

به نام خدا



دانشگاه علامه طباطبائی

پروژه درس سری زمانی

موضوع: بررسی و تحلیل مدل سری زمانی برای داده‌های مجموع ماهانه مسافران
خطوط هوایی ایالات متحده از سال ۱۹۴۹ تا ۱۹۶۰ با استفاده از نرم افزار SPSS

استاد درس: دکتر محمدرضا صالحی راد

پژوهشگر: علی شکارچی

تابستان ۱۴۰۳

مقدمه:

در این پژوهش الگوی سری زمانی برای داده‌های مجموع ماهانه مسافران خطوط هوایی ایالات متحده از سال ۱۹۴۹ تا ۱۹۶۰ (در مقیاس مناسب) به کمک نرم افزار SPSS مورد بررسی واقع میشود. این مجموعه داده از مجموعه داده‌های داخلی نرم افزار R به نام AirPassengers گرفته شده است. (که در سایت کگل بزرگترین انجمن علم داده جهان به ادرس <https://www.kaggle.com/datasets/chirag19/air-passengers> قابل دریافت است.) به دلیل ماهیت تاریخی و جزئیات زمانی ثابت، مجموعه داده‌های مسافران هوایی به عنوان یک منبع ارزشمند برای محققان و دانشجویان در زمینه‌های آمار، اقتصادسنجی و برنامه‌ریزی حمل‌ونقل عمل می‌کند. این پژوهش به کمک مدل‌های سری زمانی کلاسیک (نوع ARIMA) با تحلیل‌هایی بر موارد پایه‌ای برازش این مدل‌ها همچون تحلیل نمودارهای ACF و PACF و روش‌های سنتی براری برآورد مولفه‌های سری زمانی به دنبال تحلیل این داده‌ها و ارائه روشی مناسب برای پیش‌بینی مقادیر آینده آنها است.

طرح پژوهش:

داده‌های ثبتی مربوط به مجموع ماهانه مسافران خطوط هوایی ایالات متحده از سال ۱۹۴۹ تا ۱۹۶۰ (در مقیاس ۱۰۰۰۰۰ مسافر) در دسترس است. جدول زیر ۲۰ مشاهده اول تعداد مسافران را که از ماه اول سال ۱۹۴۹ ثبت شده‌اند نشان می‌دهد. جدول داده‌ها با همین الگو اطلاعات را تا ماه آخر سال ۱۹۶۰ دربر دارد.

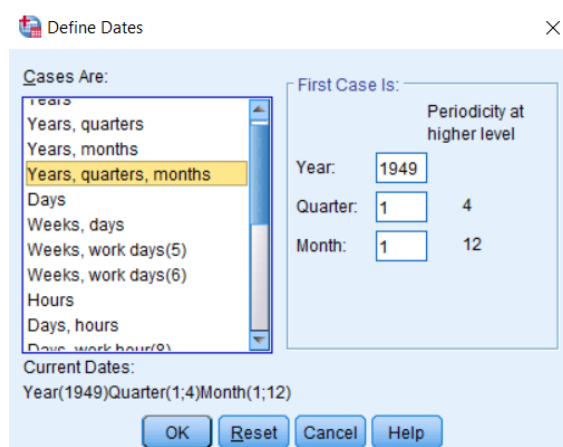
Month	Passengers
Jan-49	112
Feb-49	118
Mar-49	132
Apr-49	129
May-49	121
Jun-49	135
Jul-49	148
Aug-49	148
Sep-49	136
Oct-49	119
Nov-49	104
Dec-49	118
Jan-50	115
Feb-50	126
Mar-50	141
Apr-50	135
May-50	125
Jun-50	149
Jul-50	170

تحلیل و برازش مدل مناسب و پیش‌بینی مقادیر آینده سری زمانی داده‌ها در
مراحل گام به گام در ادامه بررسی میشوند:

گام اول:

پس از وارد کردن داده‌ها در نرم افزار SPSS تاریخ متناظر هر داده با فرمت مناسب برای نرم افزار تعریف میشود تا تحلیل صحیح داده‌ها امکان پذیر باشد.

دستور: Data » Define data and time

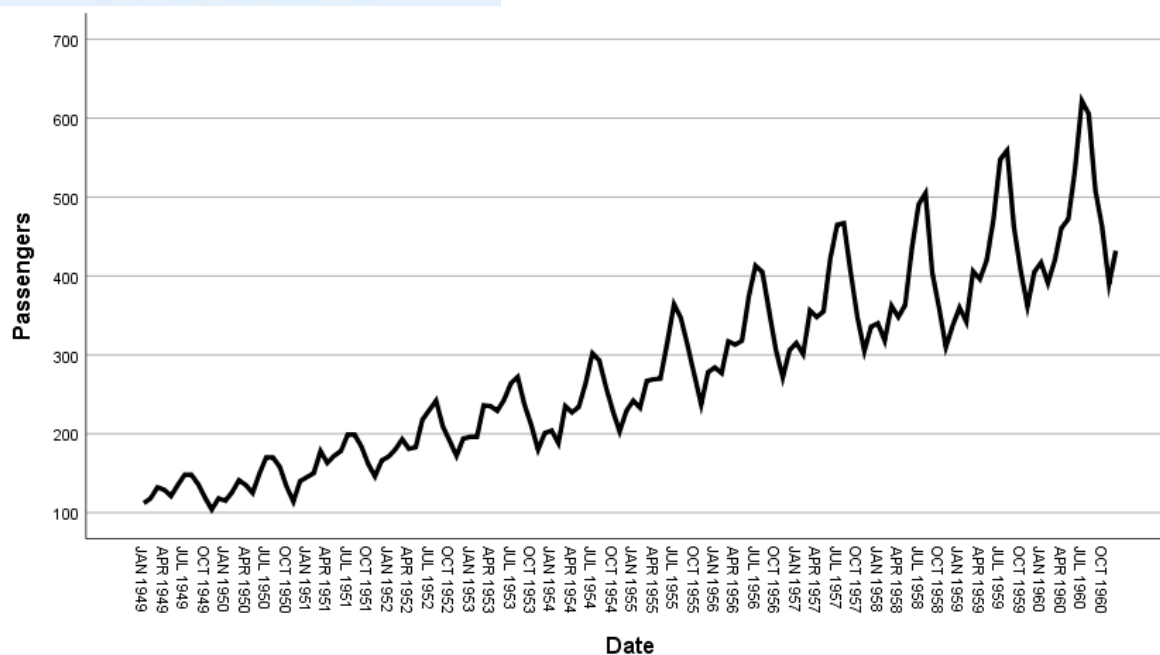
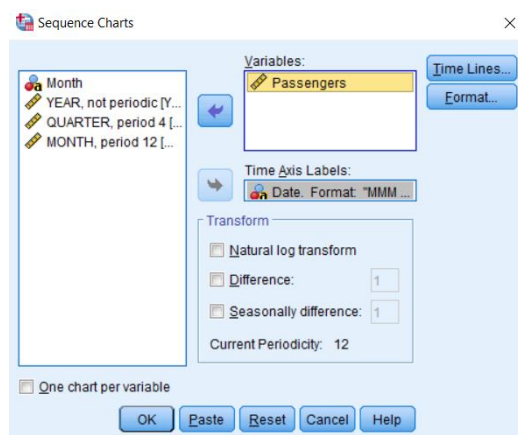


	Month	Passengers	YEAR_	QUARTER_	MONTH_	DATE_
1	1949-01	112	1949	1	1	JAN 1949
2	1949-02	118	1949	1	2	FEB 1949
3	1949-03	132	1949	1	3	MAR 1949
4	1949-04	129	1949	2	4	APR 1949
5	1949-05	121	1949	2	5	MAY 1949
6	1949-06	135	1949	2	6	JUN 1949
7	1949-07	148	1949	3	7	JUL 1949
8	1949-08	148	1949	3	8	AUG 1949
9	1949-09	136	1949	3	9	SEP 1949
10	1949-10	119	1949	4	10	OCT 1949
11	1949-11	104	1949	4	11	NOV 1949
12	1949-12	118	1949	4	12	DEC 1949
13	1950-01	115	1950	1	1	JAN 1950
14	1950-02	126	1950	1	2	FEB 1950
15	1950-03	141	1950	1	3	MAR 1950
16	1950-04	135	1950	2	4	APR 1950
17	1950-05	125	1950	2	5	MAY 1950
18	1950-06	149	1950	2	6	JUN 1950
19	1950-07	170	1950	3	7	JUL 1950
20	1950-08	170	1950	3	8	AUG 1950

گام دوم:

نمودار داده‌ها در مقابل زمان رسم میشود و به یک نتایج شهودی در مورد داده‌ها بررسی میشوند.

دستور: Analyse » Forecasting » Sequence charts



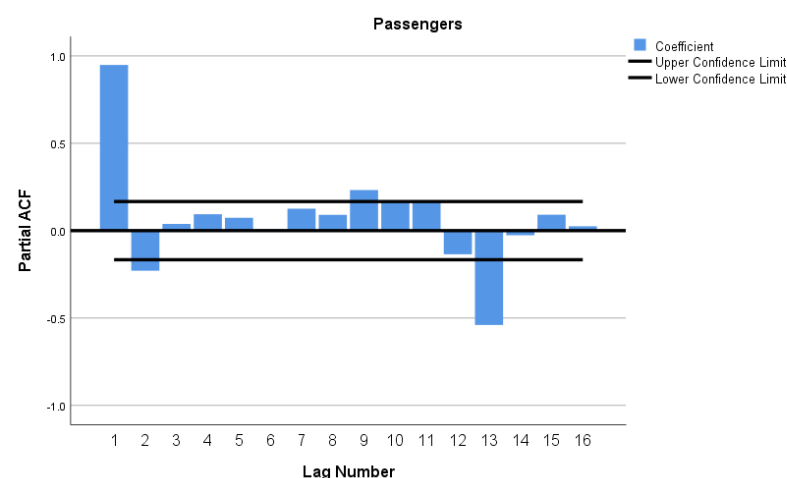
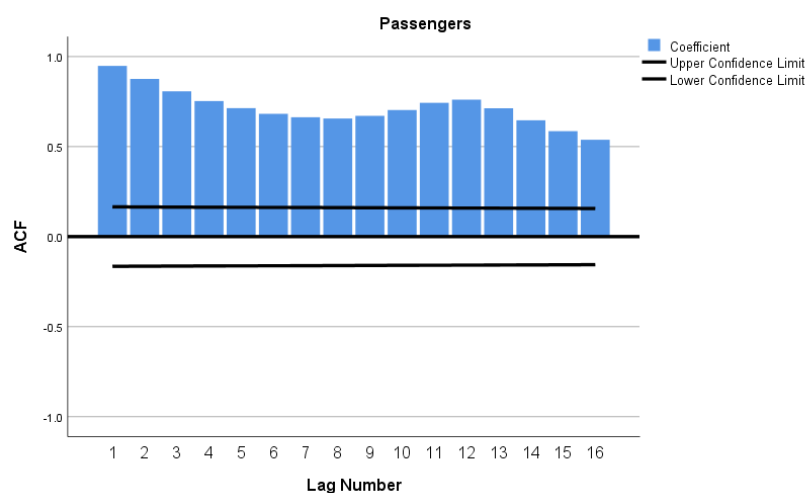
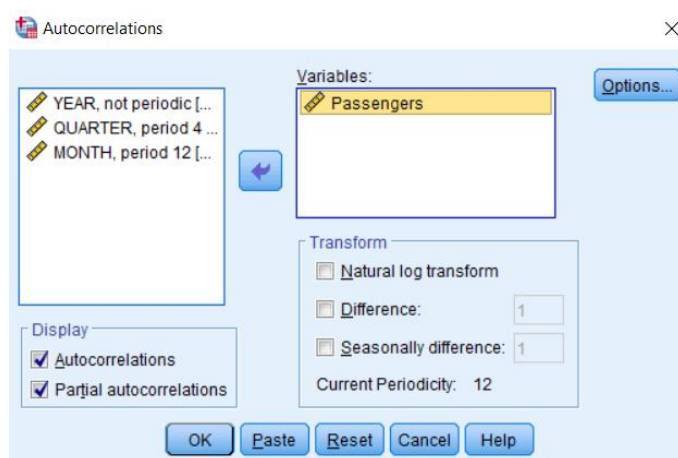
ملاحظه میشود که داده‌ها به صورت شهودی از شرایط ایستایی (میانگین و واریانس ثابت) برخوردار نیستند، و یک روند رو به بالا در داده‌ها مشاهده میشود. همچنین یک عامل فصلی در داده‌ها قابل پیش‌بینی میباشد که در ادامه به آن پرداخته میشود.

و بنظر میرسد برای رسیدن به ایستایی تفاضل‌گیری راهکاری مناسب باشد.

گام سوم:

در این گام به کمک نمودارهای ACF و PACF شرایط رسیدن به ایستایی برای سری زمانی بررسی میشود. ابتدا این نمودارها برای داده‌های اصلی رسم و تحلیل مناسب از روی نمودار انجام میشود.

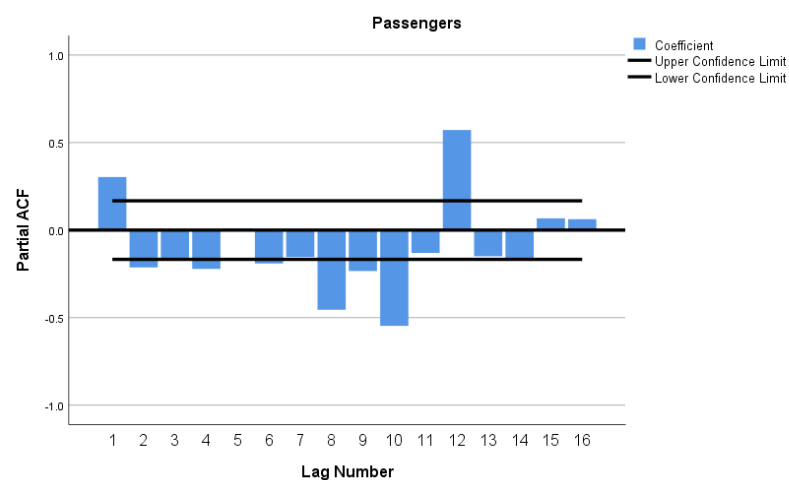
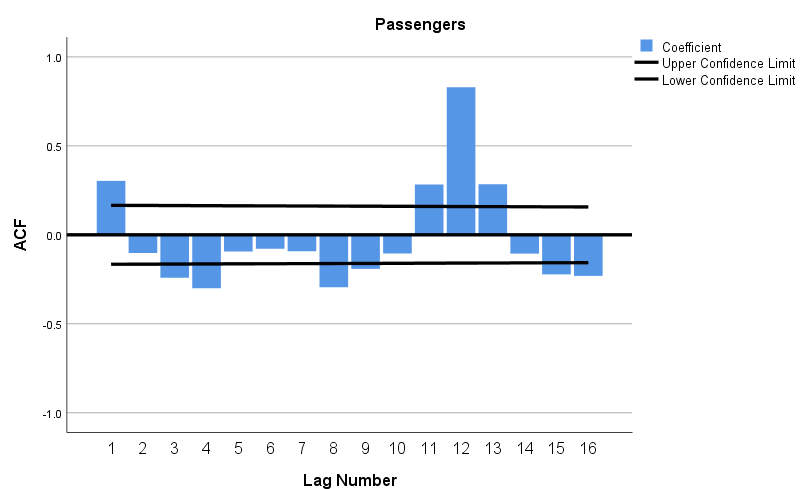
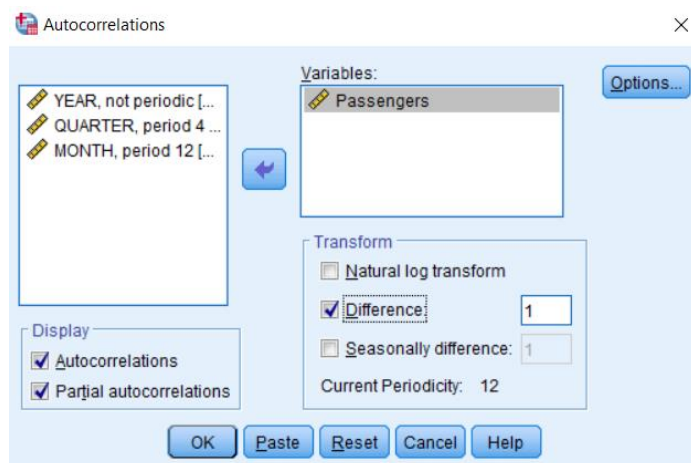
دستور: Analyse » Forecasting » Autocorrelation



ملاحظه میشود که نمودار ACF بسیار کند کاهش پیدا میکند و همانطور که از قبل انتظار میرفت سری زمانی ایستا نبوده و نیاز به تفاضل‌گیری دارد.

با انجام یک مرحله تفاضل‌گیری دوباره دو نمودار برای سری رسم و تحلیل میشوند.

دستور: Analyse » Forecasting » Autocorrelation



با کمی دقت در حرکت نوسان دار دو نمودار مشخص میشود که یک روند فصلی در ایجاد ناپایداری برای سری مورد انتظار است که در ادامه با یک تفاضل گیری فصلی این مورد بررسی و صحت آن تایید میشود.

با یک مرحله تفاضل‌گیری فصلی دو نمودار برای سری رسم و تحلیل میشوند.

دستور: Analyse » Forecasting » Autocorrelation

Autocorrelations

Variables:

- Passengers

Transform

☐ Natural log transform

☒ Difference: 1

☒ Seasonally difference: 1

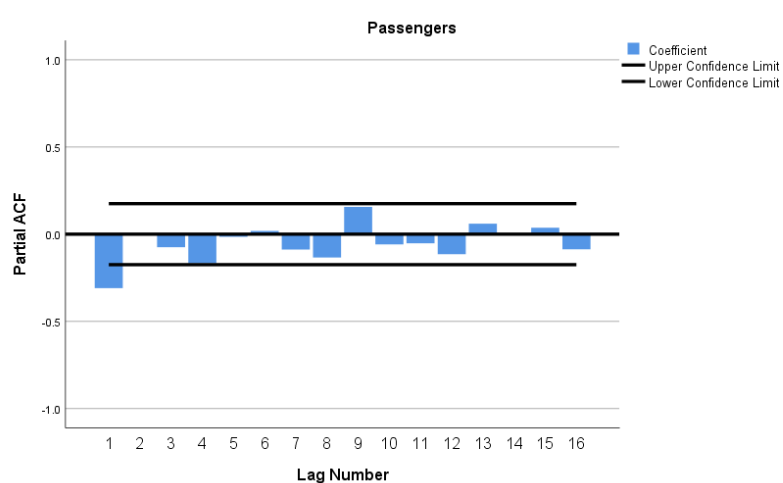
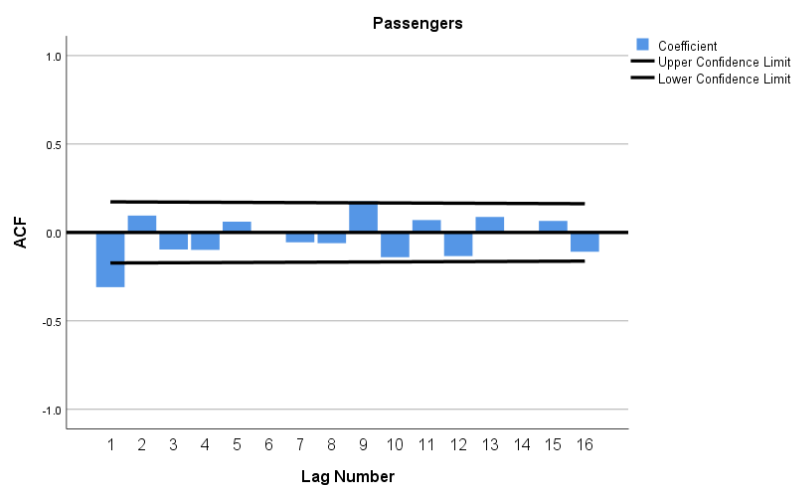
Current Periodicity: 12

Display

☒ Autocorrelations

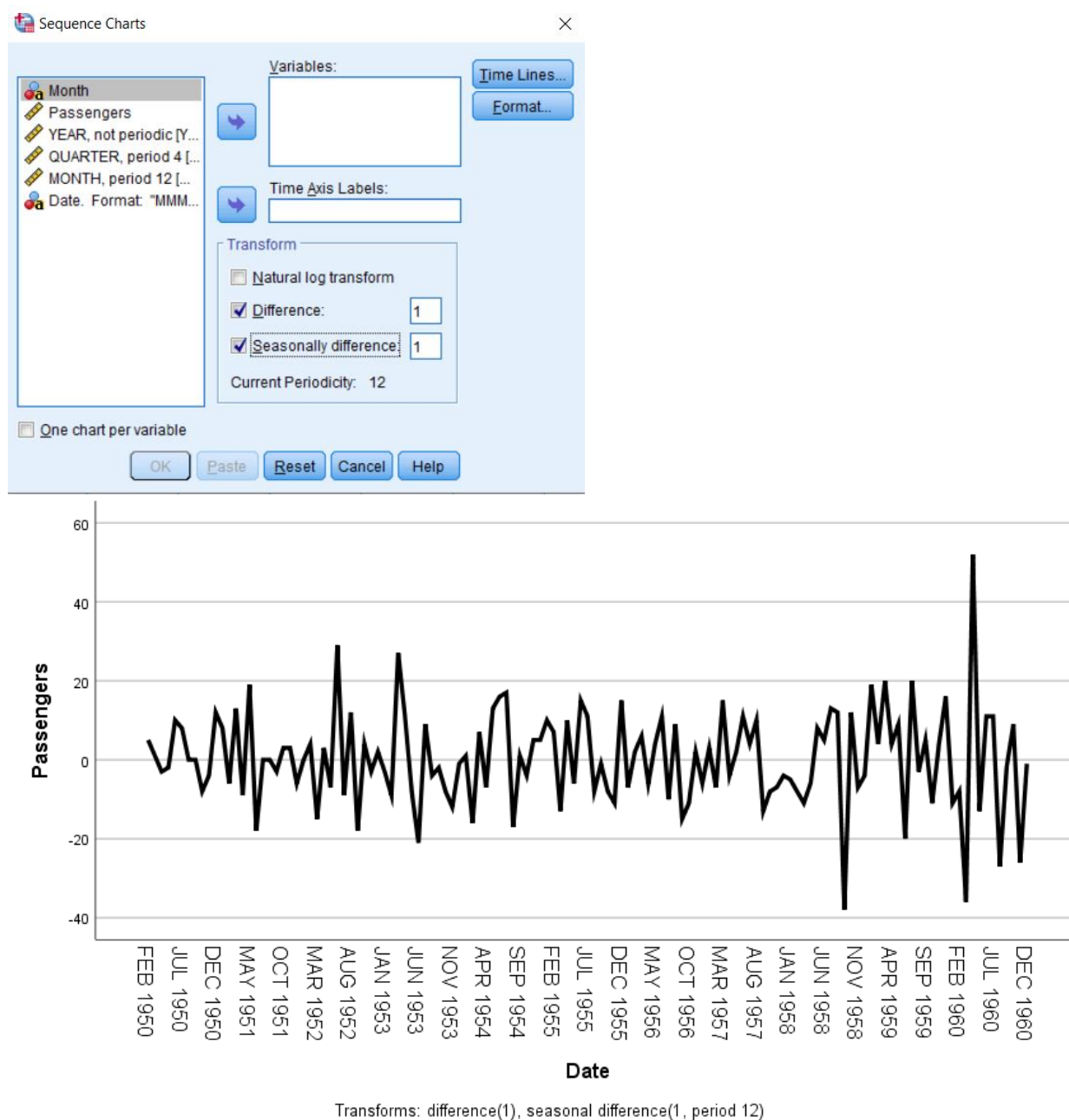
☒ Partial autocorrelations

OK Paste Reset Cancel Help



اینک مشخص شد با مرتبه اول تفاضل‌گیری از داده‌های اصلی و مرتبه اول تفاضل‌گیری فصلی، داده‌های سری به ایستایی میرسد و تنها یک مقدار معنادار در فاصله زمانی یکم در نمودار ACF (که از بازه اطمینان ۹۵ درصد خارج است) مشاهده میشود. که در برازش مدل سری زمانی ARMA حاصل از تفاضل‌گیری‌ها میتواند نشان‌دهنده پارامتر $q=1$ باشد.

برای بررسی صحت ادعای فوق نمودار حاصل از دو تفضیل‌گیری داده‌های اصلی و تفاضل‌گیری فصلی رسم میشود.
دستور: Analyse » Forecasting » Sequence charts



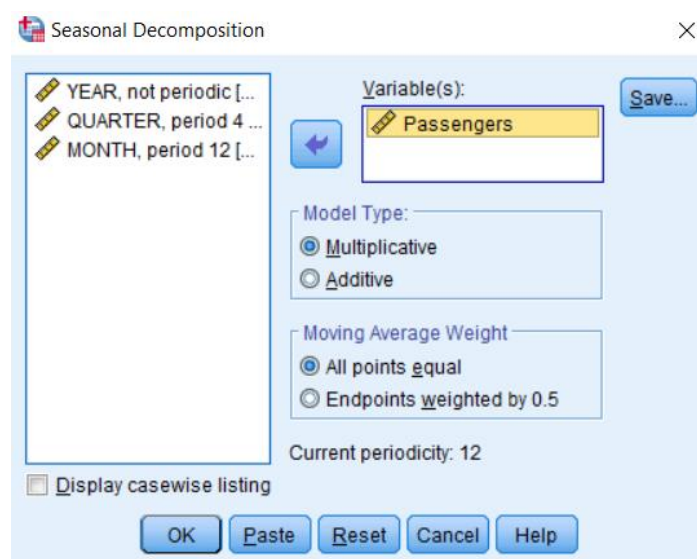
ملاحظه میشود تا حد بسیار قابل قبولی داده‌ها به شرایط ایستایی (ثابت بودن میانگین و واریانس) نزدیک شدند.

گام چهارم:

اینک به بررسی نمودارهای برآوردهای مولفه‌های سری زمانی به روش‌های سنتی یعنی برآوردهای مرکب از مولفه روند Z_t ، مولفه فصلی (S_t) و مولفه‌های نامنظم (e_t) پرداخته میشود. با فرض استقلال و جمع پذیری مولفه‌ها $Z_t = P_t + S_t + e_t$ برقرار است.

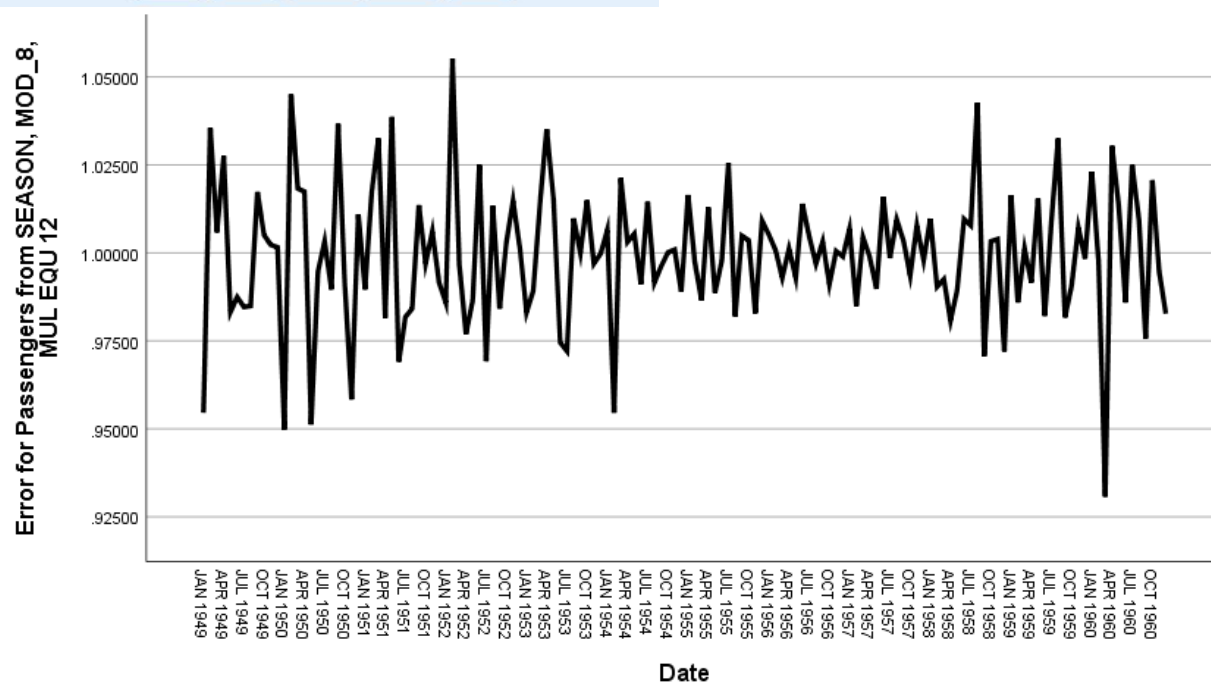
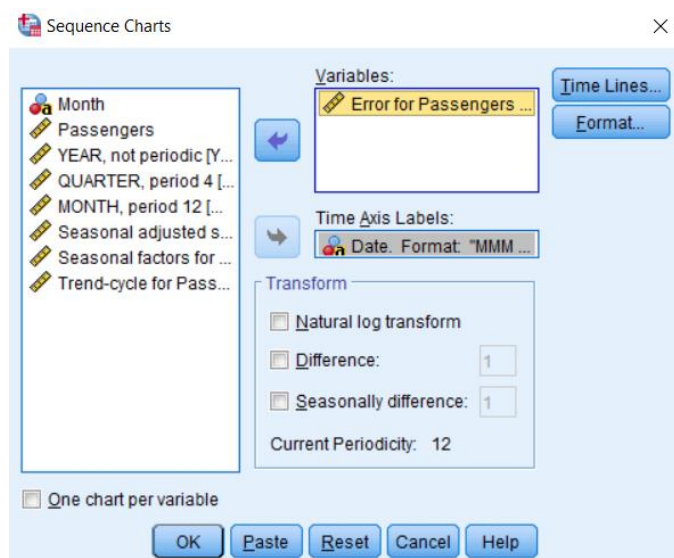
محاسبه این مولفه‌ها و برآوردهایشان به کمک دستور زیر صورت میگیرد و مقادیر برآورد شده در ستون‌هایی جدید به جدول داده‌ها اضافه میشوند.

دستور: Analyse » Forecasting » Seasonal Decomposition



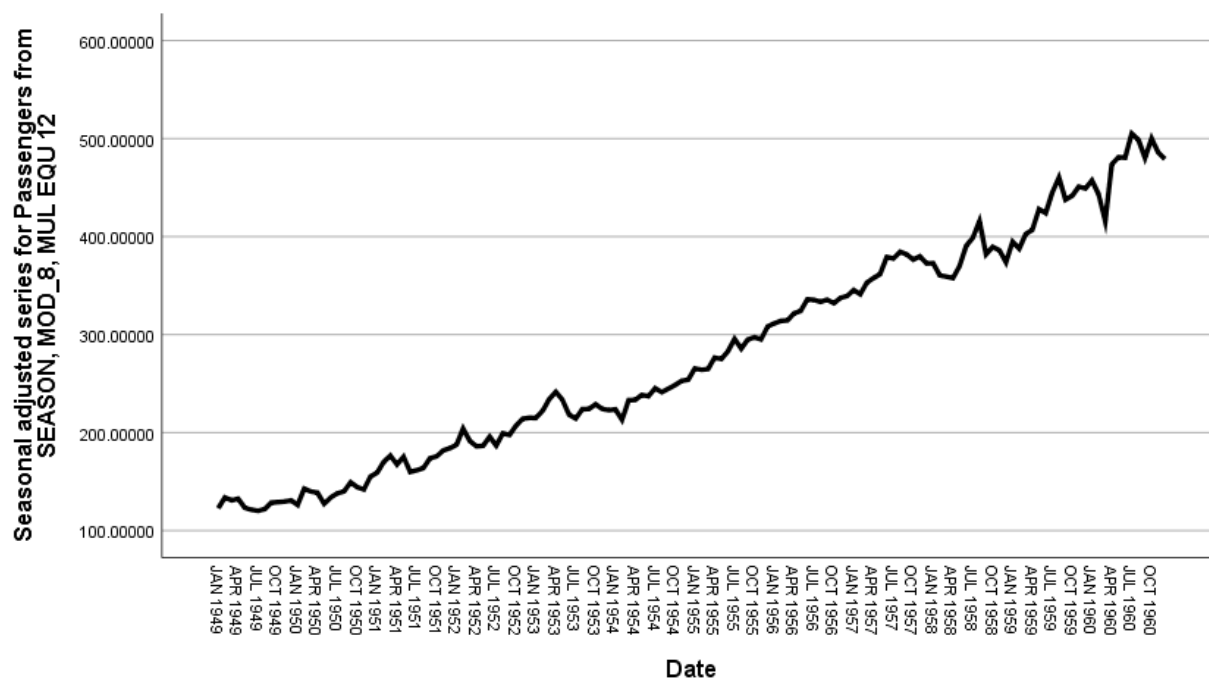
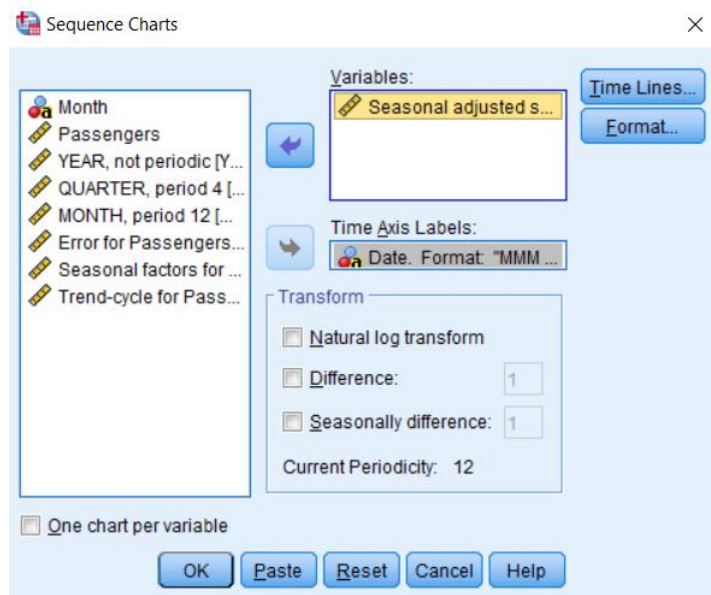
ابتدا اثر مولفه‌های نامنظم سری (e_t) یعنی نوفه‌ها یا همان عوامل تصادفی که ناشی از اختلاف مقدار نظری و تجربی هستند با اجرای دستور زیر برای متغیر خطا که پس از دستور بالا ایجاد شد اجرا میشود و نمودار حاصل نشان داده میشود.

دستور: Analyse » Forecasting » Sequence charts



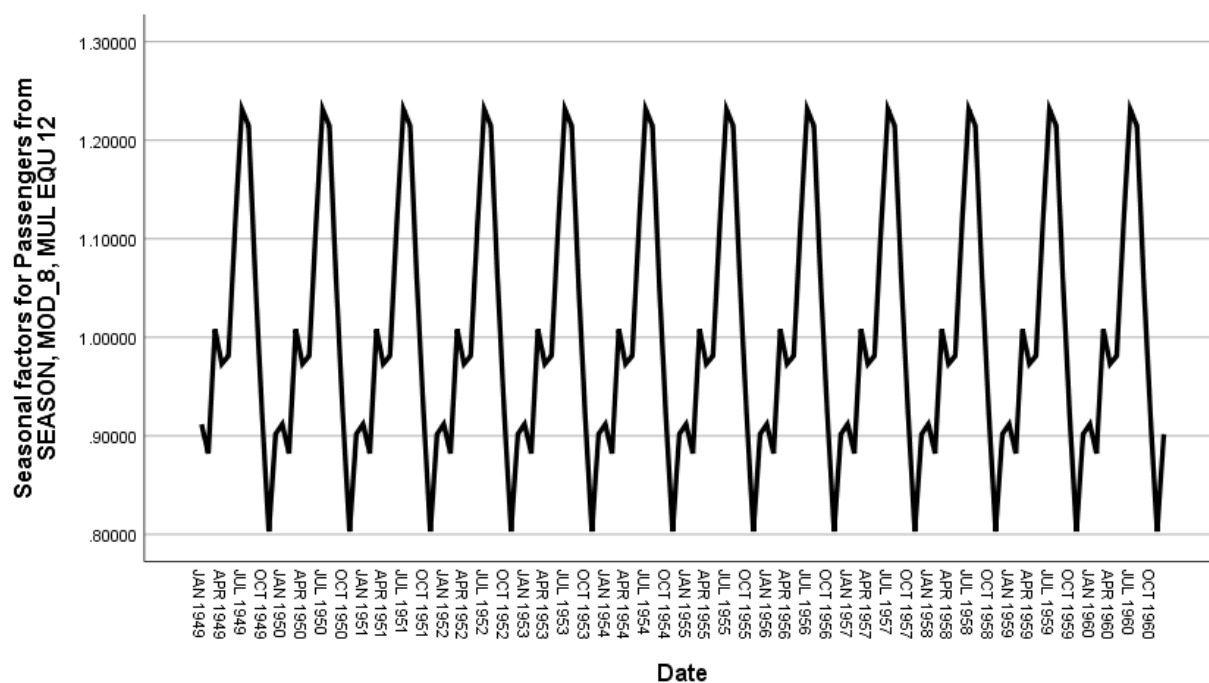
سپس نمودار سری فصلی تعدیل شده (که حاصل از حذف برآورد نوسانات فصلی است. $\hat{N}_t = Z_t - \hat{S}_t$) با اجرای دستور زیر برای متغیر داده‌های تعدیل شده که پس از دستور Seasonal Decomposition ایجاد شد اجرا میشود و نمودار حاصل نشان داده میشود.

دستور: Analyse » Forecasting » Sequence charts



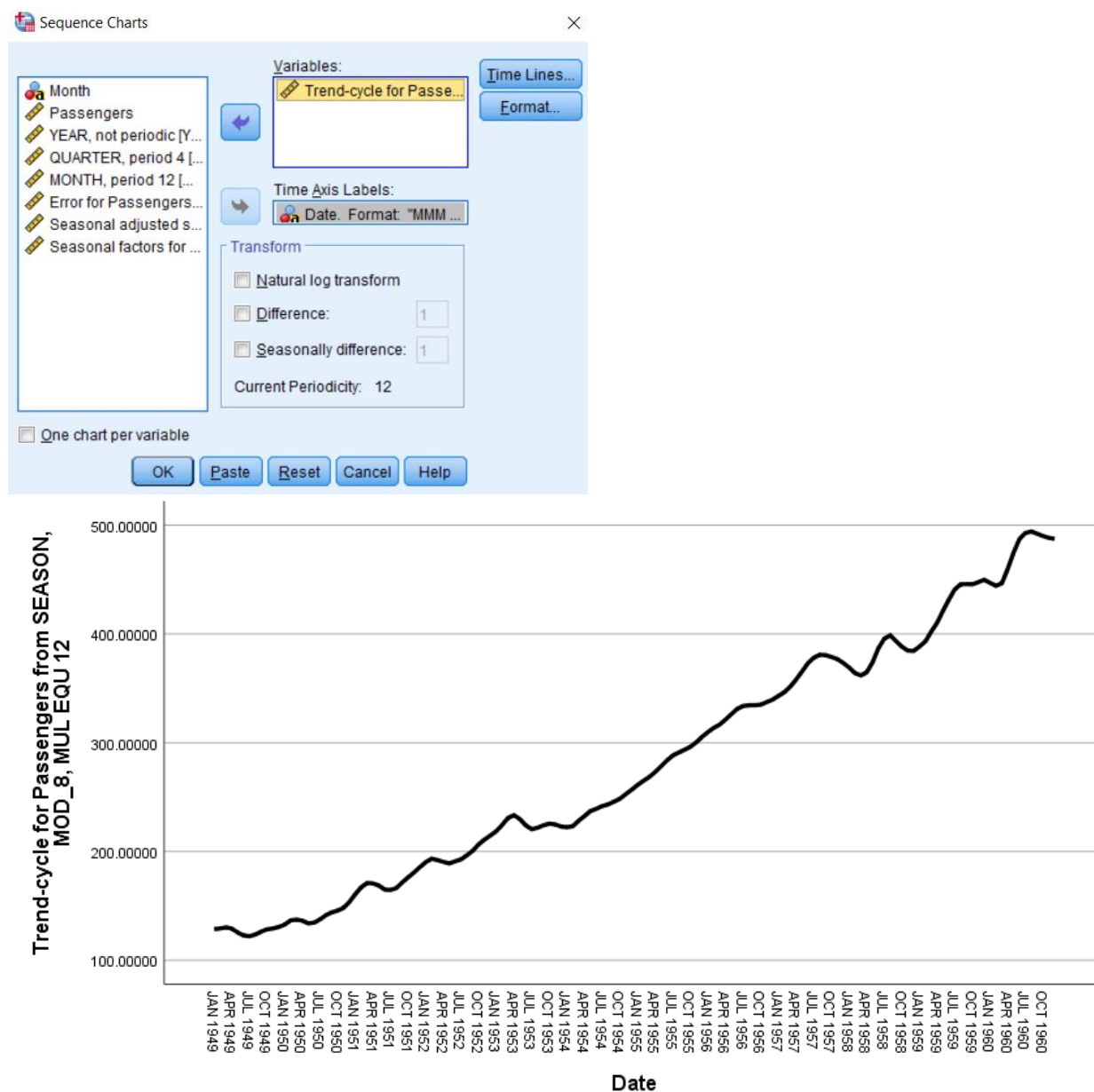
در این قسمت نمودار سری برآورد مولفه فصلی (\hat{S}_t) با دوره فصلی $s=12$ با اجرای دستور زیر برای متغیر برآورد مولفه فصلی که پس از دستور Decomposition Seasonal ایجاد شد اجرا میشود و نمودار حاصل نشان داده میشود.

دستور: Analyse » Forecasting » Sequence charts



و در آخر نمودار سری برآورد مولفه روند (\hat{P}_t) که با استفاده از روش میانگین متحرک برآورد میشود با اجرای دستور زیر برای متغیر برآورد مولفه روند که پس از دستور Decomposition Seasonal ایجاد شد اجرا میشود و نمودار حاصل نشان داده میشود.

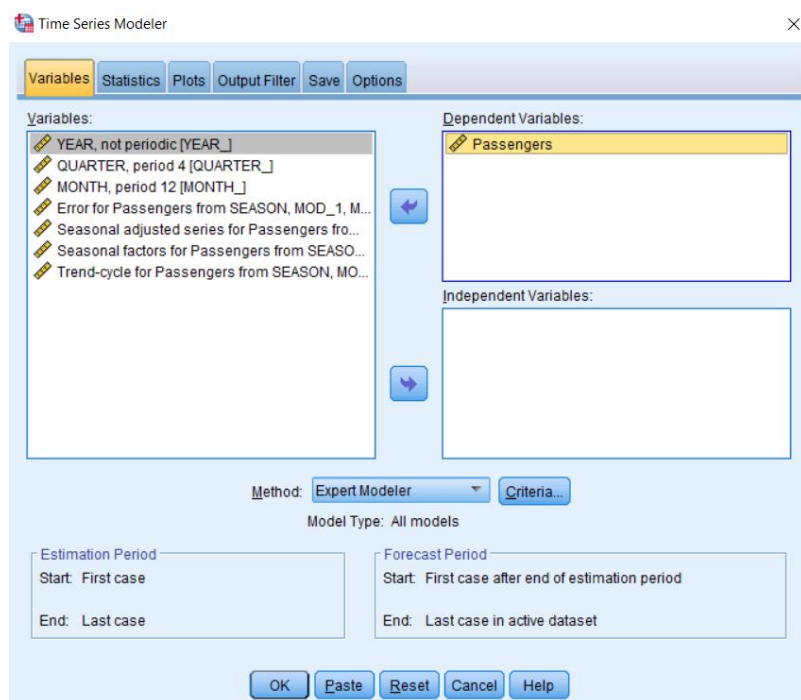
دستور: Analyse » Forecasting » Sequence charts



گام آخر:

شایان توجه است که نرم افزار SPSS توانایی یافتن مناسب‌ترین مدل برای داده‌های سری زمانی را دارد (expert modeler)، در اینجا از این قابلیت مطابق دستور زیر استفاده میشود و مناسب ترین مدل پیشنهادی نرم افزار SPSS با مدل کشف شده با کمک تحلیل‌های گام به گام نمودارهای ACF و PACF که در گام‌های قبل پیش‌بینی شدند مقایسه میشود.

دستور: Analyse » Forecasting » Create Traditional Models



Model Description

Model Type			
Model ID	Passengers	Model_1	ARIMA(0,1,1)(0,1,1)

Model Statistics

Model	Number of Predictors	Model Fit statistics Stationary R-squared	Ljung-Box Q(18)			Number of Outliers
			Statistics	DF	Sig.	
Passengers-Model_1	0	.324	12.571	16	.704	0

ملاحظه میشود که مدل پیشنهادی نرم افزار با مدل پیش‌بینی شده از تحلیل نمودارها کاملاً مطابقت دارد و نتیجه میدهد تحلیل‌ها به درستی انجام شده و مدل سری زمانی مناسب برای داده‌های مجموع ماهانه مسافران خطوط هوایی ایالات متحده از سال ۱۹۴۹ تا ۱۹۶۰ مدل $ARIMA(0,1,1)(0,1,1)$ است.

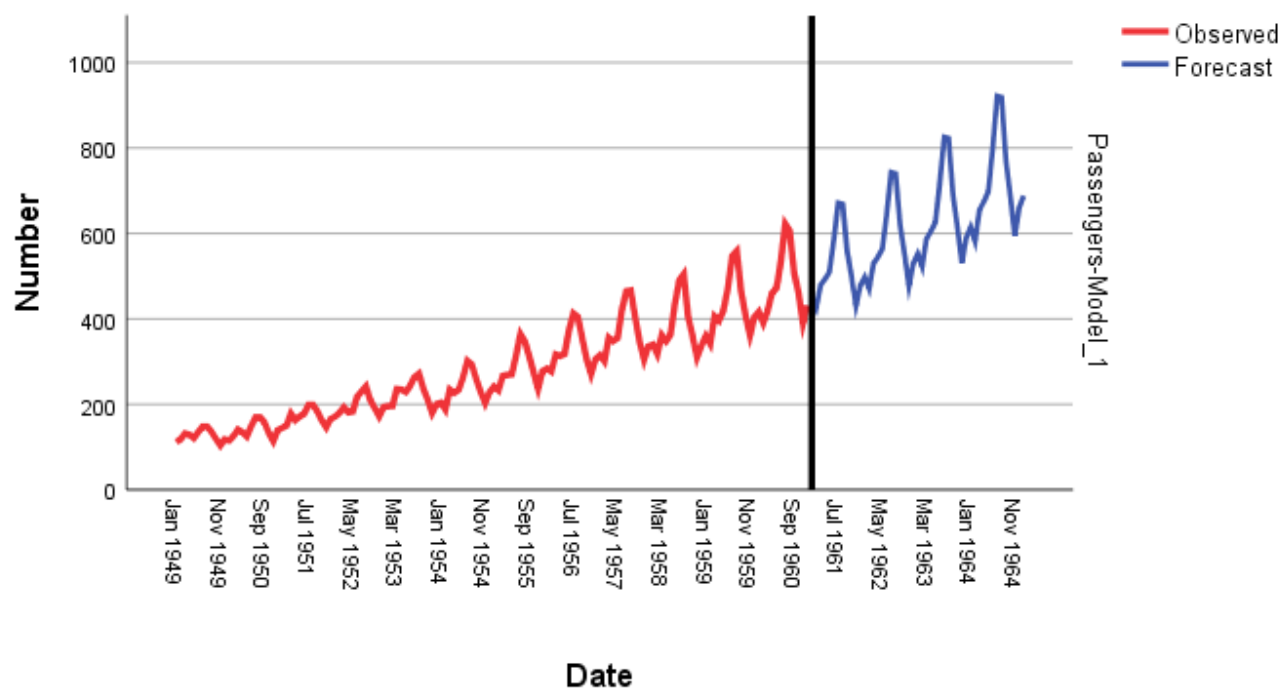
همچنین با توجه به اطلاعات جدول Model Statistics و آماره $Ljung-Box\ Q(18)$ که همبستگی را در باقیمانده‌های مدل آزمون میکند، طبق اینکه $p\text{-value} (<0.05)$ نشان‌دهنده همبستگی معناداری در خطاها است که نشانه‌ای از ضعف مدل مورد برازش است، با توجه به مقدار این آماره در مدل برازش داده شده یعنی ۰/۷۰۴، که بسیار بالاتر از ۰/۰۵ است احتمالاً هیچ ارتباطی بین باقیمانده‌ها در ۱۸ تاخیر وجود ندارد. همچنین تعداد پیش‌بینی کننده‌ها و تعداد نقاط پرت هر دو صفر هستند که به ترتیب بیانگر ساده بودن مدل (مقادیر آینده فقط بر اساس مقادیر گذشته خود سری پیش‌بینی میشوند) و عدم وجود مشاهدات بیرونی در سری زمانی مورد مطالعه است.

و لازم به ذکر است یکی از قابلیت‌های قسمت Create Traditional Models نرم افزار SPSS پیش‌بینی مقادیر آینده سری زمانی مطابق با مدل سری زمانی برازش داده شده در بازه زمانی دلخواه است. با استفاده از دستور زیر برای بازه زمانی ۴ ساله مطابق با مدل سری زمانی برازش داده شده پیش‌بینی صورت میگیرد.

دستور: **Analyse » Forecasting » Create Traditional Models(options)**

The screenshot shows the 'Options' tab of the 'Time Series Modeler' dialog box. The 'Forecast Period' section has two radio buttons: 'First case after end of estimation period through last case in active dataset' (unselected) and 'First case after end of estimation period through a specified date' (selected). Below this, a 'Date' table is shown with columns 'Year', 'Quarter', and 'Month'. The 'Year' column contains '1964', 'Quarter' contains '1', and 'Month' contains '1'. The 'User-Missing Values' section has two radio buttons: 'Treat as invalid' (selected) and 'Treat as valid' (unselected). The 'Confidence Interval Width (%)' is set to '95'. The 'Prefix for Model Identifiers in Output' is set to 'Model'. The 'Maximum Number of Lags Shown in ACF and PACF Output' is set to '24'. At the bottom, there are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

Year	Quarter	Month
1964	1	1



این پیش‌بینی به طور شهودی نیز مورد تایید است چرا که روند رو به بالا (رو به رشد) تعداد مسافران خطوط هوایی و اثر روند فصلی سری زمانی در پیش‌بینی انجام شده به وضوح قابل درک است.

و در آخر یکی دیگر از قابلیت‌های شایان توجه نرم افزار SPSS در قسمت Create Traditional Models افزودن ستون‌های جدید به جدول داده‌ها شامل مواردی چون داده‌های مورد پیش‌بینی مطابق با مدل برازش داده شده، تعیین کران بالا و پایین برای هر پیش‌بینی (با اطمینان ۹۵ درصد) و تعیین مقدار نوفه برای هر داده میباشد که موارد نامبرده با اجرای دستور زیر به جدول داده‌ها اضافه میشوند.

دستور: Analyse » Forecasting » Create Traditional Models(save)

Time Series Modeler

Variables Statistics Plots Output Filter **Save** Options

Save Variables


Variables:

Description	Save	Variable Name Prefix
Predicted Values	<input checked="" type="checkbox"/>	Predicted
Lower Confidence Limits	<input checked="" type="checkbox"/>	LCL
Upper Confidence Limits	<input checked="" type="checkbox"/>	UCL
Noise Residuals	<input checked="" type="checkbox"/>	NResidual


For each item you select, one variable is saved per dependent variable.

Export Model File

XML File: [Browse...](#)

 XML files are only compatible with SPSS applications.

PMML File: [Browse...](#)

 PMML files are compatible with PMML-compliant applications, including SPSS.

OK Paste Reset Cancel Help

	Month	Passengers	YEAR_	QUARTER_	MONTH_	DATE_	Predicted_Passengers_Model	LCL_Passengers_Model_1	UCL_Passengers_Model_1	NResidual_Passengers_Model
2	1949-02	118	1949	1	2	FEB 1949	-	-	-	-
3	1949-03	132	1949	1	3	MAR 1949	-	-	-	-
4	1949-04	129	1949	2	4	APR 1949	-	-	-	-
5	1949-05	121	1949	2	5	MAY 1949	-	-	-	-
6	1949-06	135	1949	2	6	JUN 1949	-	-	-	-
7	1949-07	148	1949	3	7	JUL 1949	-	-	-	-
8	1949-08	148	1949	3	8	AUG 1949	-	-	-	-
9	1949-09	136	1949	3	9	SEP 1949	-	-	-	-
10	1949-10	119	1949	4	10	OCT 1949	-	-	-	-
11	1949-11	104	1949	4	11	NOV 1949	-	-	-	-
12	1949-12	118	1949	4	12	DEC 1949	-	-	-	-
13	1950-01	115	1950	1	1	JAN 1950	-	-	-	-
14	1950-02	126	1950	1	2	FEB 1950	121	111	133	0
15	1950-03	141	1950	1	3	MAR 1950	139	128	151	0
16	1950-04	135	1950	2	4	APR 1950	137	126	149	0
17	1950-05	125	1950	2	5	MAY 1950	128	117	139	0
18	1950-06	149	1950	2	6	JUN 1950	141	129	153	0
19	1950-07	170	1950	3	7	JUL 1950	160	147	173	0
20	1950-08	170	1950	3	8	AUG 1950	166	152	180	0
21	1950-09	158	1950	3	9	SEP 1950	155	142	168	0
22	1950-10	133	1950	4	10	OCT 1950	137	126	149	0
23	1950-11	114	1950	4	11	NOV 1950	118	108	128	0
24	1950-12	140	1950	4	12	DEC 1950	131	120	142	0
25	1951-01	145	1951	1	1	JAN 1951	134	123	145	0

مشاهده میشود که موارد ذکر شده به جدول اضافه شدند. همچنین عدم وجود پیش‌بینی برای ۱۳ داده اول سری به دلیل یک مرتبه تفاضل‌گیری داده‌های اصلی و یک مرتبه تفاضل‌گیری فصلی میباشد.