بسمه تعالی

موضوع

تحلیل سری زمانی بر روی داده های حداقل حقوق کارگران

گردآورنده

سارا معصومی

استاد راهنما

دکتر مهرداد تقی پور

دانشگاه قم – بهمن۱۴۰۰

فهرست

مقدمه

تعریف سری زمانی

دنباله‌ای از داده‌ها که در یک محدود زمانی جمع‌آوری شده‌اند، یک سری زمانی را تشکیل می‌دهند. این داده‌ها تغییراتی که پدیده در طول زمان دچار شده را منعکس می‌کنند. بنابراین می‌توانیم این مقدارها را یک بردار وابسته به زمان بدانیم. در این حالت اگر X یک بردار باشد، سری زمانی را می‌توان به صورت زیر نشان داد؛ که در آن t ، بیانگر زمان و X نیز یک متغیر تصادفی است $t=0,1,2,\ldots$

طبق این تعریف زمان $t=0$ نیز قابل تعریف است. این لحظه می‌تواند زمان تولد یک پدیده یا هنگامی باشد که اولین اطلاعات در آن لحظه ثبت شده است. به این ترتیب $X(t)$ متغیر تصادفی X را در زمان t نشان می‌دهد. مقدارهای مشاهده شده این متغیر تصادفی دارای ترتیبی هستند که زمان وقوع هر داده را نشان می‌دهند.روش‌های تحلیل سری زمانی به دو دسته تقسیم می‌شوند: روش‌های دامنه فرکانس و روش‌های دامنه زمان.

افزون بر این می‌توان روش‌های تحلیل سری زمانی را به دو دسته پارامتری و ناپارامتری تقسیم کرد .همچنین می‌توان روش‌های تحلیل سری زمانی را به دستهٔ روش‌های خطی و غیر خطی یا روش‌های تک‌متغیره و چندمتغیره تقسیم کرد. همچنین اگر تغییرات پدیده را در مدل سری زمانی برای زمان‌های منقطع در نظر بگیریم، سری را زمان-گسسته (Discreet Time) و برعکس اگر زمان را به صورت پیوسته در مدل فرض کنیم، سری را زمان-پیوسته (Continuous Time) می‌نامند. برای مثال ثبت دما، دبی رودخانه و ... از گروه سری‌های زمان-پیوسته هستند و تعداد جمعیت، تولیدات کارخانه و ... از نوع سری زمان- گسسته محسوب می‌شوند. معمولا در سری زمان-گسسته، داده‌ها در مقاطع مشخصی از زمان مثل ساعت، روز یا هفته و حتی سال جمع‌آوری می‌شوند. غالباً ایجاد مدل‌ها برای سری‌های زمان-گسسته انجام می‌شود زیرا با استفاده از گروه‌بندی و ایجاد فاصله‌های زمانی ترتیبی، امکان تبدیل سری‌های زمانی-پیوسته به زمان-گسسته وجود دارد.

اهداف تحلیل سری زمانی

۱. توصیف :با استفاده از نمودار ها و اندازه های توصیفی مانند میانگین ، واریانس ، میانه و...
۲. تشریح :تغییرات یک سری زمانی برای تشریح تغییرات سری دیگر به کار میرود.مثل فیلتر های خطی.
۳. پیش بینی :براساس اطلاعاتی که از داده های گذشته به دست می‌آید می‌توان مشاهدات آینده را پیش بینی کرد.
۴. کنترل: متغیر های ورودی را طوری تنظیم میکند تا فرایند را در هدف مورد نظر نگاه دارد.

"فصل اول"

مولفه‌های یک سری زمانی

معمولا می‌توان الگوی رفتار یا مدل تغییرات یک سری زمانی را به چهار مولفه تفکیک کرد. روند (Trend)، تناوب(Cyclic) ، فصل (Seasonal) و تغییرات نامعمول .(Irregular) اگر نمودار مربوط به داده‌های سری زمانی را برحسب زمان ترسیم کنیم می‌توانیم این مولفه‌ها را تشخیص دهیم در نتیجه شناخت بهتری از داده‌های سری زمانی خواهیم داشت .

- روند (Trend) : تمایل سری زمانی به افزایش، کاهش یا حتی ثابت بودن، روند را تشکیل می‌دهد. در یک سری زمانی با روند افزایشی، انتظار داریم مقدارهای سری زمانی در زمان‌های $t=1$ و $t=2$ به صورت $X(1) \leq X(2)$ باشند. برای مثال روند برای سری زمانی مربوط به میزان جمعیت یا سرمایه در بازار بورس به صورت افزایشی، ولی روند برای میزان مرگ و میر با توجه به پیشرفت در امور پزشکی، کاهشی است.

- فصل (Seasonal) : در سری زمانی، تغییراتی که در دوره‌ای کوتاه‌تر از یک تناوب به صورت تکراری رخ می‌دهد، به تغییرات فصلی معروف است. برای مثال در طول یک سال میزان فروش لباس‌های گرم در زمستان افزایش داشته و سپس در فصل‌های دیگر کاهش داشته است. این تناوب در سال‌های بعد نیز به همین شکل تکرار می‌شود. همانطور که مشخص است دوره تکرار تغییرات فصلی کوتاه‌تر از دوره تکرار برای تغییرات تناوبی است.

- تناوب (Cyclic) : تغییرات یکسان و تکراری در مقاطع میان‌مدت، تناوب در سری زمانی نامیده می‌شود. معمولا این تناوب ممکن است هر دو سال یا بیشتر اتفاق بیافتد. برای مثال تناوب در کسب و کار دارای یک چرخه چهار مرحله‌ای است که باعث می‌شود داده‌های مربوط به کسب و کار در یک دوره تناوب ۳ ساله تکرار شوند .

- تغییرات نامعمول (Irregular) : این گونه تغییرات بر اثر عوامل تصادفی و غیرقابل پیش‌بینی ایجاد می‌شوند. برای مثال زلزله یا سیل در بررسی رشد جمعیت ممکن است اثرات بزرگی داشته باشد. این مولفه بعد از شناسایی توسط نمودار ترسیم شده از سری زمانی باید حذف شود. در غیر اینصورت نتایج حاصل از تحلیل سری زمانی ممکن است گمراه کننده باشند.

$$X_t = M_t + S_t + Y_t$$

سری زمانی ایستا

هنگامی امکان پیش‌بینی برای سری زمانی وجود دارد، که به صورت «ایستا (Stationary) «در آمده باشد. منظور از یک سری زمانی ایستا، دنباله‌ای از مقادیر وابسته به زمان است که میانگین و واریانس آن به زمان وابسته نباشند. در حقیقت در یک سری زمانی ایستا، قوانین حاکم بر تغییرات مقدارها، وابسته به زمان نیست.

پس ابتدا باید مولفه‌هایی مانند روند یا تغییرات فصلی را از سری زمانی خارج کرد تا سری زمانی تبدیل به یک سری ایستا شده و امکان پیش‌بینی و مدل سازی بوجود آید

سری زمانی {Xt} را ایستا (Stationary) یا ایستای ضعیف (Weak Stationary) می‌نامند اگر شرایط زیر برایش برقرار باشد.

- امید ریاضی سری زمانی برحسب زمان تغییر نکند، یعنی E(Xt)=cE(X_t)=c.
- واریانس سری زمانی برحسب زمان تغییر نکند، یعنی V(Xt)=cV(X_t)=c.
- کوواریانس (همبستگی) بین X_tو X_{t-h} به زمان بستگی نداشته باشد.

زمانی که سری زمانی ایستا باشد، قادر هستیم براساس مدل‌های مرتبط با سری زمانی، رفتار فرآیند را برحسب زمان توضیح دهیم. در غیر اینصورت تا زمانی که سری زمانی ایستا نشود، امکان استفاده از مدل‌های سری زمانی معمول وجود ندارد. حال با توجه مفهوم ایستایی سری زمانی، به معرفی تابع خودهمبستگی می‌پردازیم. از آنجایی که در محاسبه تابع خودهمبستگی از کوواریانس و همچنین واریانس سری زمانی استفاده می‌شود، اگر سری ایستا نباشد، امکان محاسبه تابع خودهمبستگی وجود نخواهد داشت.

تعریف تابع خودهمبستگی

فرض کنید X_t یا $X(t)$ مقدار سری زمانی را در زمان t نشان دهد. تابع خودهمبستگی برای این سری زمانی، ضرایب همبستگی بین مشاهدات X_tو X_{t-h} را براساس $h=1,2,3,\cdots$ نشان می‌دهد. بنابراین از لحاظ محاسباتی خودهمبستگی بین X_tو X_{t-h} به صورت زیر قابل محاسبه است.

$$ACF = \mathrm{corr}(X_t, X_{t-h}) = \frac{\mathrm{cov}(X_t, X_{t-h})}{\sqrt{\mathrm{var}\ (X_t)}\ \sqrt{\mathrm{var}(\ X_{t-h})}} = \frac{\gamma(h)}{\mathrm{var}(X_t)} = \frac{\gamma(h)}{\gamma(0)}$$

به این ترتیب مخرج کسر ثابت بوده و از طرف دیگر می‌دانیم یکی از شرایط سری ایستا، ثابت بودن کوواریانس نیز هست. در این صورت انتظار داریم در یک سری زمانی ایستا، مقدار تابع خود همبستگی براساس تاخیر h تابعی از h باشد. این موضوع اساس تشکیل تابع خود همبستگی را نشان می‌دهد. بسیاری از سری‌های زمانی ایستا، دارای الگوی مشخصی برای تابع خود همبستگی هستند.

روش‌های ایستا کردن سری زمانی

از آنجایی که امکان پیش‌بینی برای سری‌های زمانی ناایستا (Non-stationary) به راحتی امکان پذیر نیست، بهتر است عواملی که باعث خارج شدن سری زمانی از حالت ایستایی هستند، حذف شوند. به این ترتیب باید مولفه‌های شناسایی شده در سری زمانی را حذف کنیم.

به این کار هموار سازی (Smoothing) یا صافی (Filtering) می‌گویند. روش‌های مختلفی برای هموارسازی سری زمانی وجود دارد از جمله :

۱. عملگرهای میانگین متحرک
۲. هموارسازی نمایی ساده
۳. برازش چند جمله ای
٤. روش‌های تفاضل‌گیری

و ... به حذف مولفه‌های سری زمانی کمک می‌کنند.

هموار سازی با یک فیلتر میانگین متحرک

در این روش برای حذف نوسانات غیر ضروری یک سری و نمایان شدن مدل اصلی که توسط این نوسانات پنهان مانده است از فیلتر ها و یا صافی هایی به نام میانگین متحرک (MA) استفاده میکنیم.تا جلوی نوسانات با فرکانس های بالا را بگیرد و تنها نوسانات ریز را عبور دهد. این فیلتر به صورت زیر معرفی میشود :

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-j}$$

به عبارت دیگر یک میانگین حول X_t به شعاع q خواهیم داشت. W_t را یک میانگین متحرک دو طرفه با دامنه q نیز می‌نامند. به‌طوری که q یک عدد صحیح غیر منفی است.هر زمان که جلو میرویم مشاهدات قدیمی حذف و مشاهدات جدید به آن اضافه میشود به همین دلیل W_t را میانگین متحرک مینامند.

هموار سازی نمایی (Exponential Smoothing)

هموار سازی نمایی نیز از همان منطق هموار سازی میانگین متحرک پیروی می کند. به این ترتیب میتوان این روش را به صورت میانگین وزنی برای داده های سری زمانی در نظر گرفت که به داده های دورتر وزن کمتری در محاسبه میانگین می دهد.

کاهش وزن مقدار های گذشته دور در حقیقت از اهمیتشان در محاسبه و پیش بینی مقادیر مربوط به آینده می کاهد و داده های مربوط به حال حاضر تاثیر بیشتری خواهند داشت.

در صورتی که سری زمانی دارای روند باشند، استفاده از روش هموار سازی نمایی مضاعف نتایج بهتری را ارائه خواهد داد. به نظر می رسد که می توان هموار سازی نمایی مضاعف را به صورت دو بار استفاده از هموار سازی نمایی ساده در نظر گرفت. با اضافه شدن فاکتور یا عامل هموار سازی فصلی، می توان برای سری های زمانی فصلی نیز مدلی ارائه داد.

فرم ریاضی این هموار سازی به صورت زیر میباشد :

$$W_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + ... + \alpha(1-\alpha)^{t-2}X_2 + (1-\alpha)^{t-1}X_1$$

$$= \sum_{j=0}^{t-2} \alpha(1-\alpha)^j X_{t-j} + (1-\alpha)^{t-1}X_1$$

بنابراین W_t میانگین وزن با وزن های زیر میباشد :

$$\alpha(1-\alpha)^j \quad j = 0,1,2,\dots,t-2 \quad , \quad (1-\alpha)^{t-1}$$

هموار سازی با روش برازش چند جمله ای

ایده اصلی برازش چند جمله ای از تقریب بسط تیلر حاصل می‌شود بدین معنی که هدف برآورد تابع بر حسب یک چند جمله ای از درجه t می‌باشد. بارسم کردن داده ها در مقابل زمان درجه این چند جمله ای را حدس میزنیم سپس با استفاده از روش کمترین مربعات خطا ، ضرایب مجهول را برآورد میکنیم.مزیت این روش بر این است که در این حالت فرم خاصی برای m_t تعیین میشود.و عیب این روش این است که درجه این چند جمله ای بستگی به نظر تحلیلگر دارد.برای حذف روند کافی است روند برآورد شده را از داده ها کم کنیم تا باقی مانده ها محاسبه شوند.

هموار سازی با تفاضل گیری

در این روش برای حذف روند ، برآورد روند را مستقیما میتوان با تفاضل گیری از داده ها حذف نمود.برای این منظور ابتدا دو عملگر زیر را معرفی میکنی.

۱.عملگر تفاضل با تاخیر یک

$$\nabla X_t = X_t - X_{t-1}$$

۲. عملگر شیفت به عقب

$$BX_t = X_{t-1}$$

با یک تفاضل گیری در تاخیر یک ، یک سری ایستا با میانگین c_1 حاصل میشود. بنابراین در مدل کلاسیک یعنی $X_t = m_t + Y_t$ که در آن m_t یکی چند جمله ای از درجه k بوده و Y_t یک سری ایستا با میانگین صفر باشد .

با استفاده از عملگر ∇^k خواهیم داشت :

$$\nabla^k X_t = \nabla^k m_t + \nabla^k Y_t = K!\ c_k + \nabla^k Y_t$$

که سری $\nabla^k Y_t$ نیز یک سری ایستا است با میانگین صفر زیرا $K!\ c_k$ به t بستگی ندارد پس سری نیز ایستا است و میانگین آن برابر $K!\ c_k$ است.

بنابراین پیشنهاد میشود که برای هر دنباله از داده ها که فقط دارای روند هستند، عملگر ∇ را مرتبا با مرتبه های مختلف تکرار میکنیم.تا اینکه دنباله حاصله یعنی $\nabla^k X_t$ به عنوان یک مصداق از یک فرایند ایستا مشاهده گردد. در عمل معمولا با درجه $k=1$ و یا $k=2$ نتیجه مطلوب حاصل میشود.

تحلیل سری زمانی

بعد از شناسایی و حذف مولفه‌های اصلی سری زمانی، وقت آن رسیده که بتوانیم عمل پیش‌بینی را انجام دهیم، یعنی مدل ریاضی برای ارتباط بین مقدارهای سری زمانی را پیدا کنیم. در اینجا به مدل های پیش‌بینی سری زمانی به نام های میانگین متحرک(Moving Average) ، اتورگرسیو(Autoregressive) و آرما (ARMA) که تلفیقی از دو مدل قبل است می‌پردازیم.

الف: مدل تصادفی محض

مدل سری زمانی برای داده های $\{X_t\}$ یک تابع چگالی احتمال توام است.

تابع $F(x_1, x_2, \dots, x_n)$ بیان کننده آرایه $(X_1, X_2, \dots, X_n)$ است. خواص مدلI.I.D به شرح زیر است:

- روندی ندارد
- عدم حضور تغییرات فصلی
- مشاهدات مستقل و از یک توزیع یکسان پیروی می کنند
- توزیع احتمال توام به صورت زیر است :

$$F(x_1, x_2, \dots, x_n) = F(x_1) \times F(x_2) \times \cdots \times F(x_n) = \prod_{i=1}^{n} F(X_i) .$$

ب: مدل اتورگرسیو مرتبه p

این مدل به صورت AR(p) بیان میشود که در آن p،مرتبه اتورگرسیو بوده و ساختار آن مطابق رابطه زیر است(.فرم باز شده) با شرط $|\varphi|<۱$

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + Z_t$$

یک مدل ایستا با نام اتورگرسیو مرتبه p است که در آن Z_t ها فرایند نویز سفید هستند با میانگین صفر و واریانس σ^2 و φ پارامتر مدل است.این مدل یک فرایند کازال است زیرا توانستیم Y_t هارا بر اساس مقادیر حال و گذشته بنویسیم.

ج) مدل میانیگین متحرک از مرتبه q

X_t را یک فرایند میانگین متحرک از مرتبه q مینامند هرگاه X_t را به توان به فرم زیر نوشت :

$$X_t = \sum_{j=0}^{q} \theta_j Z_{t-j}$$

که در آن $\theta_1, \theta_2, \dots, \theta_q$ ثابت های حقیقی هستند و $\theta_0 = 1$ است و Z_t ها دارای توزیع نویز سفید هستند با میانگین ۰ و واریانس σ^2 . و به صورت $MA(q)$ مینویسند.

مدل $MA(q)$ ایستای ضعیف و q-وابسته است و در صورتی که Z_t ها مستقل و هم توزیع باشند ایستای قوی خواهد بود.در عمل اگر تابع خودهمبستگی نمونه ای از سری را رسم کنیم و ببینیم که از مرتبه q به بعد $\hat{\rho}(h)$ تقریبا صفر است میتوان مدل $MA(q)$ را پیشنهاد داد.

د: مدل اتورگرسیو میانگین متحرک[۲] (ARMA)

این مدل تلفیقی از مدل اتورگرسیو و میانگین متحرک است. که به صورت $ARMA(p,q)$ بیان میشود که در آن p و q به ترتیب، مرتبه اتورگرسیو و مرتبه میانگین متحرک بوده و ساختار آن مطابق رابطه زیر است .

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} - Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \cdots - \theta q Z_{t-q}$$

Y_t متغیر پیشبینی شده در زمان t ام و φ پارامتر مدل AR و θ پارامتر مدل MA میباشد. و p , q به ترتیب مرتبه اتورگرسیو و میانگین متحرک است. t زمان و Z_t ها نویز سفید هستند با میانگین صفر و واریانس σ^2.

Xt را یک فرایند ARMA(p,q) با میانگین μ مینامند اگر Xt-μ یک فرایند ARMA(p,q) باشد.مدل معرفی شده با شرط $|\varphi|<۱$ یک فرایند ایستا ، کازال و معکوس‌پذیر میباشد.

روش تشخیص سری تصادفی محض (I.I.D)

۱. تابع خود همبستگی نمونه ای :

زمانی که نمودار تابع خودهمبستگی نمونه ای را رسم میکنیم لازم است ۹۵٪ از داده ها در بازه ی $\pm\frac{1.96}{\sqrt{n}}$ قرار گیرند.آنگاه میتوانیم فرض I.I.D بودن را برای داده ها در نظر بگیریم.

روش تشخیص مدل اتورگرسیو AR(p)

میتوانیم نمودار تابع خودهمبستی را رسم کنیم اگر داده ها به صورت نمایی کاهش پیدا کرده باشند میتوانیم مدل اتورگرسیو را برایشان پیشنهاد بدهیم.همچنین برای اطمینان از پیشنهاد خود میتوانیم نمودار پراکنش را برای Yt ها در مقابل Yt-1 رسم کنیم اگر رابطه خطی بین آنها مشاهده کنیم یعنی مدل اتورگرسیو یک مدل مناسب برای باقی مانده ها است.

روش تشخیص مدل میانگین متحرک MA(q)

در این روش نیز میتوانیم با رسم نمودار خود همبستگی و توجه به تابع خودهمبستگی مدل میانگین مترح مدل را تشخیص دهیم به این طریق که اگر در q مقادیر غیر صفر حضور داشته باشند دارد و بعد از تاخیر q مقدار داده ها برابر صفر باشد.

"فصل دوم "

چکیده :

در این بخش میخواهیم داده های مربوط به حداقل حقوق کارگران از سال ۱۳۴۷ تا ۱۳۹۶ را که به طور سالانه جمع آوری شده اند را با استفاده از روش های مربوط به سری زمانی تحلیل کنیم چرا که داده های ما وابسته به زمان هستند.

این داده ها ماهیت پیوسته دارند و به دلیل این که طی ۳۹ سال به طور سالانه یعنی در زمانهای مقطعی جمع آوری شده اند بنابراین زمان-گسسته هستند.

نکته۱: به علت بزرگی داده ها از آنها ln گرفته ایم و با داده های ln گرفته شده تحلیل هارا پیش برده ایم.

نکته ۲ : منبع داده های مورد استفاده در این پروژه سایت مرکز آمار ایران میباشد.

نکته ۳ : برای انجام تحلیل سری زمانی از نرم افزار SPSS استفاده میکنیم.

تحلیل سری زمانی را در سه مرحله پیش برده ایم.مرحله اول بررسی حضور روند و عامل فصلی است مرحله دوم حذف روند است و مرحله سوم ارائه مدل مناسب برای باقی مانده های سری میباشد.

در مرحله اول به کمک رسم نمودار سری زمانی دیدیم که یک روند خطی صعودی در داده ها وجود دارد. در مرحله دوم تلاش کردیم تا این روند را از سری حذف کنیم.این کار را به دو شیوه مختلف انجام دادیم.ابتدا با استفاده از روش تفاضل گیری روند را حذف نمودیم و سپس با استفاده از روش برازش چند جمله ای به حذف روند پرداختیم.اما نتایجی که از این دو روش به دست آمد یکسان نبود چرا که در روش تفاضل گیری با مشاهده ی نمودار تابع خودهمبستگی برای باقی مانده ها نتیجه گرفتیم که باقی مانده ها سری تصادفی محض هستند و میانگین و واریانس را برایشان برآورد کردیم.اما در روش برازش چند جمله ای زمانی که تابع خودهمبستگی را برای باقی مانده ها رسم نمودیم دیدیم که باقی مانده ها سری تصادفی محض نیستند.بنابراین لازم شد تا مدل مناسبی را برای با قی مانده ها پیدا کنیم.با توجه به شکل نمودار تابع خودهمبستگی برای باقی مانده ها که به صورت نمایی نزول کرده بود مدل اتو رگرسیو(1) را پیشنهاد دادیم و با رسم نمودار پراکنش برای مقادیر باقی مانده ها در زمان حال (t)در مقابل باقی مانده ها در زمان t-1 و مشاهده ی وجود

رابطه ی خطی بین آنها از پیشنهاد خود مطمئن شدیم.اما در کنار آن مدل اتورگرسیو(2) را نیز ارزیابی نمودیم.

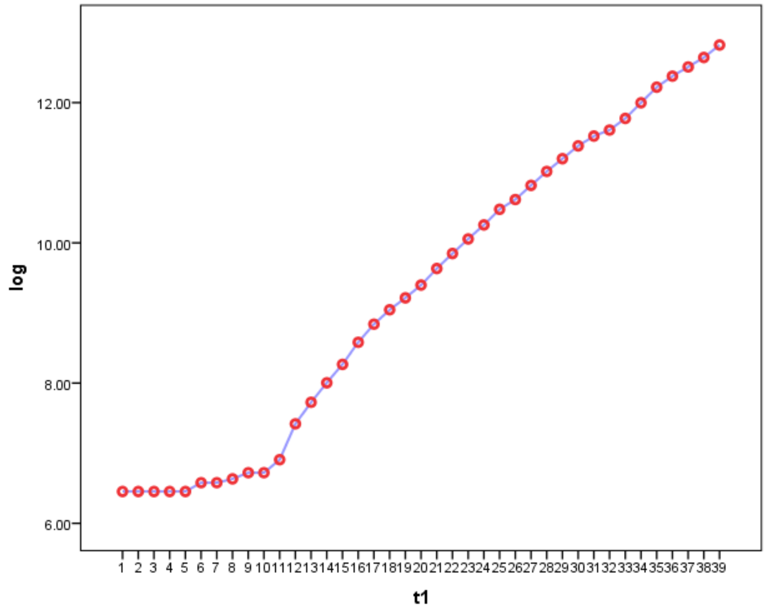
برای هر کدام از این دو مدل ضرایب φ را محاسبه نمودیم و سپس به بررسی سری Zt ها پرداختیم تا مطمئن شویم که در هر دو مدل Zt ها I.I.D باشند. همچنین مقدار واریانس را برای هر دو مدل اتورگرسیو ۱ و ۲ به دست آوردیم.هر دو مدل برای سری باقی مانده ها مناسب بودند اما ما ملاک برتری را بر کمتر بودن مقدار واریانس قرار دادیم و چون واریانس حاصل از مدل اتورگرسیو(2) کمتر از واریانس حاصل از مدل اتورگرسیو(1) بود بنابراین مدل اتورگرسیو(2) مدل مناسب تری برای سری باقی مانده ها بود.

مدل میانگین متحرک را بررسی نکردیم چرا که با توجه به نمودار تابع خودهمبستگی مدل میانگین متحرک مدل مناسبی برای باقی مانده های ما نبود.

بنابراین در مدل ARMA(p,q) مقدار q=0 است و زمانی که ARMA(p,0) داشته باشیم معادل این است که AR(p) را داریم بنابراین تنها به بررسی مدل اتورگرسیو پرداختیم.

در انتها نیز در قسمت پیوست روش ها و مسیر های تحلیل و محاسبات همچنین داده های اصلی بدون اعمال Ln را قرار داده ایم.

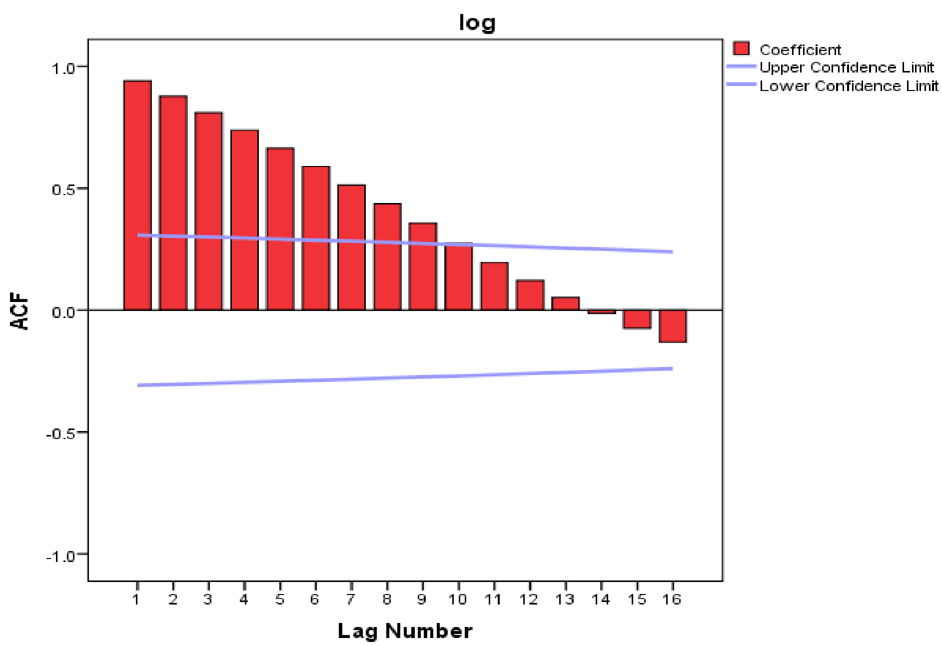
مرحله اول : رسم نمودار سری زمانی با هدف تشخیص روند و یا مولفه فصلی



تفسیر: همانطور که در نمودار بالا مشاهده میکنیم.مولفه فصلی در وجود ندارد اما داده ها دارای روند صعودی میباشند و این یعمی سری ما یک سری ایستا نیست و نیاز است که به حذف روند بپردازیم.

بنابراین مدل ما به صورت $X_t = m_t + Y_t$ میباشد که در آن m_t روند و Y_t مقدار تصادفی مستقل و هم توزیع با میانگین صفر میباشد.

تابع خود همبستگی X_t ها :

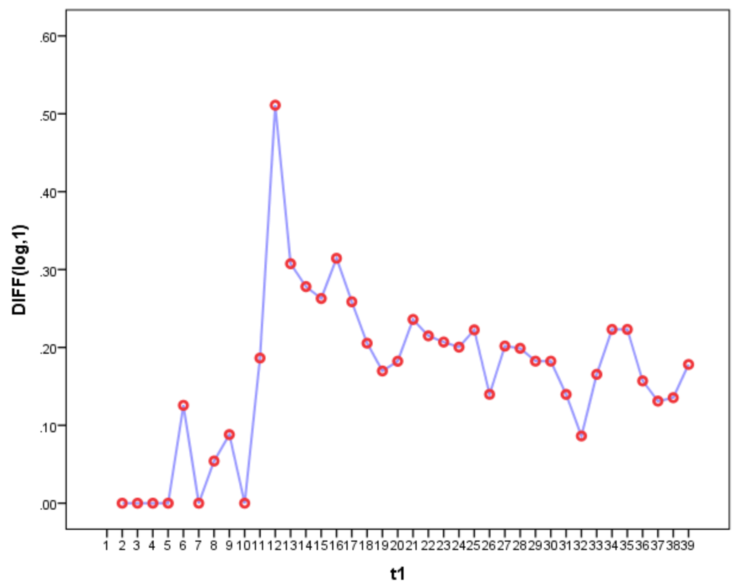


مرحله دوم : برآورد و حذف روند

در اینجا میتوانیم از روش های معرفی شده در فصل اول برای حذف روند از سری داده ها استفاده کنیم.با استفاده از ۳ روش تفاضل گیری ، هموارسازی نمایی و برازش چند جمله ای به حذف روند میپردازیم.

روش اول : تفاضل گیری

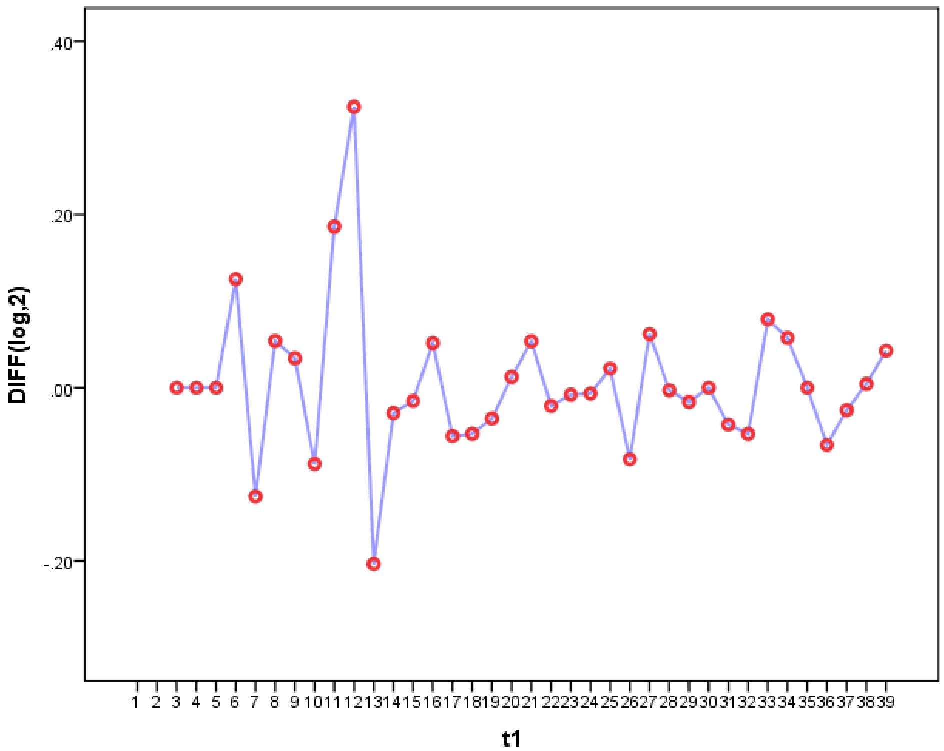
تفاضل گیری مرتبه ۱ را در نرم افزار SPSS انجام دادیم و نمودار سری زمانی باقی مانده های حاصل از آن را در زیر میبینیم.



تفسیر: همانطور که در نمودار فوق مشاهده میکنیم همچنان به طورمیانگین یک روند صعودی در داده ها مشاهده میشود و روند به طور کامل حذف نشده است پس نیاز است یک بار دیگر عمل تفاضل گیری را روی داده ها اعمال کنیم.

تفاضل گیری مرتبه ۲ :

در اینجا نمودار سری زمانی برای باقی مانده های حاصل از روش تفاضل گیری مرتبه ۲ را میبینیم:



تفسیر: همانطور که در نمودار فوق میبینیم روند داده ها به طور کامل حذف شده است و سری ما تبدیل به یک سری ایستا شد😊

روش دوم : برازش چند جمله ای

با توجه به نمودار سری زمانی برای داده ها (در صفحه ۱۴ مرحله ی اول) میتوانیم یک رابطه خطی درجه ۱ را برای روند موجود در سری پیشنهاد بدهیم. بنا براین معاله ی m_t (روند) به صورت زیر میباشد.

$X_t = m_t + Y_t \quad , \quad m_t = a_\circ + a_1 t$

حال نیاز است تا ضرایب $a_\circ$, a_1 را با استفاده از روش کمترین مربعات خطا محاسبه کنیم.

Coefficients[a]

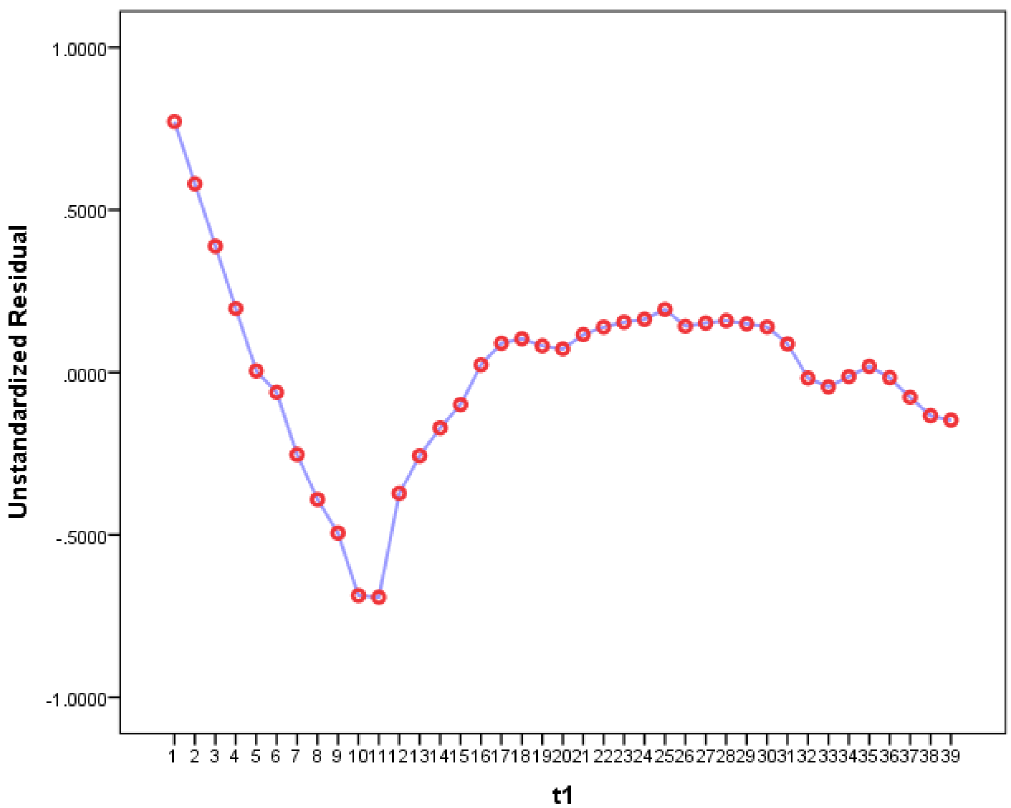
Model		Unstandardized Coefficients		Standardized Coefficients
		B	Std. Error	Beta
1	(Constant)	5.490	.094	
	t1	.192	.004	.992

a. Dependent Variable: log

تفسیر: مقدار ضرایب در جدول فوق محاسبه شده است حال میتوانیم روند را برای تمام زمان ها حساب کنیم و از سری کم کنیم تا با قی مانده ها محاسبه شوند که این کار را SPSS برای ما انجام میدهد و مقادیر باقی مانده را محاسبه میکند.

$$Y_t = X_t - m_t \qquad m_t = 5.490 + 0.192\, t$$

حال میخواهیم نمودار سری زمانی برای باقی مانده های حاصل از روش برازش چند جمله ای رسم کنیم تا ببینیم روند حذف شده است یا خیر.به نمودار زیر توجه کنید.



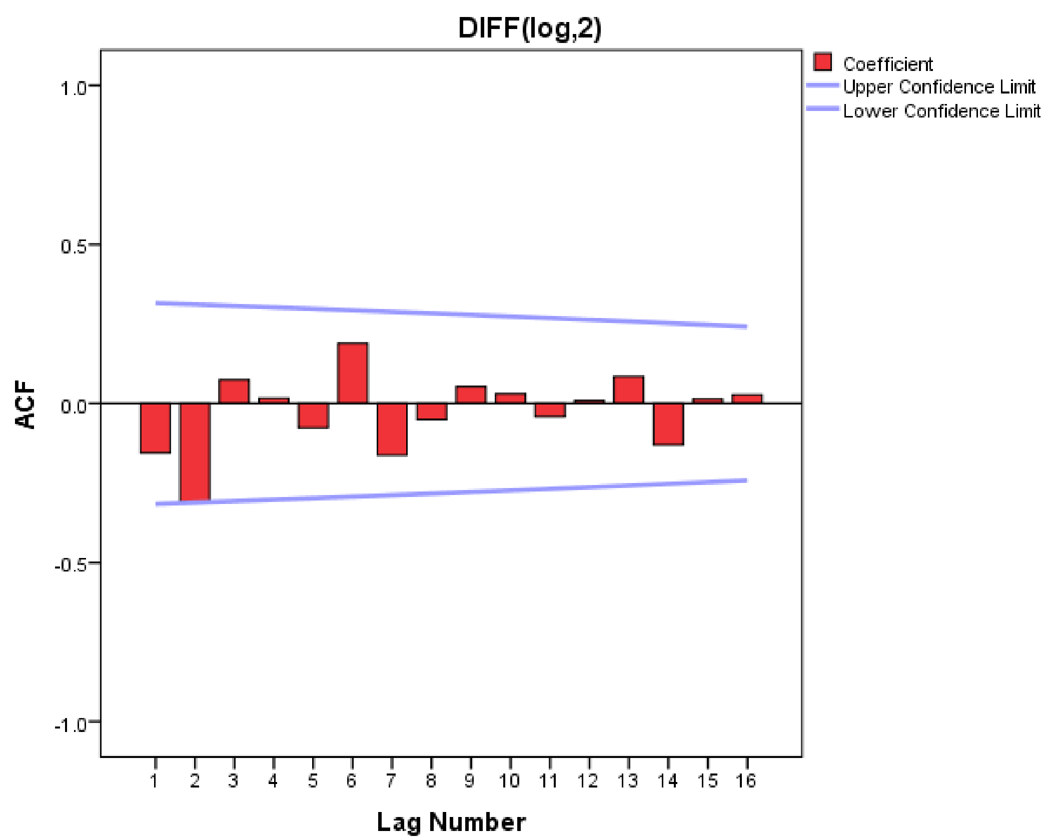
تفسیر: همانطور که در نمودار فوق میبینیم با استفاده از روش برازش چند جمله ای روند تقریبا حذف شده و سری ایستا شده است.

مرحله سوم : تشخیص مدل مناسب برای باقی مانده ها

در مرحله قبلی با استفاده از دو روش تفاضل گیری و برازش چند جمله ای به حذف روند پرداختیم حال وقت آن رسیده تا با استفاده از تکنیک های معرفی شده در فصل اول مدل مناسب را برای باقی مانده ها معرفی کنیم.

به منظور این کار ، تابع خود همبستگی را برای باقی مانده های حاصله از هر دو روش انجام شده (تفاضل گیری و برازش چند جمله ای) رسم میکنیم.

۱. رسم نمودار تابع خود همبستگی برای باقی مانده های حاصل از روش تفاضل گیری مرتبه ۲



تفسیر : همانطور که در نمودار فوق میبینیم بیش از ۹۵٪ مقادیر در بازه ی $\pm\frac{1.96}{\sqrt{39}}$ قرار دارند و خارج بازه نیستند. بنا براین میتوانیم سری تصادفی محض را برای باقی مانده های حاصل از روش تفاضل گیری مرتبه ۲ در نظر بگیریم.

مقادیر تابع خود همبستگی نمونه برای باقی مانده های حاصل از روش تفاضل گیری مرتبه ۲

Lag	Autocorrelation
1	-.156
2	-.310
3	.074
4	.017
5	-.076
6	.189
7	-.162
8	-.050
9	.053
10	.030
11	-.042
12	.009
13	.084
14	-.130
15	.014
16	.027

برآورد پارامتر های مدل :

تا اینجا توانستیم مدل سری تصادفی محض (I.I.D) را برای باقی مانده های حاصل از روش تفاضل گیری مرتبه ۲ در نظر بگیریم حال وقت آن است که پارامتر های مدل یعنی میانگین و واریانس را برآورد کنیم.

$$E(Y_t) = 0$$

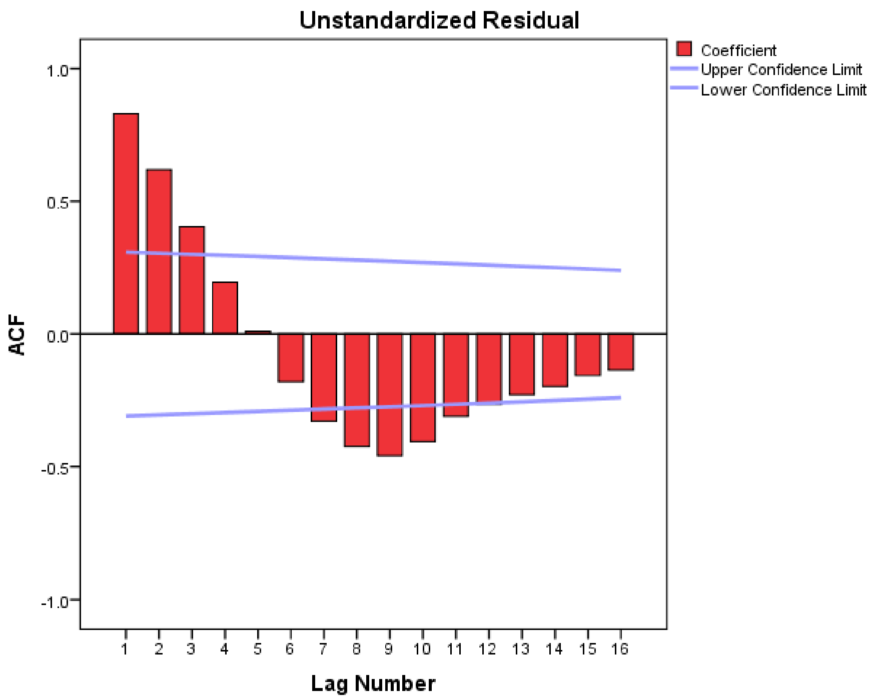
DIFF(log,2)

N	Valid	37
	Missing	2
Variance		.008

$$\widehat{\delta^2} = 0.008$$

نتیجه : بنا براین باقی مانده های حاصل از حذف روند با استفاده از روش تفاضل گیری یک سری تصادفی محض با میانگین صفر و واریانس ۰/۰۰۸ میباشند.

۲. رسم نمودار تابع خودهمبستگی برای باق مانده های حاصل از روش برازش چند جمله ای



تفسیر: با توجه به نمودار فوق که نمودار تابع خودهمبستگی برای باق مانده های حاصل از حذف روند با روش برازش چند جمله ای است میباشد. همانطور که مشاهده میکنیم بیشتر از ۵٪ داده ها خارج از فاصله اطمینان قرار دارند در نتیجه نمیتوانیم فرض I.I.D بودن را برای باق مانده ها در نظر بگیریم.با توجه به این که به صورت نمایی نزول کرده اند مدل اتورگرسیو ۱ یا ۲ را برای سری باق مانده ها پیشنهاد میدهیم.اما لازم است این پیشنهاد مورد بررسی و ارزیابی قرار گیرد.

برآورد پارامتر مدل

برای مدل اتورگرسیو لازم است پارامتر φ برآورد شود که این کار را با استفاده از روش حداقل مربعات برای هر دو مدل اتورگرسیو (1) و اتورگرسیو(2) انجام میدهیم :

۱. مدل اتورگرسیو (1) :

$$Y_t = \varphi Y_{t-1} + Z_t$$

برای ای مدل نیاز داریم پارامتر φ را برآورد کنیم و ثابت کنیم Z_t ها I.I.D نویز با واراینس مجهول δ^2 است که باید واریانس را نیز برآورد کنیم. φ را از دو روش حداقل مربعات خطا و تابع خودهمبستگی در زمان ۱ محاسبه میکنیم که از هر دو روش تقریبا به یم مقدار یکسان میرسیم.

برآورد φ با استفاده از روش حداقل مربعات خطا :

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.003	.026		-.123	.903
	Yt_1	.836	.091	.833	9.168	.000

a. Dependent Variable: Unstandardized Residual

برآورد φ با استفاده از مقدار تایع خود همبستگی در زمان ۱:

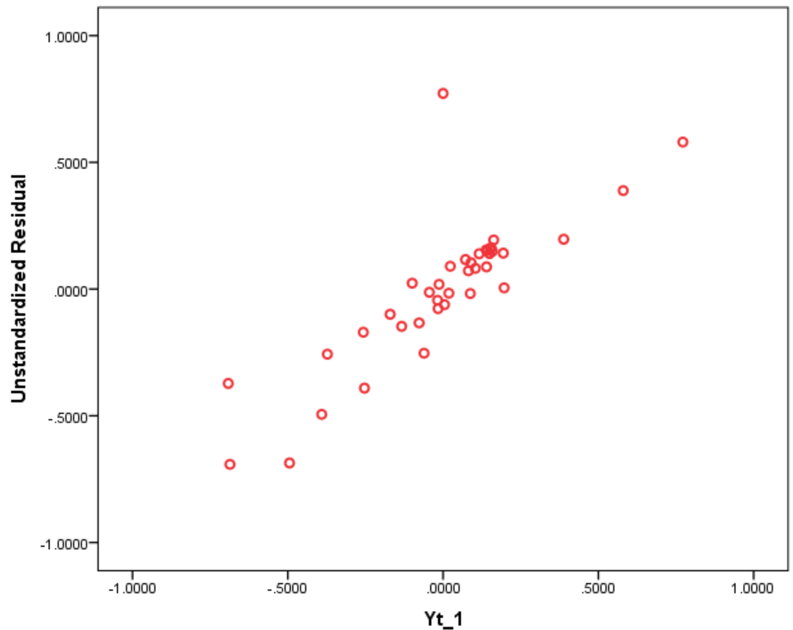
Autocorrelations

Series: Unstandardized Residual

Lag	Autocorrelation
1	.830
2	.619
3	.404
4	.196
5	.009
6	-.180
7	-.328
8	-.423
9	-.458
10	-.406
11	-.309
12	-.265
13	-.229
14	-.197
15	-.157
16	-.135

نتیجه : دیدیم که در هر دو روش بالا مقدار φ برابر ۰٫۸۳ میباشد.

برای این که مطمئن شویم مدل اتوروگرسیو (1) مدل مناسبی برای باقی مانده های حاصل از حذف روند با استفاده از رابزش چندجمله ای میباشد لازم است نمودار پراکنش برای Y_t ها در مقابل Y_{t-1} رسم کنیم و لازم است یک رابطه خطی بین آنها برقرار باشد.



تفسیر : میبینیم که یک رابطه خطی بین Y_t ها و Y_{t-1} برقرار است بنابراین مدل اتورگرسیو(1) برای این داده ها مدل مناسبی است پس به سراغ بررسی Z_t ها میرویم.

لازم است Z_t ها I.I.D با میانگین صفر و واریانس δ^2 باشند

$$E(Z_t) = 0$$

Zt1

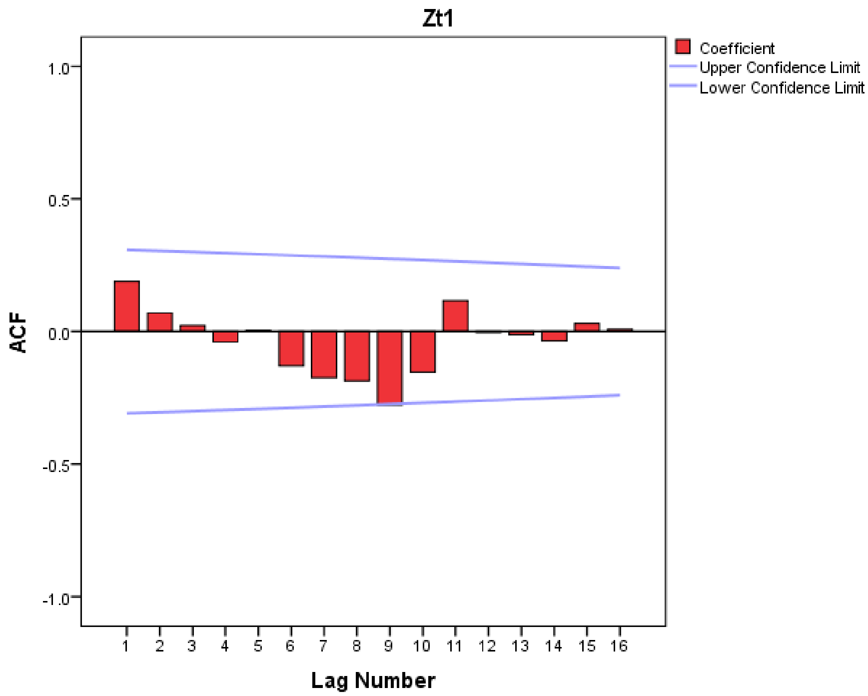
N	Valid	39
	Missing	0
Variance		.025

$$\widehat{\delta^2} = 0.025$$

همچنین برای بررسی i.i.d بودن توزیع Z_t ها از رسم نمودار تابع خودهمبستگی برای Z_t ها کمک میگیریم. به این گونه که ابتدا Z_t هارا از فرمول زیر برآورد میکنیم:

$$\widehat{Z}_t = Y_t - \varphi Y_{t-1}$$

رسم نمودار تابع خود همبستگی برای Z_t های حاصل از مدل اتورگرسیو(1)



تفسیر : همانطور که در نمودار بالا مشاهده میکنیم Z_t ، I.I.D هستند چرا که تمام مقادیر خودهمبستگی برای آنها داخل فاصله اطمینان قرار گرفته اند.

$$Z_t \sim I.I.D(0\ ,0.025)$$

۲. مدل اتورگرسیو (2)

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + Z_t$$

در این بخش نیز لازم است φ_1 و φ_2 را برآورد کنیم و بررسی کنیم که Z_t ها I.I.D باشند.سپس δ^2 را برای Z_t های حاصل از اتورگرسیو(2) را برآورد میکنیم.

برای برآورد φ_1 و φ_2 از روش کمترین مربعات خطا استفاده میکنیم.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-.002	.025		-.088	.931
	Yt_1	1.028	.162	1.024	6.334	.000
	YT_2	-.231	.163	-.229	-1.419	.165

a. Dependent Variable: Unstandardized Residual

تفسیر : همانطور که در جدول بالا میبینیم مقدار $\varphi_1 = 1.028$ و مقدار $\varphi_2 = -0.231$ محاسبه شده است.

حال باید Z_t ها را از طریق فرمول زیر برآورد کنیم و سپس میانگین و واریانس را برا Z_t ها محاسبه کنیم:

$$Y_t - \varphi_1 Y_{t-1} - \varphi_2 Y_{t-2} = \widehat{Z}_t$$

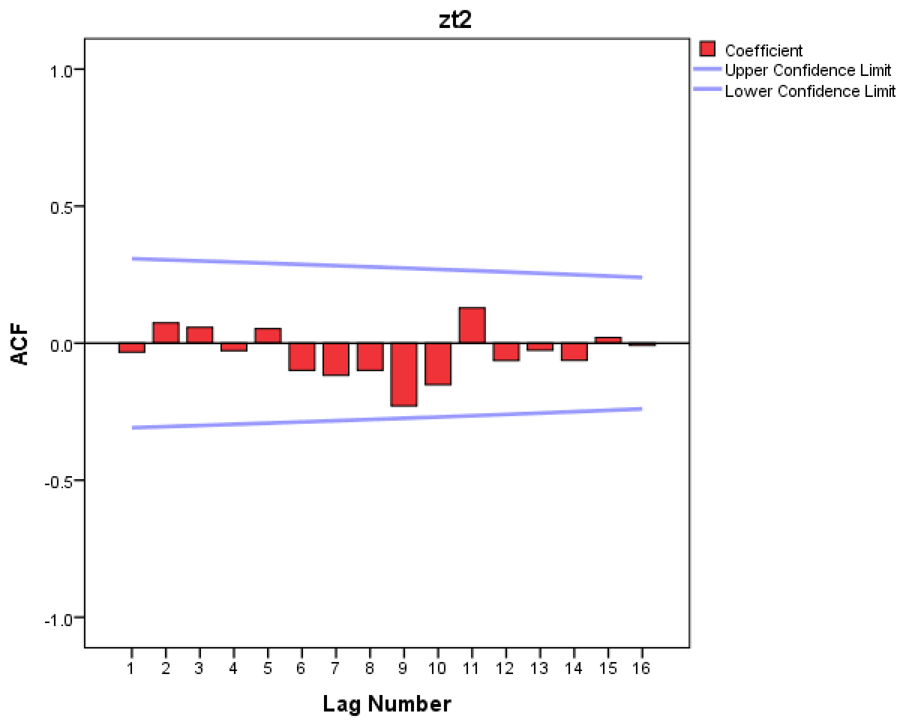
$$\mathrm{E}(\mathrm{Z_t}) = 0$$

zt2

N	Valid	39
	Missing	0
Variance		.024

$$\widehat{\delta^2} = 0.025$$

حال باید ثابت کنیم که Z_t ها I.I.D هستند برای این کار از نمودار تابع خودهمبستگی کمک میگیریم.



تفسیر : با توجه به این که تمام مقادیر همبستگی برای Z_t های حاصل از مدل اتورگرسیو(2) داخل فاصله اطمینان قرار گرفته اند میتوان نتیجه گرفت که Z_t ها ، I.I.D هستند. در نتیجه :

$$Z_t \sim I.I.D(0\,, 0.024)$$

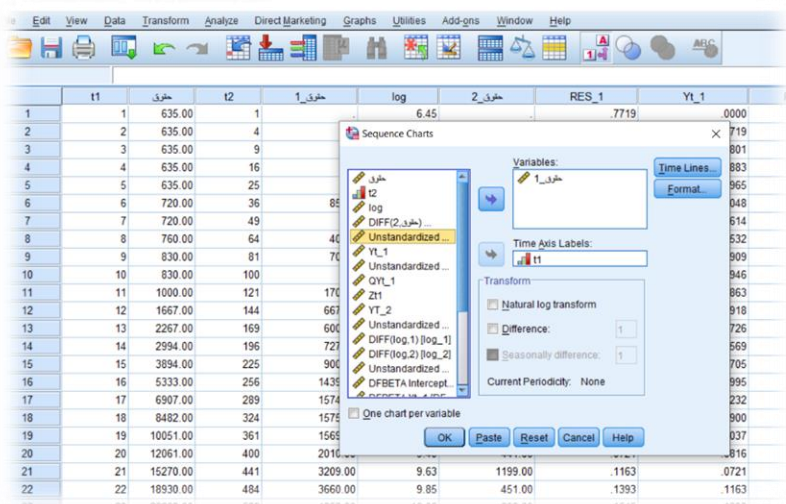
نتیجه کلی :

با توجه به این که هر دو مدل AR(2) , AR(1) برای باقی مانده ها مناسب بودند ما میتوانیم ملاک برتری برای انتخاب مدل را میزان واریانس آنها در نظر بگیریم که در مدل AR(2) واریانس برابر ۰/۰۲۴ است اما در AR(1) واریانس برابر ۰/۰۲۵ است که کمی بیشتر میباشد. بنابراین مدل AR(2) مدل مناسبی برای باقی مانده ها است.

"پیوست "

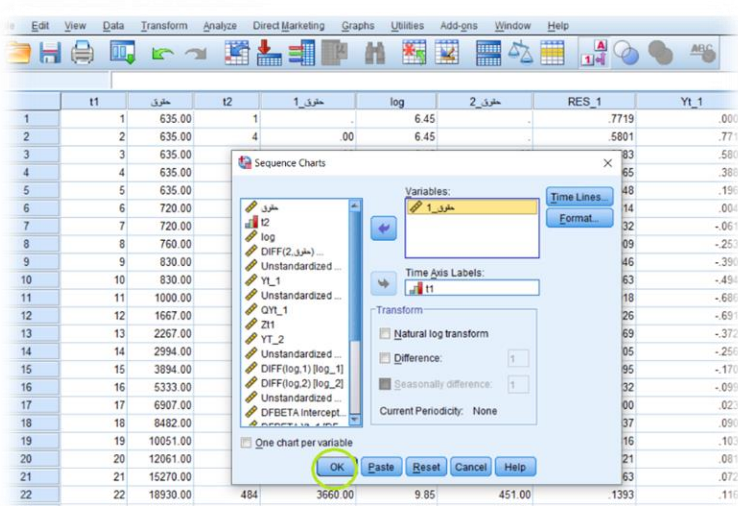
۱.مسیر رسم نمودار سری زمانی در SPSS

Analyze > Forecasting > sequence charts



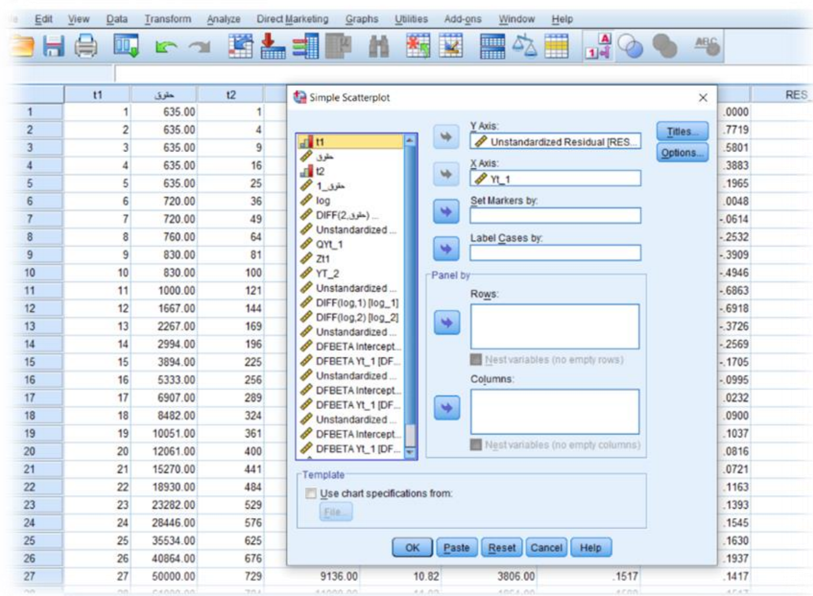
۲.مسیر رسم تابع خودهمبستگی در SPSS

Analyze > Forecasting > Autocorrelations



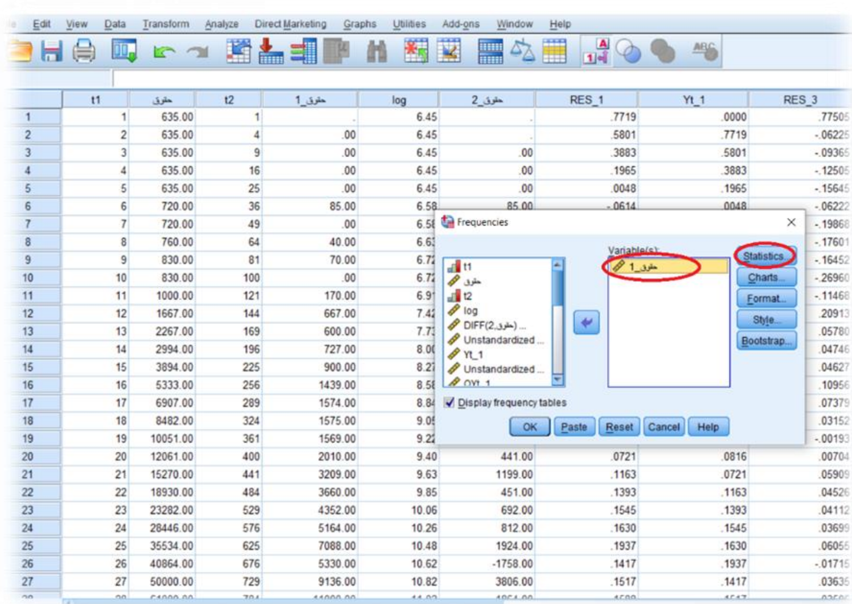
۳.مسیر رسم نمودار پراکنش در SPSS

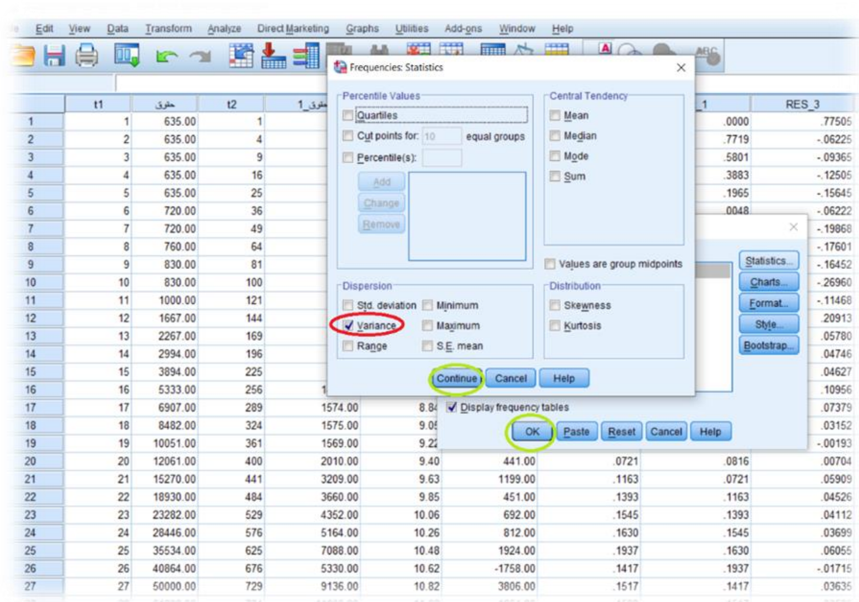
Graphs > legasy Dialogs > Scatter



۴.مسیر محاسبه ی واریانس در SPSS

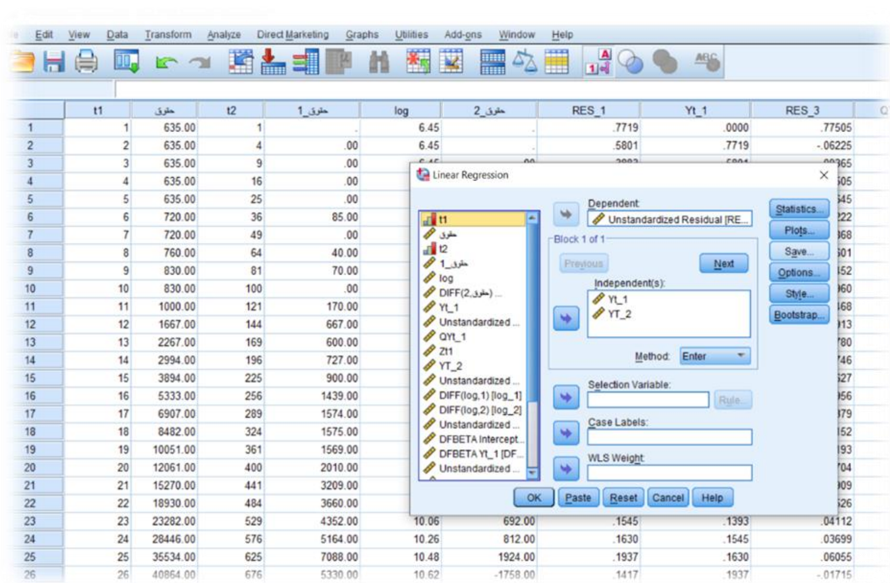
Analyze > Descriptive Statistics > Frequencies

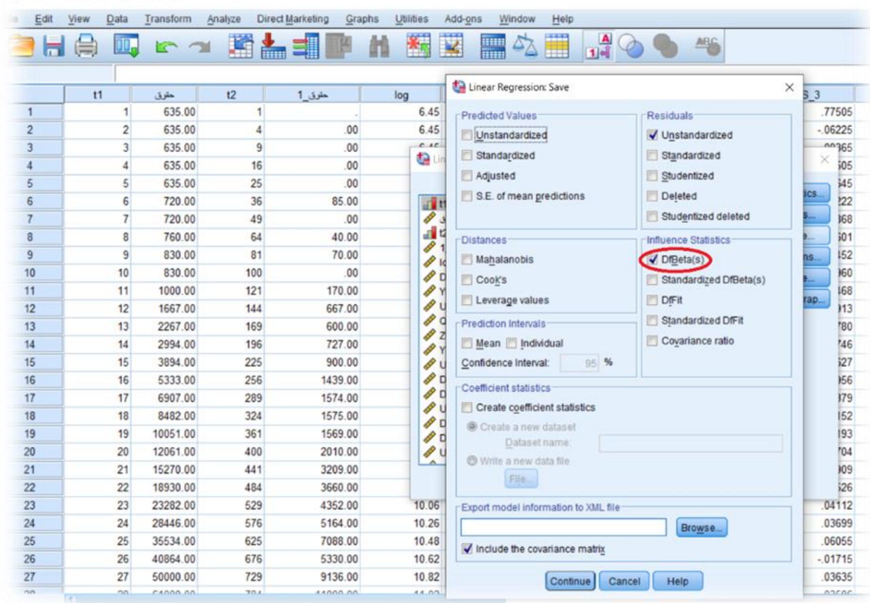




۵.روش براورد ضرایب در روش برازش چندجمله ای در SPSS

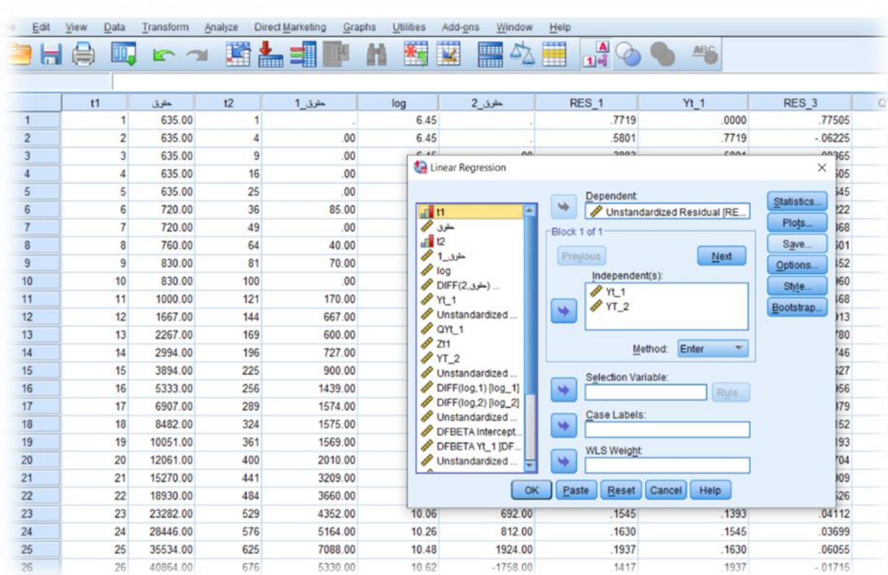
Analyze > Regression > linear

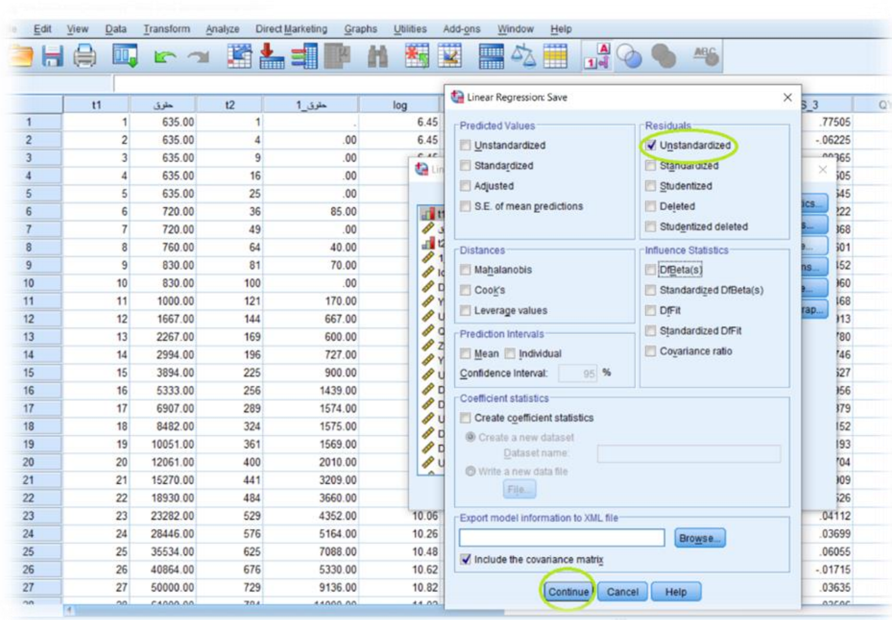




۶.محاسبه باقیمانده ها در روش برازش چند جمله ای در SPSS

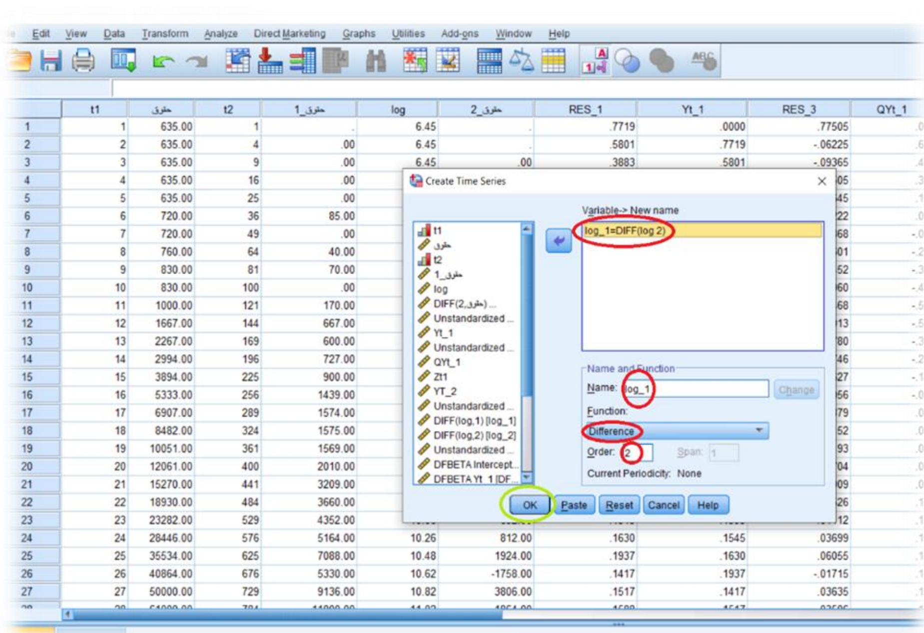
Analyze > Regression > linear





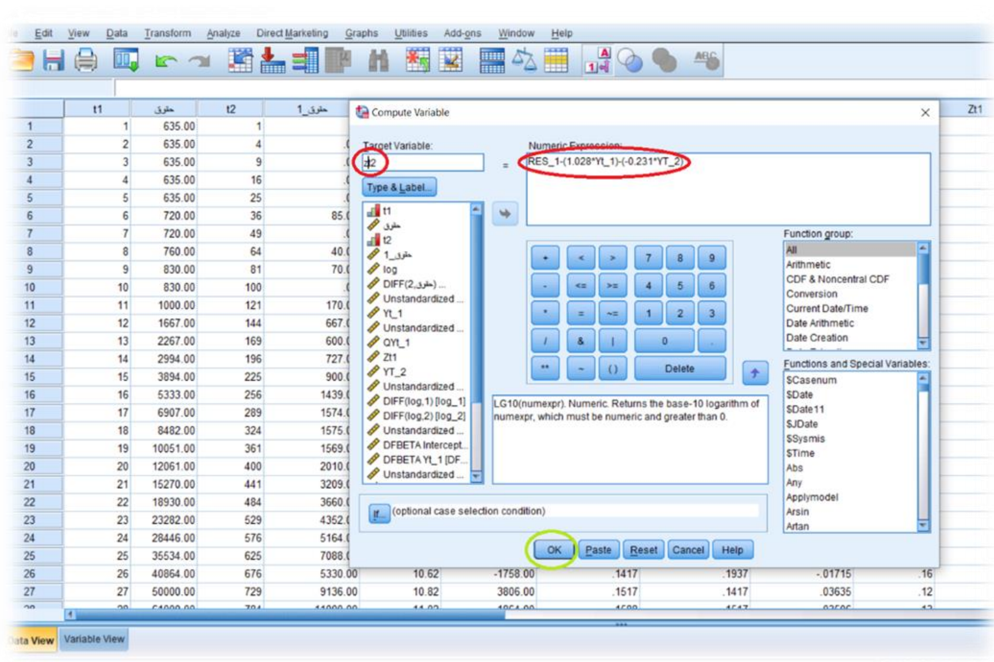
۷.روش تفاضل گیری برای حذف روند در SPSS

Transform > Create Time Series



۸.انجام اعمال جبری بر روی متغیر ها در SPSS

Transform > Compute Variable



داده های اصلی

۶۳۵/۰۰
۶۳۵/۰۰
۶۳۵/۰۰
۶۳۵/۰۰
۶۳۵/۰۰
۷۲۰/۰۰
۷۲۰/۰۰
۷۶۰/۰۰
۸۳۰/۰۰
۸۳۰/۰۰
۱۰۰۰/۰۰
۱۶۶۷/۰۰
۲۲۶۷/۰۰
۲۹۹۴/۰۰
۳۸۹۴/۰۰
۵۳۳۳/۰۰
۶۹۰۷/۰۰
۸۴۸۲/۰۰
۱۰۰۵۱/۰۰
۱۲۰۶۱/۰۰
۱۵۲۷۰/۰۰
۱۸۹۳۰/۰۰
۲۳۲۸۲/۰۰
۲۸۴۴۶/۰۰
۳۵۵۳۴/۰۰
۴۰۸۶۴/۰۰
۵۰۰۰۰/۰۰
۶۱۰۰۰/۰۰
۷۳۲۰۰/۰۰
۸۷۸۴۰/۰۰
۱۰۱۰۰۰/۰۰
۱۱۰۱۰۰/۰۰
۱۲۹۹۰۰/۰۰
۱۶۲۳۷۵/۰۰
۲۰۲۹۷۰/۰۰
۲۳۷۴۷۵/۰۰
۲۷۰۷۲۲/۰۰
۳۰۹۹۷۷/۰۰
۳۷۰۴۲۳/۰۰