Correlation	-.606	.965	.962	1.000				
	Sig. (1-tailed)	.018	.000	.000					
MIN.T	Pearson Correlation	-.610	.979	.975	.957	1.000			
	Sig. (1-tailed)	.018	.000	.000	.000				
MEAN.T	Pearson Correlation	-.601	.972	.982	.934	.939	1.000		
	Sig. (1-tailed)	.019	.000	.000	.000	.000			
WIND	Pearson Correlation	-.298	-.611	-.674	-.671	-.655	-.647	1.000	
	Sig. (1-tailed)	.173	.017	.008	.008	.010	.012		
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 رابطه خطی وجود دارد. از طرفی مشاهده می شود که متغیرهای مستقل با هم رابطه خوبی را نشان می دهند. قدرت رابطه خطی بین متغیرهای مستقل توسط شاخصی اندازه گیری می شود که تولرانس<sup>1</sup> 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<sup>2</sup> در نرم افزار مینی تب از عامل افزایش واریانس<sup>3</sup> (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

<sup>1</sup> - Tolerance

<sup>2</sup> - Multicollinearity

<sup>3</sup> - Variance Inflation Factor

هم خطی وجود دارد. با توجه به این که تولرانس یک نسبت است، مقدار آن بین صفر و یک می‌باشد. مقدار نزدیک به یک به این معنی است که در یک متغیر مستقل بخش کوچکی از پراکنش آن با سایر متغیرهای مستقل توجیه می‌کند و مقدار نزدیک به صفر نشان می‌دهد که یک متغیر تقریباً "یک ترکیب خطی از سایر متغیرهای مستقل است".

جدول شماره 4-11: آماره‌های تولرانس و عامل افزایش واریانس سری‌های مؤثر در بارش خرم‌آباد:

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

## تحلیل عاملی<sup>1</sup>:

در برخی تحقیقات به دلایل مختلف با حجم زیادی از متغیرها روبرو هستیم. برای تحلیل دقیق‌تر و رسیدن به نتایج علمی‌تر و در عین حال عملیاتی‌تر، 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 می‌کند.

<sup>1</sup> - برای اطلاع بیشتر به کتاب مقدمه‌ای بر روش‌های تحقیق در علوم انسانی با تأکید بر جغرافیا 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 هتلینگ (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:

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

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 بر گرفته و در نتیجه توصیف و 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 واحدهای 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 ميزانی 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 اندازه ای است که تعیین می کند چه مقدار واریانس در کل داده ها به وسیله یک عامل تبیین می شود. در تحلیل عاملی مقدار ویژه برابر 1 می 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 منظور دستیابی به عاملهای معنی دار و ساده است. محورهای مختصات عاملها را به دور مبدأ 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  $y$ ها 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 را تبیین 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 و اساسی تبیین کرد. از این رو، هدف های عمده تحلیل عاملی را می توان به شرح زیر بیان کرد:

1. کاهش تعداد زیادی متغیر به تعداد کمتری عامل به منظور مدل سازی. زیرا اگر تعداد متغیرها زیاد باشد، مدل سازی برای همه اندازه ها به گونه انفرادی امکان پذیر نیست. به همین دلیل، تحلیل عاملی با مدل سازی های مربوط به تکنیک معادلات ساختاری یکپارچه شده و به 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 بر روی 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- رابطه 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 آنها نیز قادر به تعیین و تشخیص مناسب بودن داده ها برای تحلیل عاملی می باشد از جمله این روشها، مقدار کفایت نمونه یا شاخص  $KMO^1$  می باشد که مقدار آن همواره بین 0 و 1 می باشد و در صورتی که این مقدار کمتر از 0.50 باشد داده ها برای تحلیل عاملی مناسب نخواهد بود. یعنی:

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

در فرمول  $r_{ij}$  ضریب همبستگی بین متغیرهای  $i$  و  $j$  (دو عنصر بردار  $X$  هستند) و  $a_{ij}$  ضریب همبستگی جزئی (شرطی یا ناقص)

بین متغیرهای  $i$  و  $j$  است. اگر همبستگی خطی و قوی بین متغیرها 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:



- 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 با  $MSA^1$  می باشد. به مانند روش قبل اندازه کفایت نمونه گیری متغیر آام 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<sup>2</sup> 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  $\chi^2$  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 یک 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 0.05 باشد داده ها با جامعه مورد معنی دار است. فرضیات این آزمون به صورت زیر می باشد:

$$H_0 : \rho = I$$

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

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  $\chi^2$  دو با درجه

$$\text{آزادی} = \frac{P(P-1)}{2} \text{ می باشد. لذا:}$$

<sup>1</sup> - Measure of Sampling Adequacy

<sup>2</sup> - Bartlett

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

$l_i$  مقادیر ویژه ماتریس همبستگی می باشد. اگر  $Q$  بزرگتر از  $\alpha^2 X$  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 متغیرها محاسبه شود که این روش یکی از تکنیک های عمومی و پر کاربرد در مطالعات می باشد که به تحلیل عاملی نوع R معروف 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 گروه بندی سلسله مراتبی برای طبقه بندی پاسخگویان یا موارد استفاده می شود.

#### - انتخاب مدل عاملی

در تحلیل عاملی مدل های 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، بطور قراردادی صفر تعیین می

گردد. مدل عاملی متمایل پیچیده تر از مدل عاملی متعامد می باشد. در واقع در این روش فرآیند تحلیلی کاملاً 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 باشد باید 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<sup>1</sup>:

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<sup>2</sup>:

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 :

<sup>1</sup> Exploratory Factor Analysis  
<sup>2</sup> 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	X1	X2	X3	X4	X5	X6
نام ایستگاه						
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 متغیرها  $P$  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:

$$X'_1 = [9.28 \quad 158 \quad 67 \quad 309 \quad 42 \quad 20]$$

$$X'_5 = [11.38 \quad 198 \quad 92 \quad 361 \quad 54 \quad 20]$$

به طور کلی هر واحد آزمایشی را به صورت یک بردار سطری با مرتبه  $1 \times P$  می توان نوشت. عناصر این بردار  $X'_r$  متغیرهای تصادفی هستند. یعنی:

$$X'_r = [x_{r1} \quad x_{r3} \quad \dots \quad x_{rj} \quad \dots \quad x_{rp}]$$

ترانژاد<sup>1</sup> یک بردار سطری به صورت بردار ستونی است و در تحلیل های چند متغیره اغلب بردار ستونی بکار برده می شود. یعنی:

<sup>1</sup> 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  $10 \times 6$  می باشد. لذا ماتریس داده را می توان به صورت زیر نوشت :

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

$X$  تا  $X_p$  به ترتیب معرف متغیرهای اندازه گیری شده در تمام واحدهای آزمایشی هستند و  $x_{rj}$  معرف اندازه متغیر  $j$ ام در واحد آزمایشی  $r$ ام است.

پارامترهای جامعه شامل بردار میانگین<sup>1</sup>، ماتریس واریانس - کوواریانس<sup>2</sup> و ماتریس همبستگی<sup>3</sup> می باشد. در توزیع بهنجاری یا نرمال بودن یک متغیر میانگین جامعه ( $\mu$ ) 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  $\mu$  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  $\sigma^2$  نشان داده می شود و به صورت  $\sigma^2 = E[(x - \mu)^2]$  تعریف می شود. درحالیکه در توزیع نرمال چند متغیره ، ماتریس واریانس - کوواریانس با نماد  $\Sigma$  مشخص شده و به این ترتیب به صورت زیر تعریف می شود:

$$\Sigma = Cov(X) = E[(X - \mu)(X - \mu)'] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

در این ماتریس واریانس - کوواریانس عناصر قطری  $\sigma_{11}$  تا  $\sigma_{pp}$  واریانس و عناصر غیر قطری کوواریانس را مشخص می کند. هر عنصر از این ماتریس را می توان به تنهایی به صورت زیر نوشت:

$$\sigma_{ii} = Var(x_i) = E(x_i - \mu_i)^2 \quad i = 1, 2, \dots, p$$

$$\sigma_{jj} = Cva(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] \quad i \neq j = 1, 2, \dots, p$$

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

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

مقادیر  $\rho_{ij}$  برای هر  $i \neq j$  در فاصله مثبت و منفی یک می باشد. اگر دو متغیر توزیع مستقل با هم داشته باشند در این صورت  $\rho_{ij} = 0$  و  $\sigma_{ij}$  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 مثلاً " $\sigma_{12} = \sigma_{21}$  یا  $\sigma_{13} = \sigma_{31}$  است.

## داده های استاندارد شده :

در بسیاری موارد متغیرهای اندازه گیری شده مقیاس های متفاوتی دارند . برای از بین 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:

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

ماتریس داده های استاندارد شده به صورت زیر می باشد:

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

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

## توصیف 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 جدا 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 عدد ثابت (وزن) ضرب می شود و سپس حاصل ضرب ها با هم جمع می شوند. در حالت کلی یک تبدیل خطی به صورت زیر نوشته می شود:

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

به عبارت دیگر بردار سطری وزن ها  $\alpha' = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_p]$  که هر عنصر آن وزن متغیر متناظر است در بردار ستونی متغیر

$$X = \begin{bmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{pr} \end{bmatrix} \text{ های تصادفی ضرب می شود.}$$

مثلاً:"

رتبه	X1	X2	X3	X4	X5	X6	نام متغیر
Y	نام ایستگاه						نام ایستگاه
نمره تبدیل شده	$\alpha_1=1$	$\alpha_2=2$	$\alpha_3=2$	$\alpha_4=3$	$\alpha_5=1$	$\alpha_6=2$	وزن
1468.28	9/28	158	67	309	42	20	1
1860.02	12/02	188	88	404	52	16	2
1827.98	10/98	184	88	395	56	16	3
1763.76	11/76	198	96	360	52	16	4
1768.38	11/38	198	92	361	54	20	5
1669.30	10/30	168	73	365	50	16	6
1518.65	11/65	189	79	293	48	22	7
1640.14	11/14	184	83	337	48	18	8
1761.50	12/5	177	82	381	52	18	9
1830.30	12/3	174	82	406	56	16	10
1710/831	11/33	363/6	166	1083/3	51	35/6	میانگین
17935/708	0/944	651/378	299/556	13350/9	18	19/378	واریانس

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:

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



## 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 تحلیل 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، 2 و 14 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 روش KMO 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 با هیچ 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 داده ها -
- 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<sup>1</sup>(منفی بودن ضریب همبستگی تفکیکی).

### روش KMO:

مقدار آن همواره بین 0 و 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. اگر 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 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 سهمیم میباشد.

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

### 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 مبنا 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 ، از جمله 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 عاملی 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 قطره های 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 شود و 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 عبارتند از :

1- وریماکس<sup>1</sup> ( چرخش متعامد بر روی ضرایب عامل ).

<sup>1</sup> - Varimax

2- کوارتیماکس<sup>1</sup> (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 آن ها بالاترین میزان از واریانس کل مشاهده شده را نشان می دهد انتخاب می کند. 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 می کند تا تعداد عامل های استخراج شده برابر با تعداد متغیر ها گردد. همبستگی هر متغیر با هر عامل بار عاملی<sup>2</sup> نامیده می شود و مقدار آن بین -1 و +1 تغییر می کند. واریانس تبیین شده توسط هر عامل برابر است با مجذور بار های عاملی آن. این واریانس مقدار ویژه<sup>3</sup> نامیده می شود. اولین مقدار ویژه همواره بیشترین بوده و از 1 بزرگتر می باشد. مقدار ویژه برای عامل های بعدی کوچکتر می باشد. 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 می شود. علی الاصول عامل هایی باید نگه داشته شود که اعتبار صوری یا نظری داشته باشد. منتها قبل از فرایند چرخش نمی توان به معنی هر عامل به خوبی پی برد، بنابراین معمولاً از ملاک های ریاضی مانند ملاک کایزر یا آزمون اسکری کتل برای نگه داشتن عامل ها استفاده می شود. بر اساس ملاک کایزر فقط عامل هایی نگه داشته می شوند که مجموع مجذور بارهای عاملی آن ها (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 می شود. 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 می شود. 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 آن ها 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 غالباً یک عامل کلی است که تمام یا اکثر متغیر ها بار عاملی بالایی روی این عامل دارد. عامل هایی بعدی معمولاً دو قطبی است و بارهای عاملی مثبت و منفی داشته و قابل تفسیر نمی باشد با چرخش ساختار عاملی روشنتر می شود.

<sup>1</sup> - Quartimax  
<sup>2</sup> - Factor Loading

<sup>3</sup> - Eigen Value

مشهورترین ملاک برای خوبی یک ساختار عاملی، ملاک مشهور ساختار ساده ترستون 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 متغیرها روی این عامل باید اساساً 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 & Bartlett's** 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 مانند **0.5** 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 از **0.5** 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 های 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

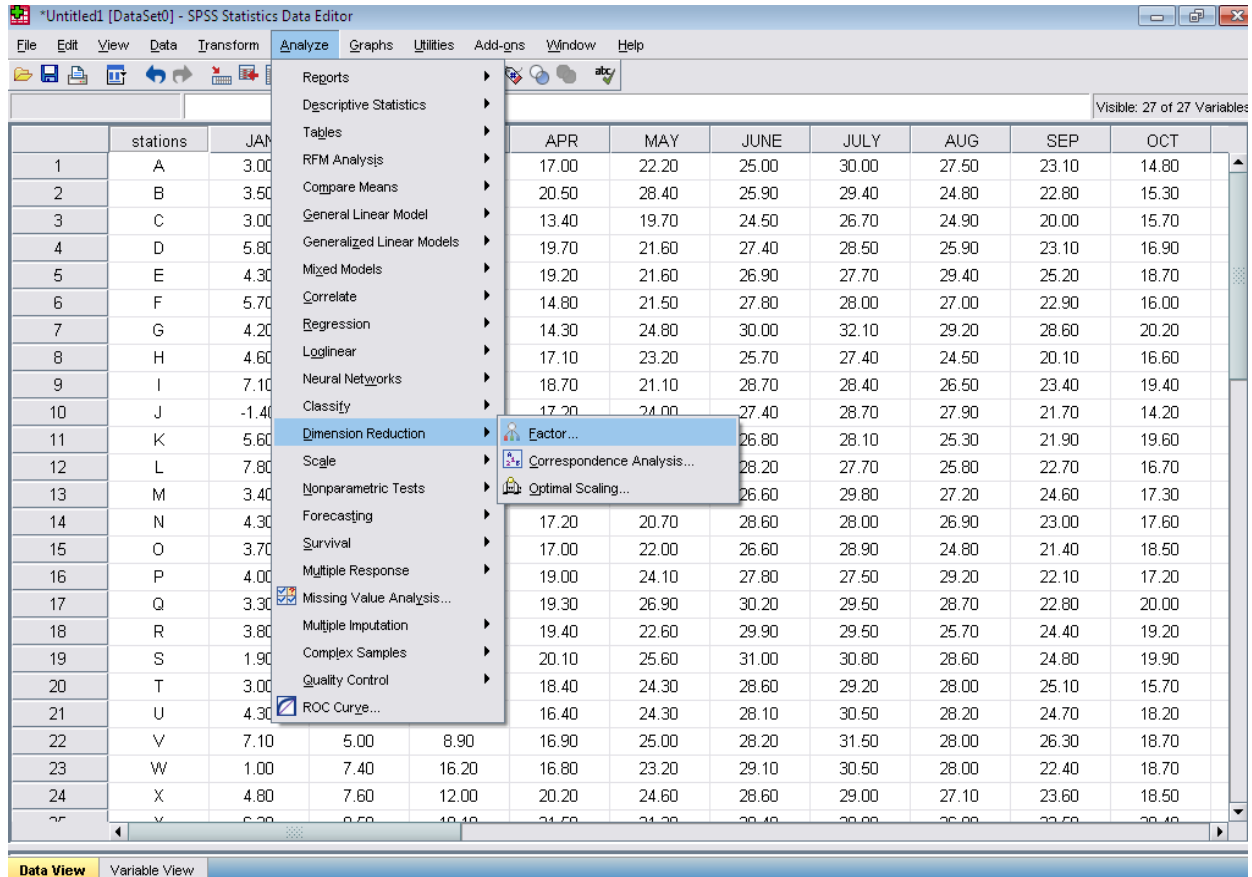
15	49.90	16.20	32.40	3.00	.00	.00	.00	.00
16	21.70	.10	1.50	.00	.10	.00	.00	.00
17	48.50	31.60	31.20	.20	.00	5.80	.00	.00
18	10.00	1.00	.40	.00	5.00	1.10	.00	19.20
19	20.20	9.20	.00	.00	.00	.00	.00	.00
20	24.10	9.70	.70	.00	.00	.00	.00	.00
21	2.40	9.70	.30	.00	.00	.00	.00	.00
22	61.20	2.80	.00	.30	.00	.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

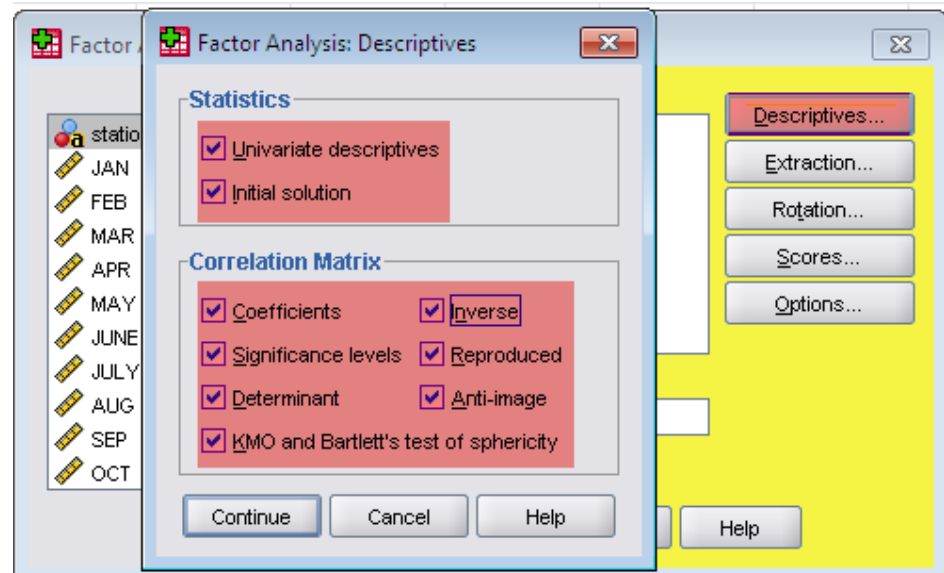
ردیف	NOV.P	DEC.P	ANNUAL.P
1	.00	10.99	143.36
2	6.06	78.35	134.66
3	.00	48.67	499.46
4	1.99	21.37	338.01
5	.00	148.85	252.43
6	6.09	85.72	309.48
7	25.03	77.21	443.90
8	72.50	43.52	244.82
9	3.70	26.40	150.30
10	3.00	78.20	236.80
11	.40	19.20	170.70
12	1.50	1.00	86.50
13	8.50	5.90	120.00
14	.10	7.40	78.70
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 **univariate 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** بر آورد های اولیه از عامل ها محاسبه می شود. در خروجی آن بر آورد های اولیه از میزان اشتراک ها ، مقادیر ویژه ماتریس همبستگی متغیرها ، درصد کل واریانس توضیح 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

ذ

*KMO and Bartlett's Test*

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



Correlation Matrix<sup>a,b</sup>

		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<sup>a,b</sup>

		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<sup>a,b</sup>

		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<sup>a,b</sup>

	DEC.P	ANNUAL.P
--	-------	----------

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

	<i>OCT.P</i>	<i>-.074</i>	<i>.038</i>
	<i>NOV.P</i>	<i>.044</i>	<i>.191</i>
	<i>DEC.P</i>	<i>1.000</i>	<i>.557</i>
	<i>ANNUAL.P</i>	<i>.557</i>	<i>1.000</i>

## Communalities

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

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 <sup>a</sup>	.418	-.369	.165	-.157	.127	.035	-.237
	FEB	.418	.769 <sup>a</sup>	.180	.198	.213	.064	.072	-.065
	MAR	-.369	.180	.809 <sup>a</sup>	.173	.210	.096	.159	.162
	APR	.165	.198	.173	.629 <sup>a</sup>	.321	.416	.185	.030
	MAY	-.157	.213	.210	.321	.678 <sup>a</sup>	.454	.425	.470
	JUNE	.127	.064	.096	.416	.454	.754 <sup>a</sup>	.630	.443

	JULY	.035	.072	.159	.185	.425	.630	.712 <sup>a</sup>	.558
	AUG	-.237	-.065	.162	.030	.470	.443	.558	.688 <sup>a</sup>
	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 <sup>b</sup>	JAN		-.083	.039	-.049	.046	-.059	-.025	.054

<i>FEB</i>	-.083		-.094	-.043	-.021	.053	.004	-.002
<i>MAR</i>	.039	-.094		.025	-.106	-.019	.005	-.018
<i>APR</i>	-.049	-.043	.025		.025	-.075	-.074	.029
<i>MAY</i>	.046	-.021	-.106	.025		-.097	-.002	-.122
<i>JUNE</i>	-.059	.053	-.019	-.075	-.097		-.029	.005
<i>JULY</i>	-.025	.004	.005	-.074	-.002	-.029		-.129
<i>AUG</i>	.054	-.002	-.018	.029	-.122	.005	-.129	
<i>SEP</i>	.029	-.026	.050	-.019	-.019	-.060	-.008	-.023
<i>OCT</i>	-.054	-.085	.074	-.029	-.076	.022	-.041	.001
<i>NOV</i>	.004	.090	-.040	.044	.030	-.010	-.066	-.011
<i>DEC</i>	.007	-.051	-.061	-.067	.043	-.056	.052	-.006
<i>ANNUAL</i>	.010	.000	.007	.031	.008	-.016	-.013	.014
<i>JAN.P</i>	-.054	.071	-.033	.060	-.007	.030	.055	-.008
<i>FEB.P</i>	-.004	.029	-.044	.078	.064	.015	-.115	-.067
<i>MAR.P</i>	.070	-.089	.000	.087	.088	-.043	-.030	.081
<i>APR.P</i>	.006	.052	-.044	-.019	.026	-.018	-.004	.032
<i>MAY.P</i>	.062	.034	-.040	-.109	.079	-.040	-.063	-.001
<i>JUNE.P</i>	.017	.001	-.051	-.054	-.017	.011	.058	-.022
<i>JULY.P</i>	-.028	.112	-.072	-.086	.036	.049	-.009	-.046
<i>AUG.P</i>	.056	.046	-.007	-.069	-.055	.037	.007	.071
<i>SEP.P</i>	.023	-.044	.003	-.066	-.023	.080	-.054	.023
<i>OCT.P</i>	.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 <sup>a</sup>	.463	.341	.213	.634	-.258	.486	-.038
	OCT	.463	.737 <sup>a</sup>	.244	.471	.711	-.240	.195	.035
	NOV	.341	.244	.849 <sup>a</sup>	.318	.422	-.291	.018	-.217
	DEC	.213	.471	.318	.810 <sup>a</sup>	.593	-.070	.221	.173

	ANNUAL	.634	.711	.422	.593	.991 <sup>a</sup>	-.504	.211	-.204
	JAN.P	-.258	-.240	-.291	-.070	-.504	.760 <sup>a</sup>	.334	.577
	FEB.P	.486	.195	.018	.221	.211	.334	.748 <sup>a</sup>	.503
	MAR.P	-.038	.035	-.217	.173	-.204	.577	.503	.728 <sup>a</sup>
	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
<i>Residual<sup>b</sup></i>	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 <sup>a</sup>	-.067	-.002	.870	-.053	-.078	-.138	-.046

	MAY.P	-.067	.782 <sup>a</sup>	.664	-.208	.298	-.093	.096	.028
	JUNE.P	-.002	.664	.852 <sup>a</sup>	-.114	-.025	-.078	.016	-.161
	JULY.P	.870	-.208	-.114	.881 <sup>a</sup>	-.108	-.045	-.040	-.020
	AUG.P	-.053	.298	-.025	-.108	.869 <sup>a</sup>	-.078	-.116	.096
	SEP.P	-.078	-.093	-.078	-.045	-.078	.857 <sup>a</sup>	.388	.668
	OCT.P	-.138	.096	.016	-.040	-.116	.388	.850 <sup>a</sup>	.609
	NOV.P	-.046	.028	-.161	-.020	.096	.668	.609	.780 <sup>a</sup>
	DEC.P	.119	.164	.106	.063	.480	.035	-.082	.058
	ANNUAL. P	.491	.158	.044	.423	.087	.131	.029	.180
<i>Residual<sup>b</sup></i>	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

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

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.

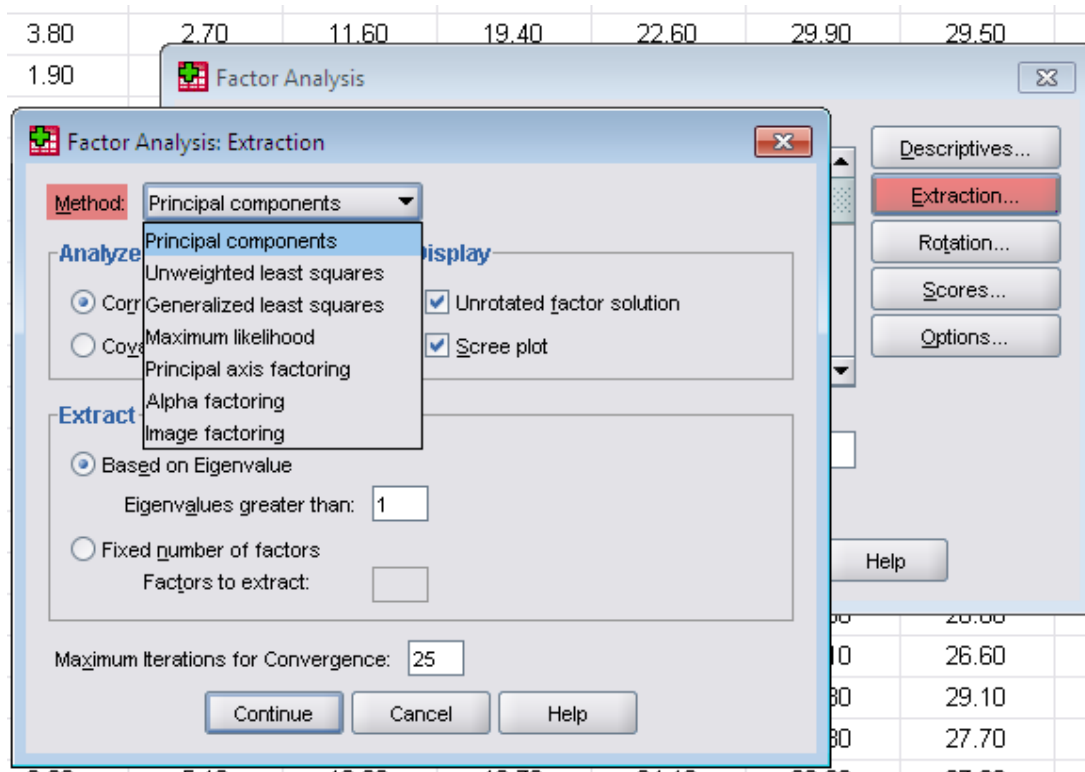
a. Reproduced communalities

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

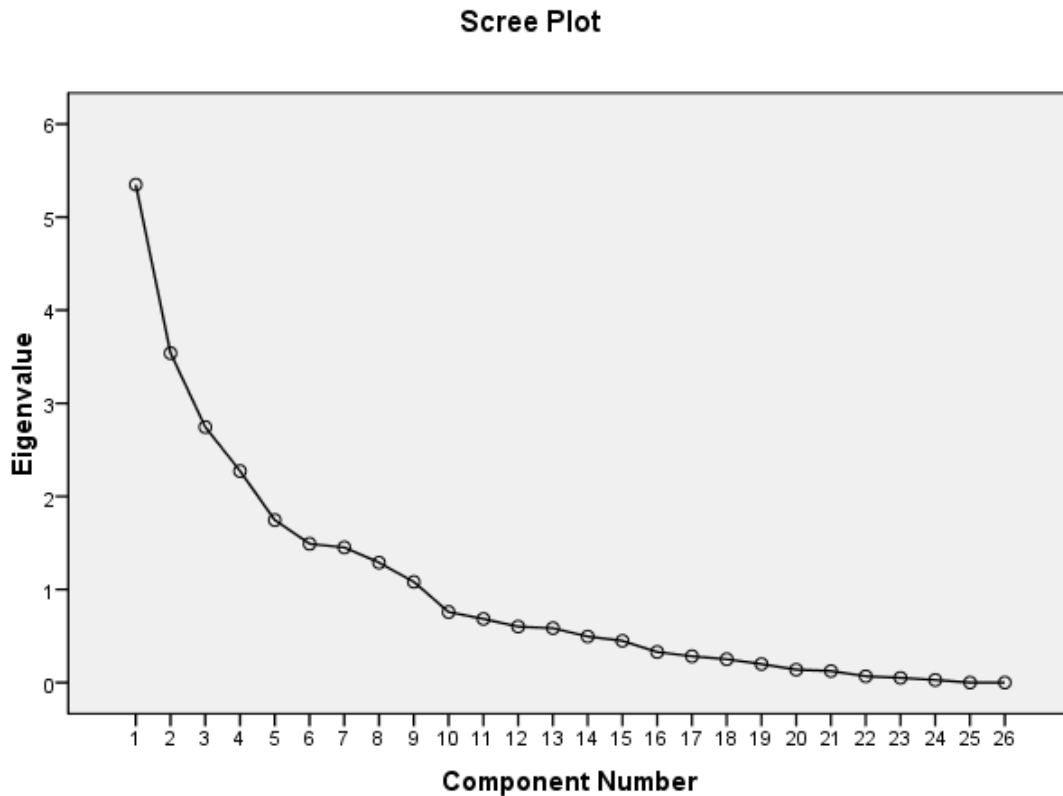
مرحله سوم: استخراج عامل ها :

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



قابل اشاره است اگر هدف محقق خلاصه کردن متغیرها و دستیابی به تعداد محدودی عامل بارش از روش مؤلفه‌های اصلی یا مبناء می‌توان استفاده کند. با انتخاب گزینه **Unrotated factor solution** ضرایب عامل‌ها قبل از دوران محاسبه می‌شوند و با انتخاب **Scree Plot** نمودار ترسیم شده مقادیر ویژه عوامل انتخابی را از بزرگ تا کوچکترین مقدار نشان می‌دهد. همچنین گزینه **Maximum Iterations for Convergence** حداکثر تعداد دفعات تکرار برای همگرایی مدل را نشان می‌دهد. یعنی:



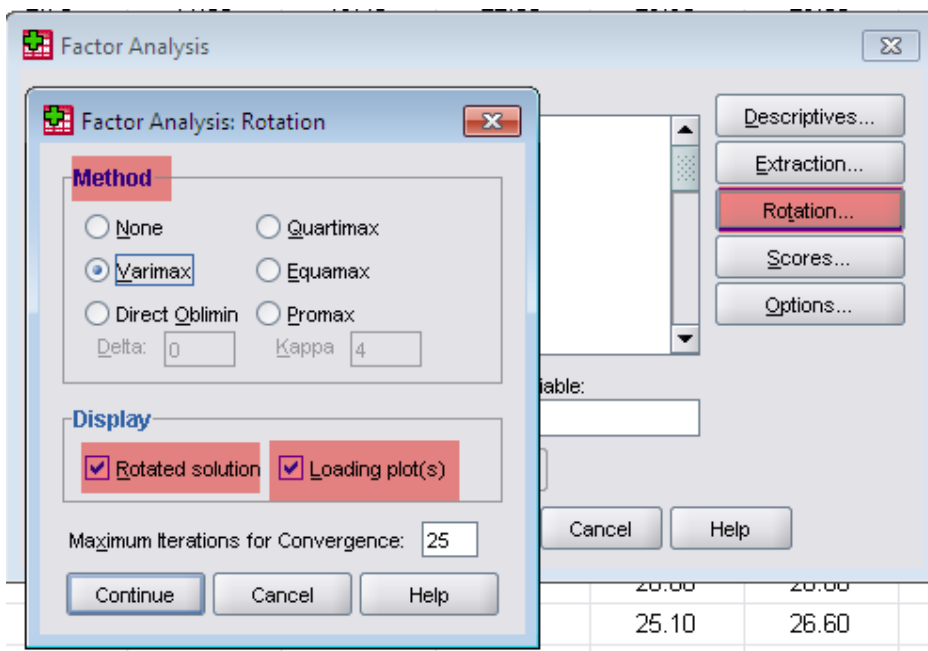
مرحله چهارم : دوران عامل ها :

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<sup>1</sup> ، **Rotated Solution** که ماتریس ضرایب عامل های دوران یافته ، ماتریس تبدیل عامل ها ، ماتریس الگو و ماتریس همبستگی بین برآوردهای دوران یافته 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 :

---

<sup>1</sup> - Direct Oblimin





خروجی قسمت دوران عامل ها:

Rotated Component Matrix<sup>a</sup>

	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.

	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

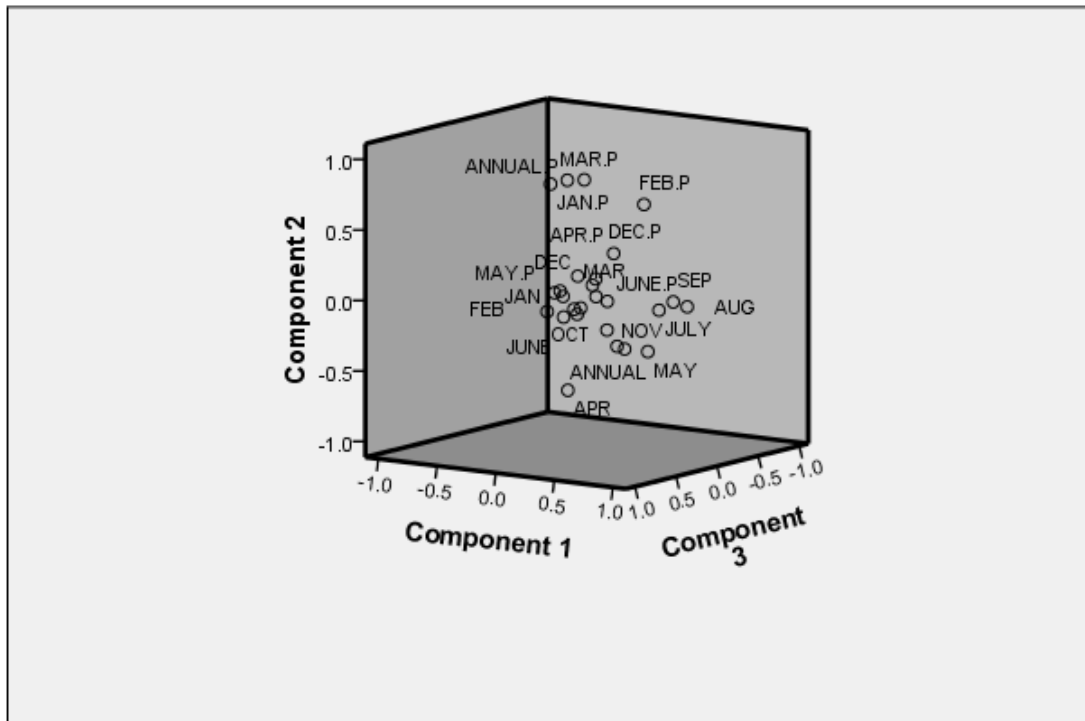
Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization.

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

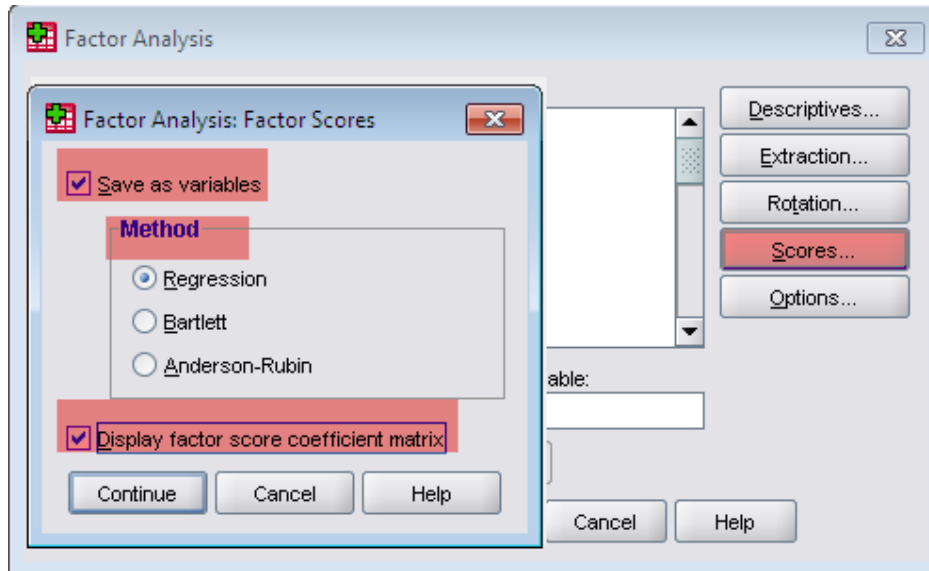
Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization.

### Component Plot in Rotated Space



مرحله پنجم: محاسبه نمره های عامل:

برای محاسبه 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 Display Factor Score Coefficient Matrix وجود دارد. روش های رگرسیون از روش کمترین مربعات برای 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



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

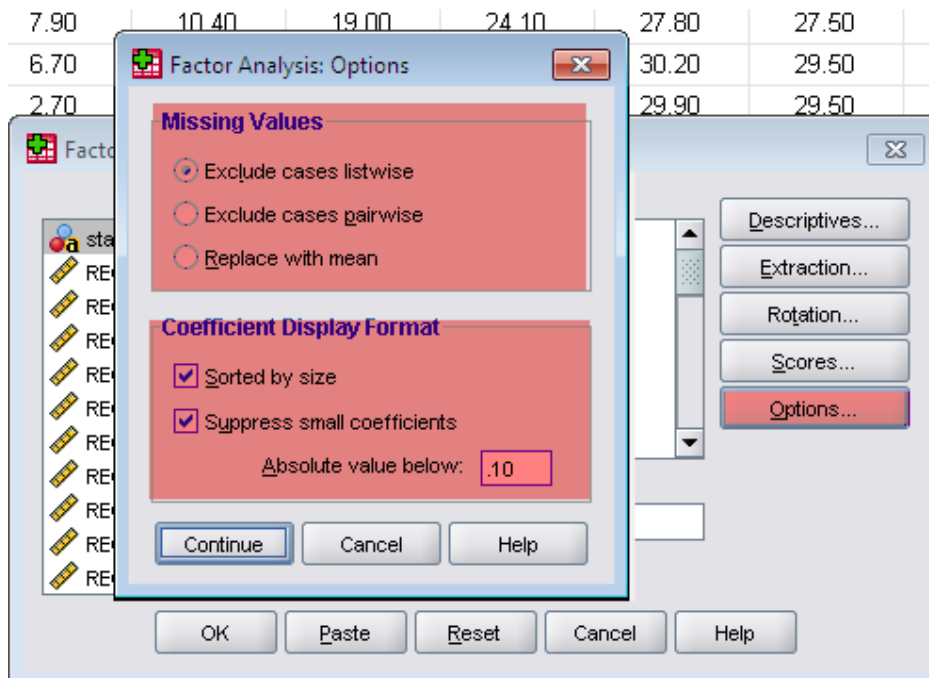
Component Score Covariance Matrix

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<sup>1</sup>

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:

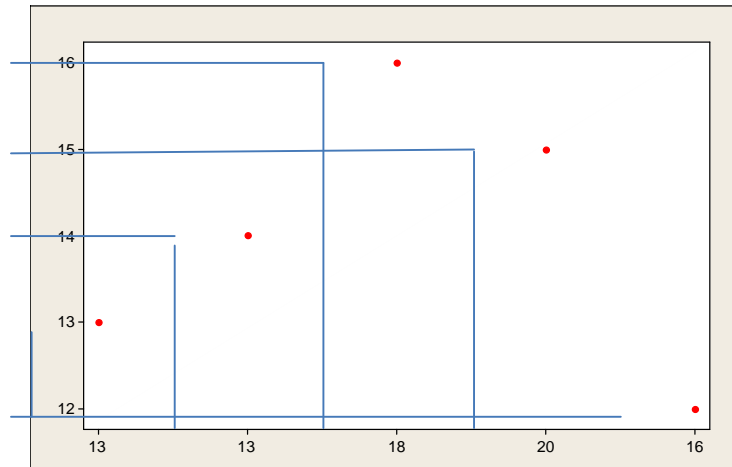
فاصله اقلیدسی:

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:

<sup>1</sup> - 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:

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

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

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

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

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

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 از بین برود). جهت بدست آوردن فاصله اقلیدسی استاندارد شده<sup>1</sup> می توان از رابطه زیر استفاده کرد:

<sup>1</sup> - Carl Pearson Distance

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

انواع تحلیل خوشه‌ای :

روش‌های متعددی برای تحلیل خوشه‌ای ارائه می‌شود که عبارتند از :

- تحلیل خوشه‌ای سلسله‌مراتبی<sup>1</sup>
- تحلیل خوشه‌ای چند میانگین<sup>2</sup>
- تحلیل خوشه‌ای دو مرحله‌ای<sup>3</sup>

در تحلیل خوشه‌ای سلسله‌مراتبی (شجره‌ای) ابتدا فاصله تکوینی هر فرد از بقیه افراد محاسبه می‌شود سپس طبق روش تجمیعی<sup>4</sup> و یا روش تقسیمی یا تجزیه‌ای<sup>5</sup> گروه‌ها تشکیل می‌شود. در روش تجمیعی همه مواد را به طور منفرد در یک دسته قرار داده می‌شود و سپس دسته‌های 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 می‌گیرد سپس این گروه به دو زیر گروه و ایندو به زیر گروه‌های دیگر تقسیم شده تا زمانی که هر ژنوتیپ در گروه خودش قرارگیرد.

تحلیل خوشه‌ای تجمیعی (ادغام) به روش‌های پیوند تکی، پیوند کامل، پیوند متوسط، وارد و روش مرکز ثقل تقسیم می‌شود. 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 میانگین‌های آنها است.

در تحلیل خوشه‌ای چند میانگینی سعی بر آن دارد تا گروه‌های همگنی از موارد را براساس خصوصیات انتخاب شده شناسایی کند. این شناسایی بر اساس الگوریتمی انجام می‌شود که تعداد قابل توجهی از موارد را شامل می‌شود. داده‌هایی که در تحلیل خوشه‌ای چند میانگینی مورد استفاده قرار می‌گیرند باید از نوع پارامتری و یا کمی باشند. چنانچه داده‌ها ناپارامتری و یا کیفی هستند باید از تحلیل سلسله‌مراتبی استفاده کرد.

در تحلیل خوشه‌ای دو مرحله‌ای برای داده‌ها و حجم زیاد داده استفاده می‌شود. در این نوع انتخاب خودکار مناسب‌ترین تعداد خوشه‌ها و همچنین بهترین شاخص برای انتخاب از بین مدل‌های مختلف خوشه بندی وجود دارد.

---

<sup>1</sup>- Hierarchical Cluster Analysis

<sup>2</sup>- K-means Cluster Analysis

<sup>3</sup>- Two-step Cluster Analysis

<sup>4</sup>- Agglomeration Cluster Analysis

<sup>5</sup>- 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 پرکاربردترین 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 تفاوت 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**<sup>1</sup> 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**<sup>2</sup> 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 در نهایت 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.

<sup>1</sup>- Forward

<sup>2</sup>- 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 :

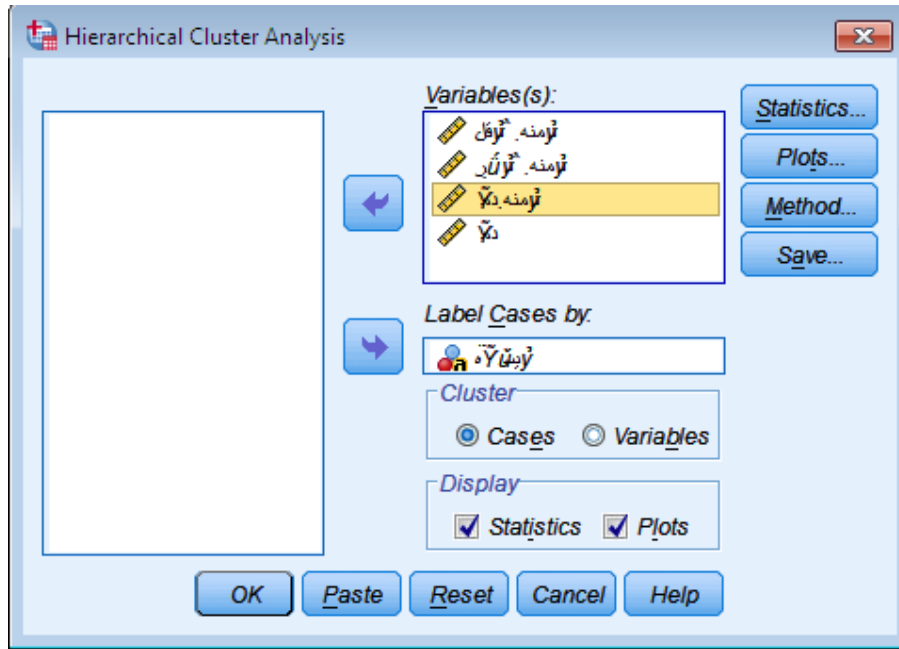
اگر به فرض 26 ایستگاه در دست باشد و هدف توجه بهینه به عناصر اقلیمی (شامل دما ، اختلاف حداکثر دما ، اختلاف حداقل دما و دامنه دما ) به صورت جدول زیر باشد . ایستگاه را راستای هدف گروه بندی نمایید:

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

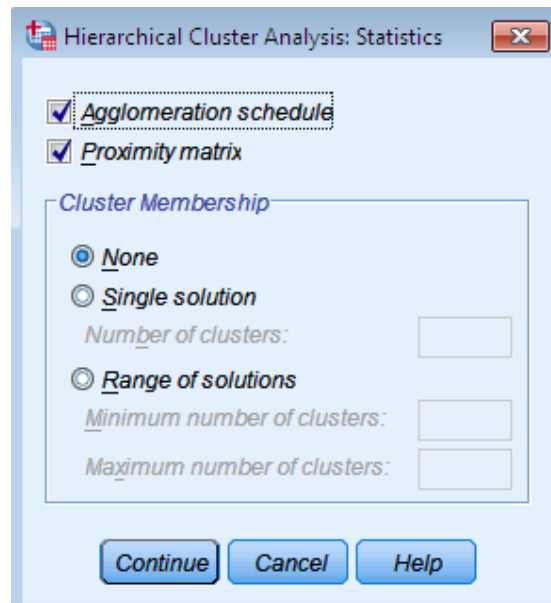
17	اردبیل	29.8	0.705	0.510	0.205
18	بوشهر	21.7	0.706	0.51	0.223
19	خرم آباد	27.6	0.680	0.499	0.207
20	همدان	26.2	0.637	0.492	0.380
21	یاسوج	33.2	0.623	0.491	0.164
22	زنجان	28.5	0.649	0.490	0.251
23	بندرعباس	27.7	0.693	0.486	0.220
24	ارومیه	30.4	0.65	0.474	0.316
25	سنندج	31.7	0.619	0.448	0.225
26	زاهدان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

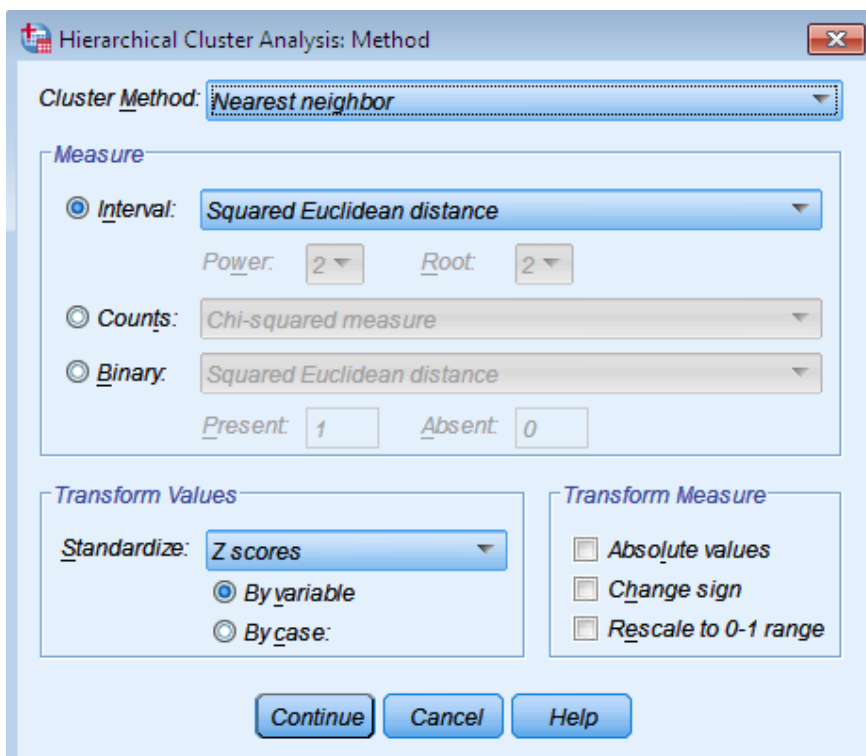
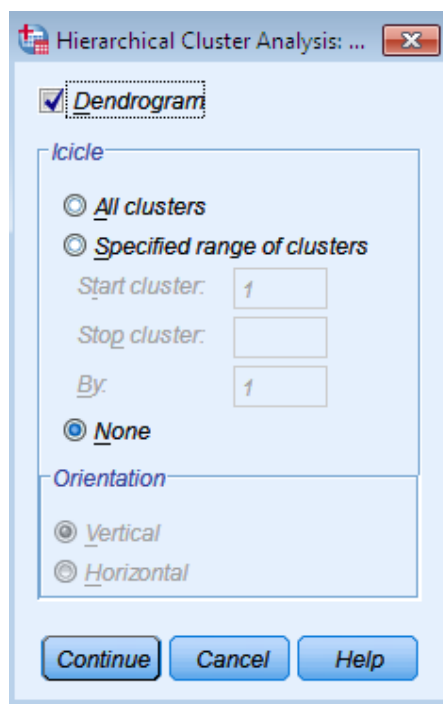
#### مراحل اجرا با نرم افزار *SPSS*:

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



سپس تنظیمات لازم به صورت زیر انجام دهید:







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. در این 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

Case	Squared Euclidean Distance							
	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

Case	Squared Euclidean Distance							
	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 یا 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 بتوان تغییرات ناگهانی و قابل توجهی را در ضرایب مشاهده نماییم. لذا قبل از مرحله 17 بهترین مرحله برای توقف خوشه بندی وجود دارد. لذا تعداد 9 خوشه از مرحله 16 تا 26 باقی می ماند که می توانیم از آنها برای خوشه بندی ایستگاه ها استفاده نمود.

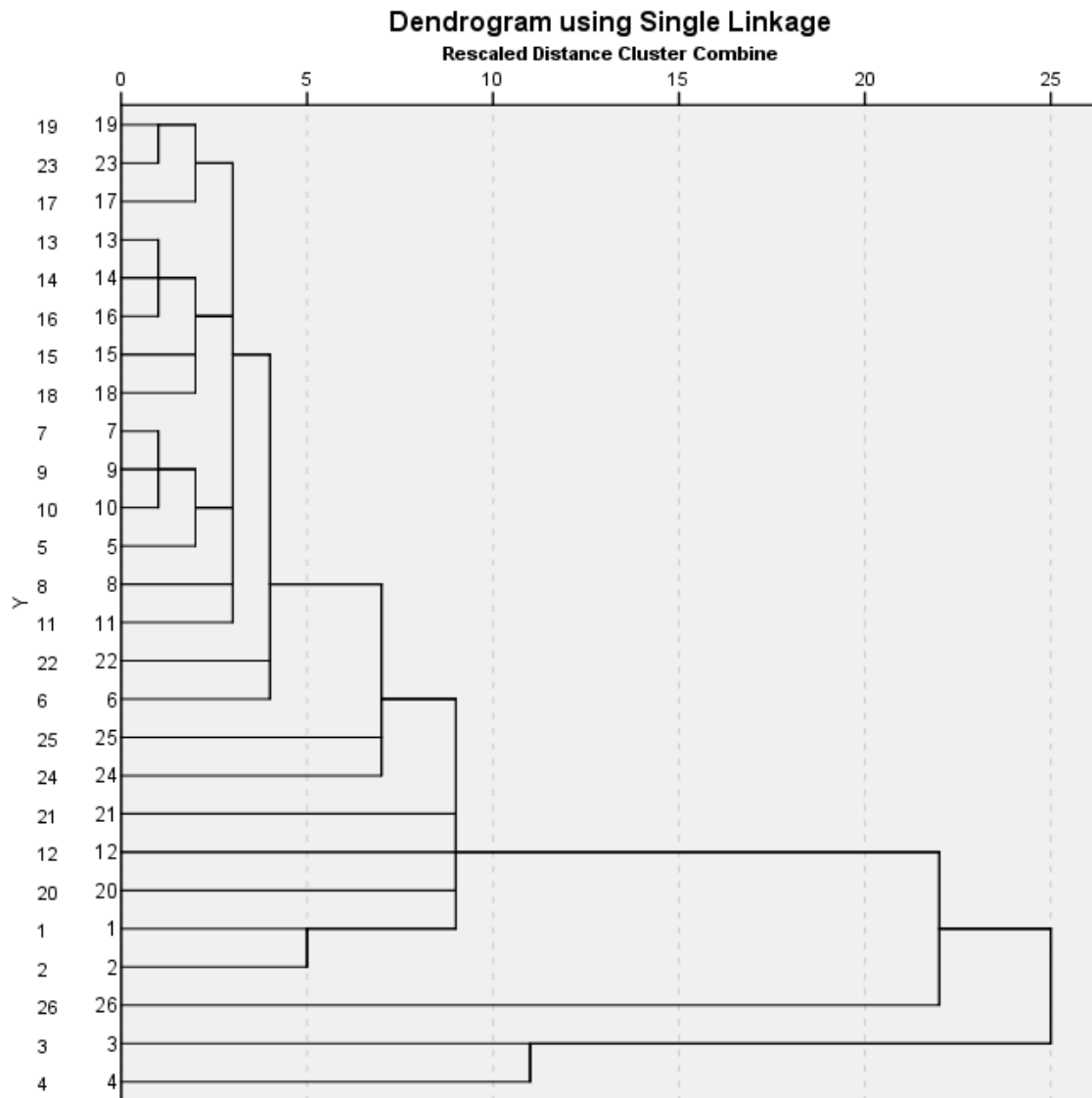
Agglomeration Schedule

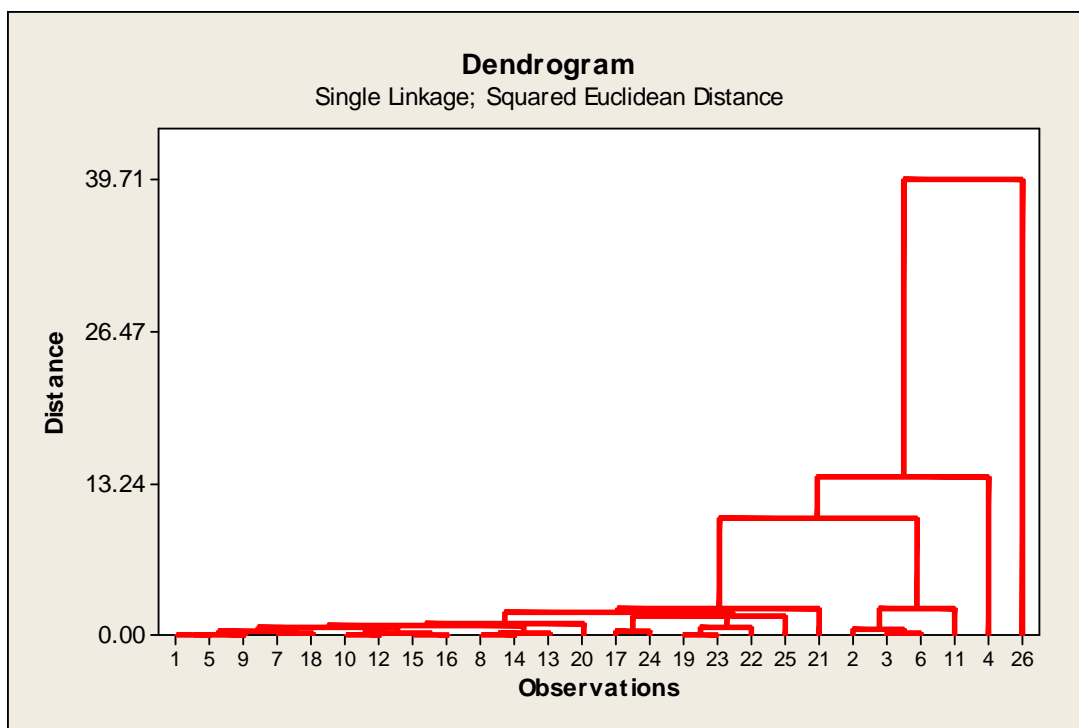
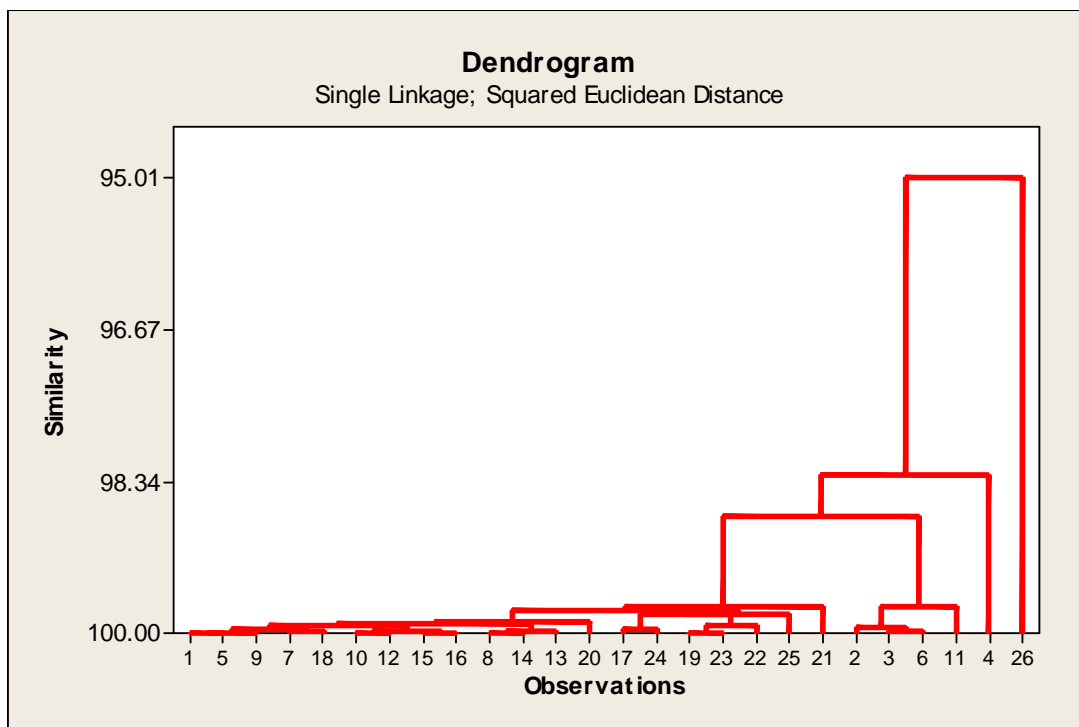
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	19	23	.130	0	0	6
2	7	9	.202	0	0	3
3	7	10	.240	2	0	9
4	13	14	.269	0	0	5
5	13	16	.269	4	0	8
6	17	19	.303	0	1	13
7	15	18	.353	0	0	8
8	13	15	.456	5	7	10
9	5	7	.458	0	3	10

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<sup>1</sup> (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 هر خوشه 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، چهار خوشه، 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:

<sup>1</sup>Dendrogram





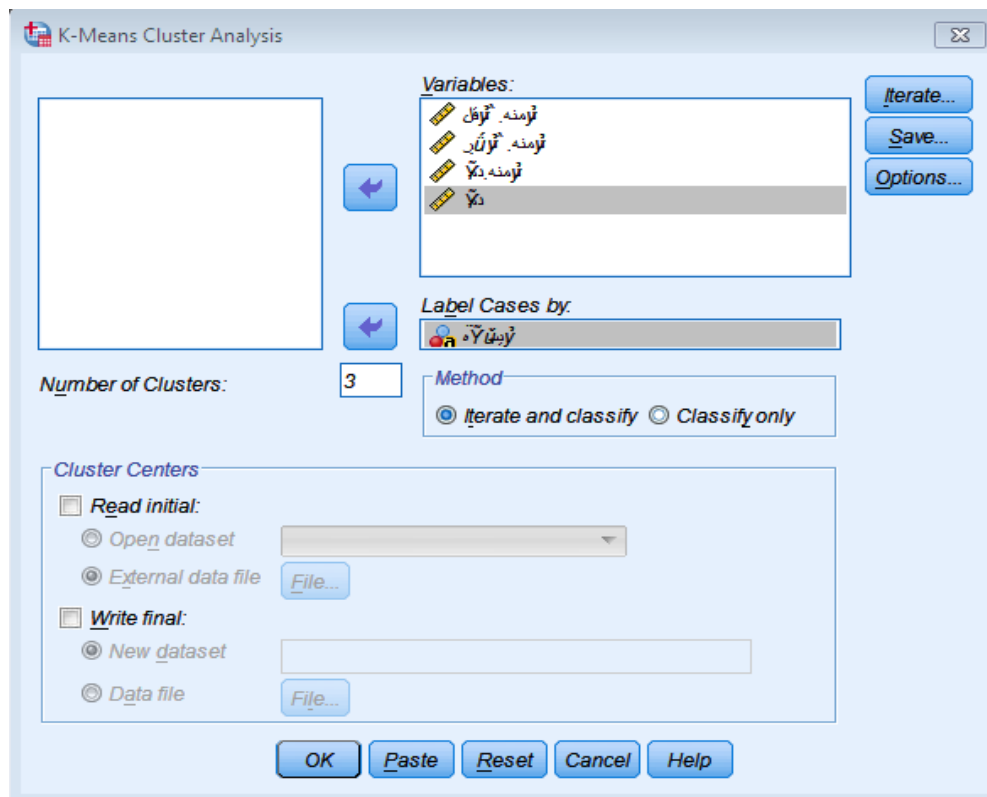
## تحلیل خوشه ای چند میانگینی :

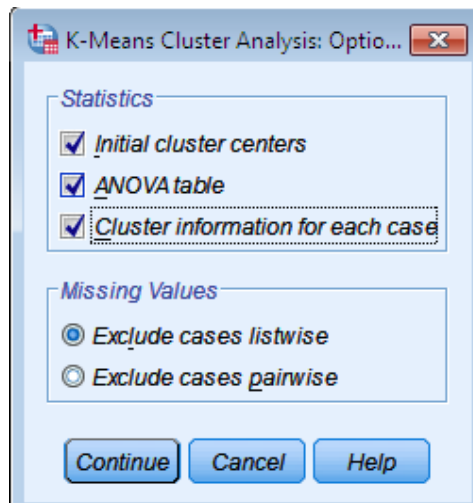
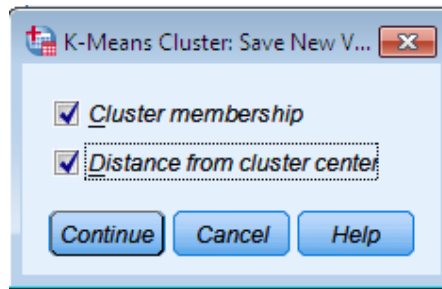
تحلیل خوشه ای چند میانگینی برای داده های بسیار بزرگ به کار می رود . در این روش فرض را بر این می گذارد که داده ها به یک تعداد خوشه مشخص و شناخته شده طبقه بندی می شوند . لذا با چنین فرضی (مشخص بودن خوشه ها) متغیرها را به چندین خوشه طبقه بندی می کند. برای طبقه بندی می توان دو روش اعمال کرد:

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 داده ها 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:







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 نسبتاً کمی تغییر می کنند.

Iteration History<sup>a</sup>

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 در خوشه 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<sup>1</sup> 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<sup>2</sup> معروف است مشاهدات را به چندین خرده خوشه<sup>3</sup> 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:

<sup>1</sup> - Scalable

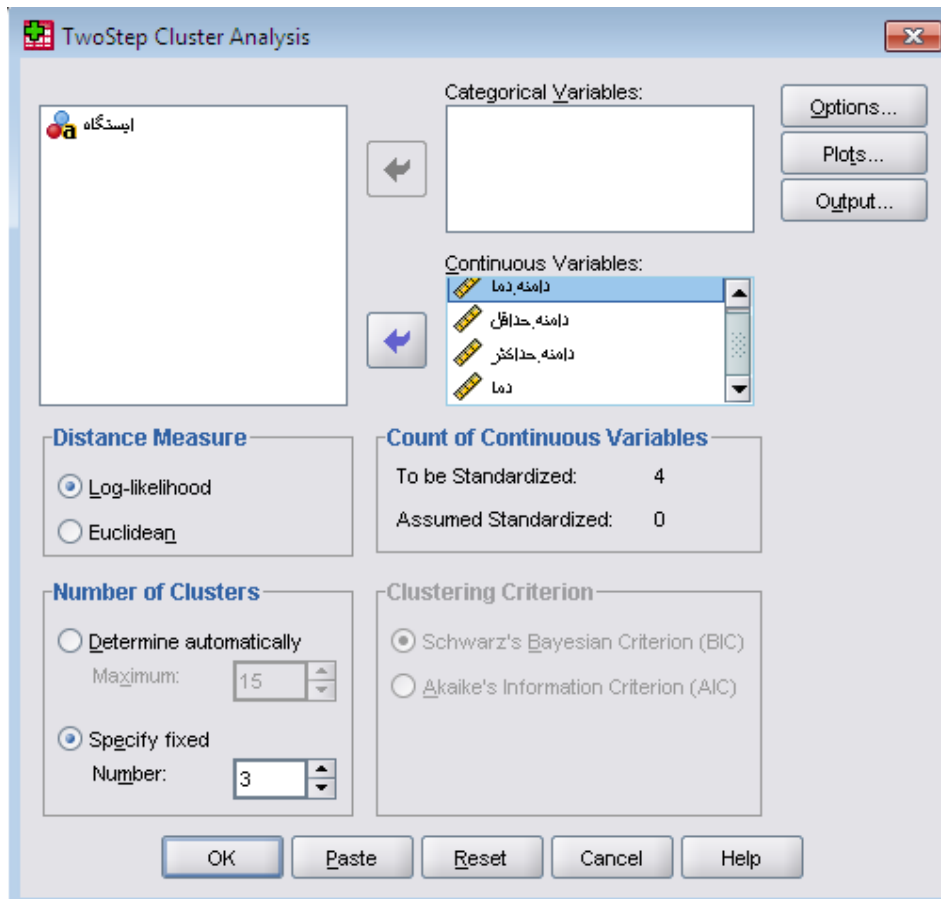
<sup>2</sup> - Pre-Cluster

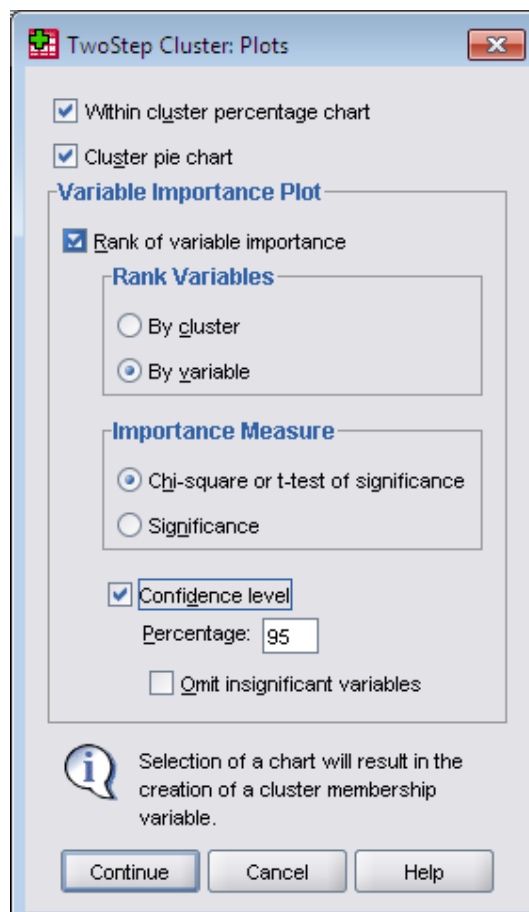
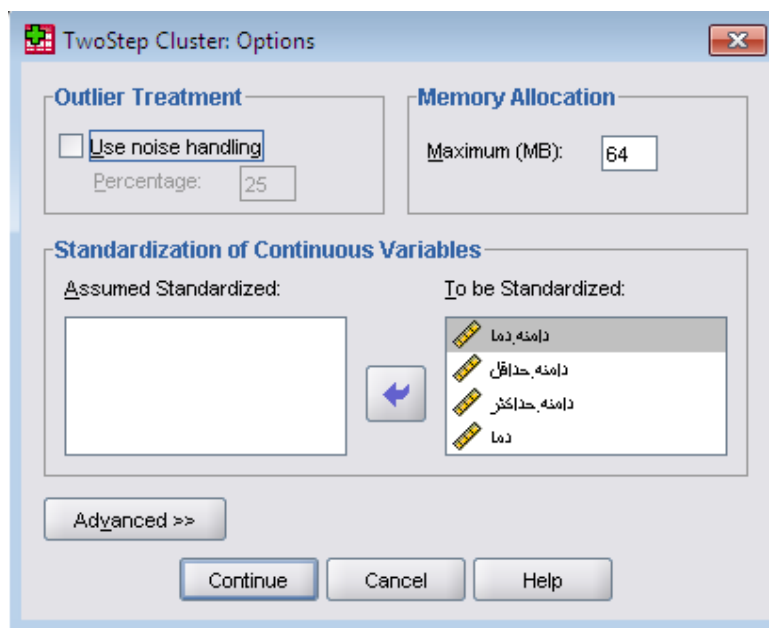
<sup>3</sup> - Sub-Cluster

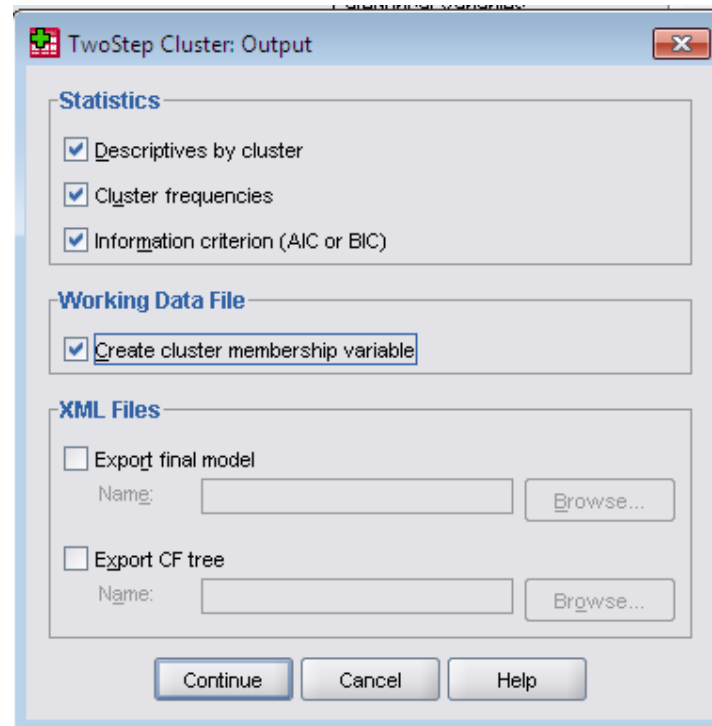
Analyze> Classify>TwoStep Cluster ابتدا دستور

فعال کنید.

مراحل زیر را به ترتیب انجام دهید:







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
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 77 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 در خوشه 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

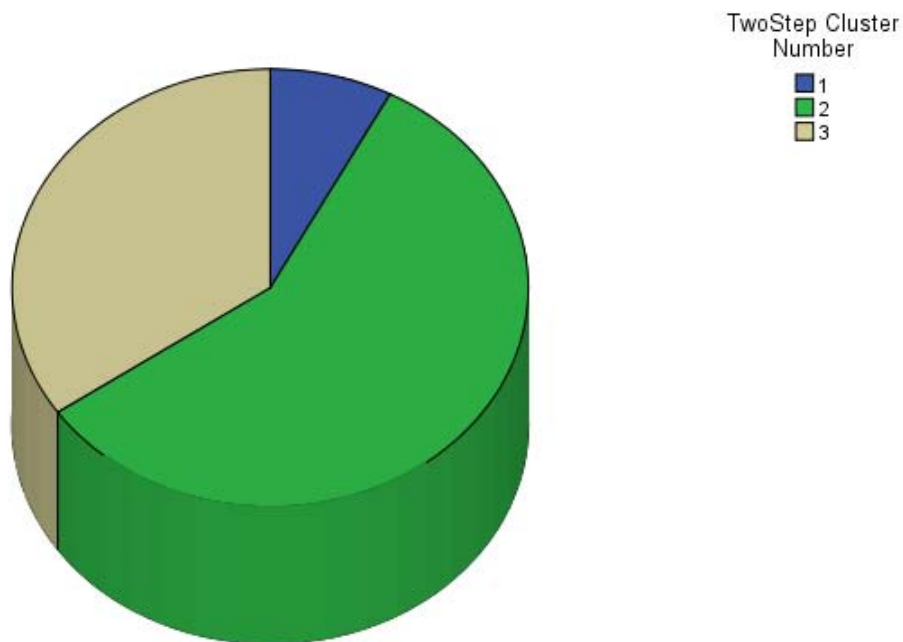
	دامنه دما		دامنه حداقل
	Mean	Std. Deviation	Mean
Cluster 1	.41350	.038891	.61200
2	.24653	.030028	.56187
3	.24311	.065693	.47589
Combined	.25819	.063557	.53596

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.

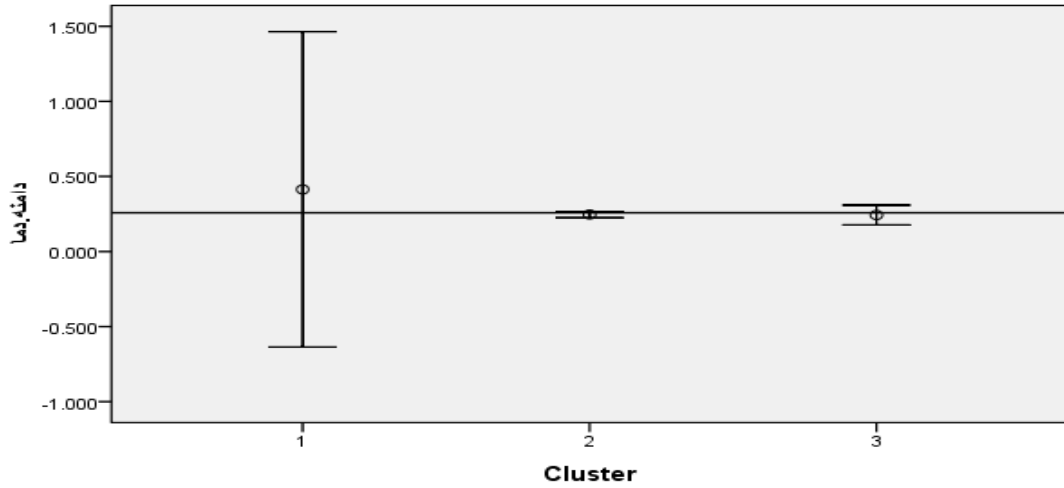
Cluster Size



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.

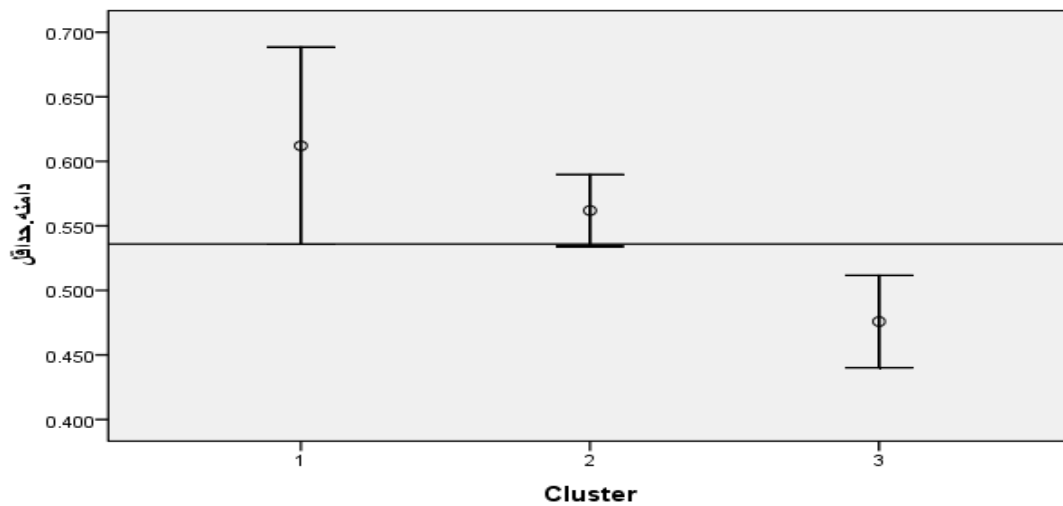
*Within Cluster Variation*

**Simultaneous 95% Confidence Intervals for Means**



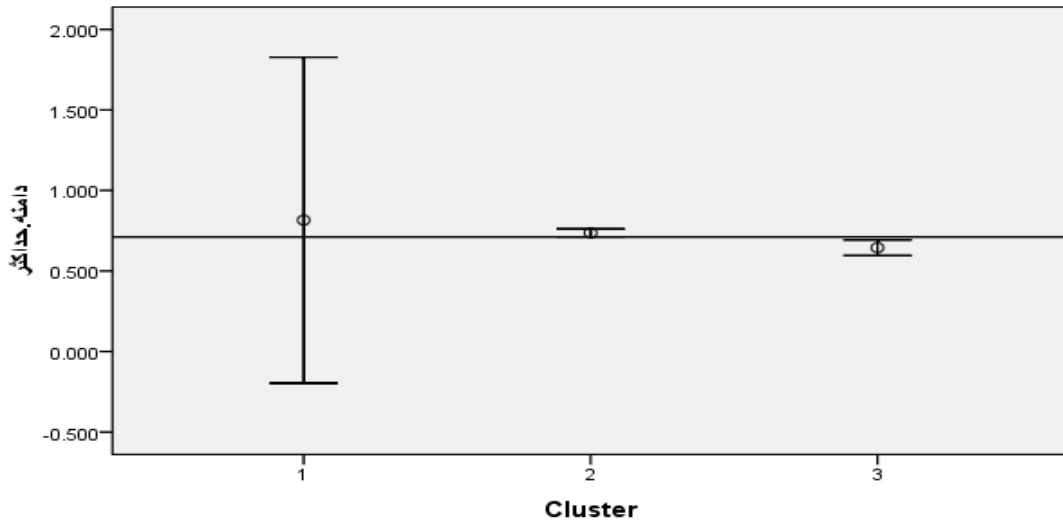
Reference Line is the Overall Mean = .258

**Simultaneous 95% Confidence Intervals for Means**



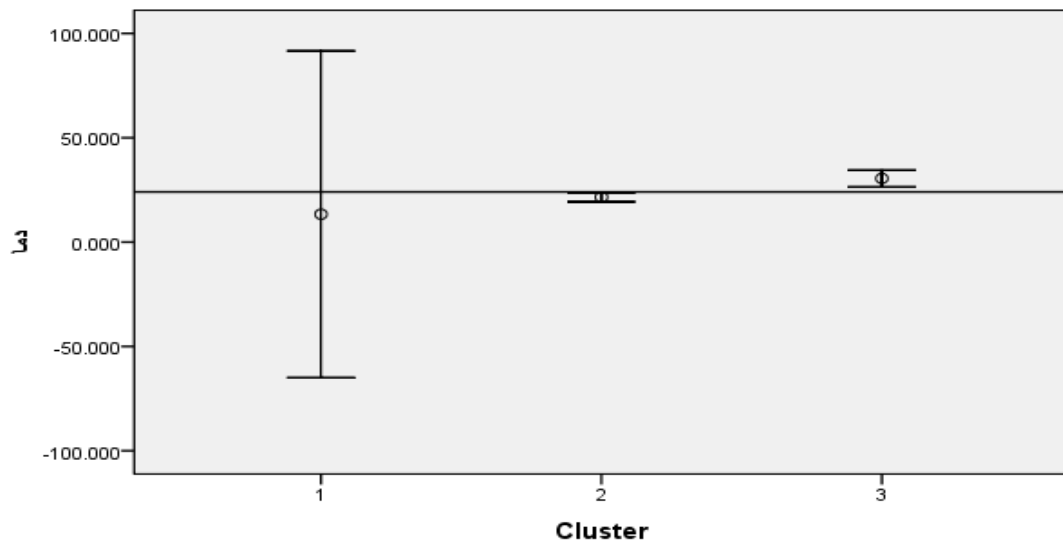
Reference Line is the Overall Mean = .536

Simultaneous 95% Confidence Intervals for Means



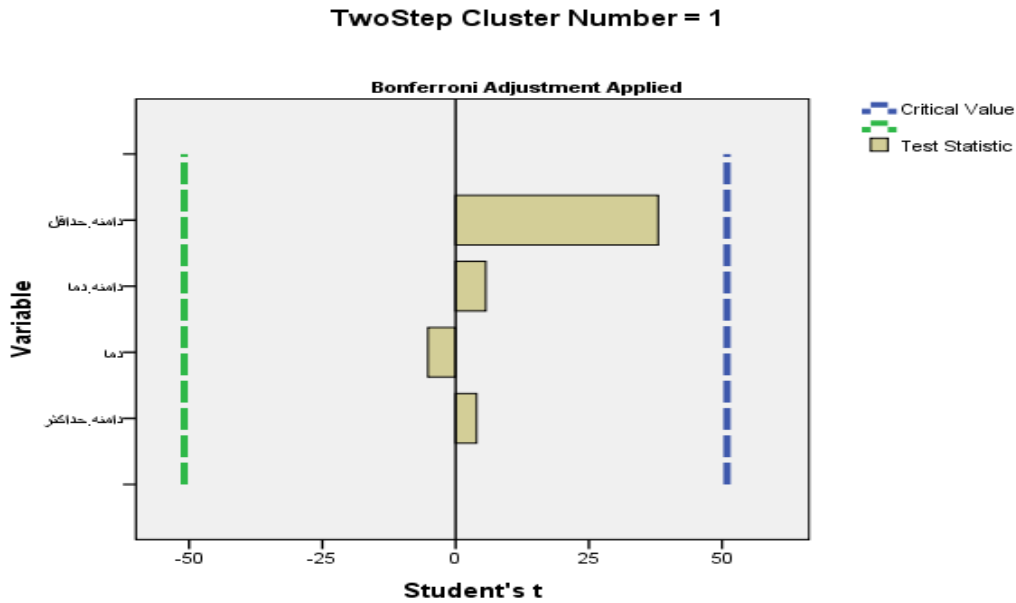
Reference Line is the Overall Mean = .711

Simultaneous 95% Confidence Intervals for Means

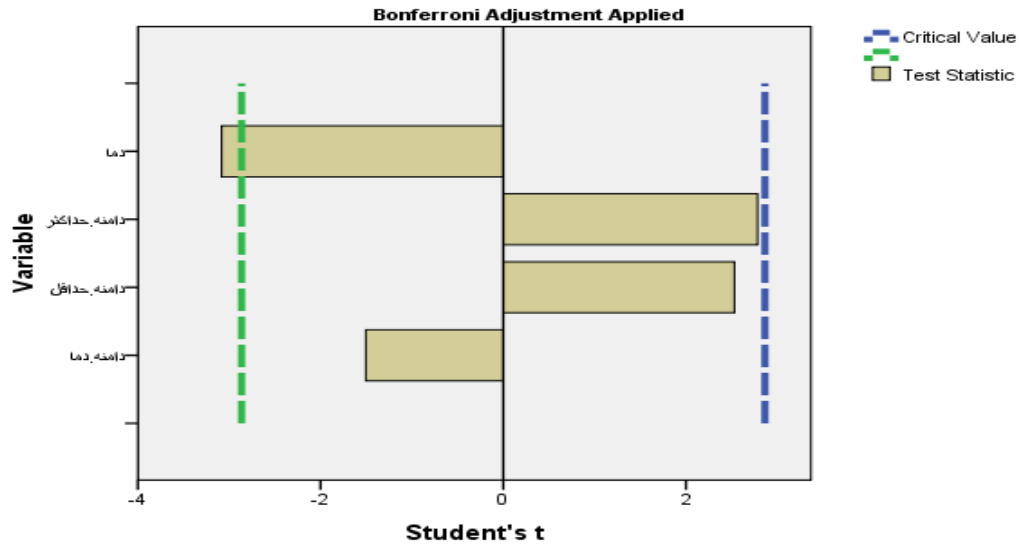


Reference Line is the Overall Mean = 24.015

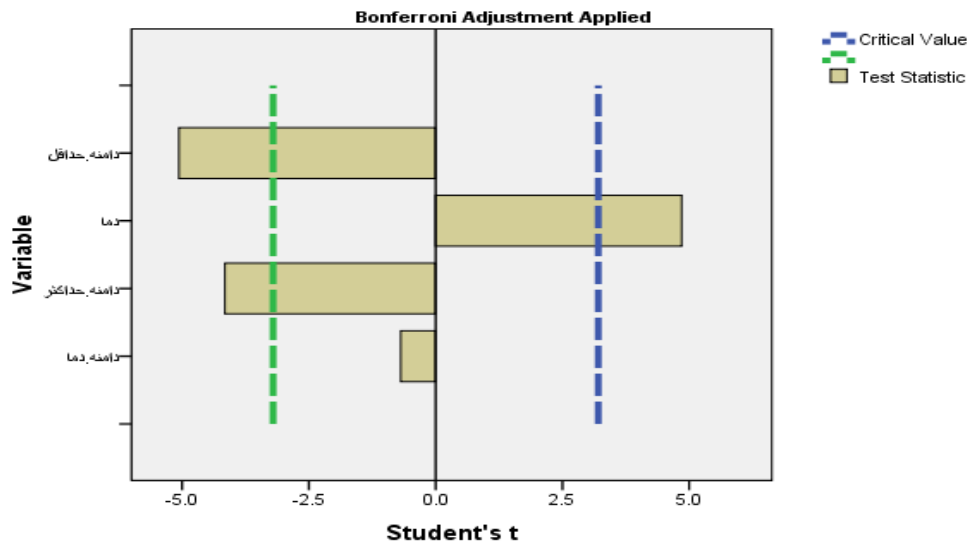
نمودارهای زیر به نمودارهای اهمیت بر حسب متغیر معروف اند. در هر نمودار متغیرها به صورت نزولی بر اساس ترتیب اهمیت شان روی محور  $Y$  نشان داده شده است. خطوط عمودی هر نمودار مقادیر بحرانی برای تعیین اهمیت هر متغیر را علامت گذاری می‌کنند. به این معنی که زمانی مقدار آماره  $t$  برای یک متغیر (چه در جهت مثبت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



TwoStep Cluster Number = 2



TwoStep Cluster Number = 3



منابع و مأخذ:

منابع فارسی :

- 1- ابراهیم پور ، ملیحه و نوروززاده رحیمی ، ماندانا: تحلیل رگرسیون همراه با Minitab<sup>14</sup>، 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 در 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<sup>15</sup> ، 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 یانرسری ، 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 بر مدل های روند) ، انتشارات پیام رسان ، 1388.
- 13- \_\_\_\_\_ : شیوه های تجزیه و تحلیل کمی در اقلیم شناسی ( با تأکید بر مدل 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 (زیر چاپ): 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 زاده مهدی : طرح و تجزیه و تحلیل آزمایشها با نرم افزار Minitab<sup>14</sup> ، 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 صنعتی اصفهان ، 1377.
- 21- صادقپور ، بهرام و مرادی وهاب : کنترل کیفیت آماری در STATISTCA<sup>R7</sup> ، انتشارات مبعث ، 1386.
- 22- \_\_\_\_\_ تحلیل پیشرفته آماری با نرم افزار STATISTCA<sup>R7</sup> ، 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- ملکیمان ، 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 متغیره چند متغیره ، 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 متغیره 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 Englon .
- 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-19<sup>th</sup> 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.