

میانگین متحرک ساده

(هموار سازی داده ها)

تدوین: مرکز آمار خوارزمی

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

مقدمه:

هموار کردن داده ها به منظور کاهش نوسان موجود در آنها بکار می رود. در این قسمت چند شیوه هموارسازی مانند میانگین متحرک ساده و هموارسازی نمایی وجود دارد. در مطالب بیان شده در پایین به بررسی روش میانگین متحرک ساده می پردازیم.

میانگین متحرک ساده

میانگین متحرک یک شیوه هموار سازی است. این روش با متوسط گیری از مشاهدات متوالی در هر لحظه، داده ها را هموار می کند. به این معنی که noise یا اختلال در مشاهدات را کاهش می دهد. میانگین های متحرک برای جستجوی الگوی داده ای زمانی که سیگنال اساسی به وسیله سرو صدا (noise) پنهان مانده است، مفیدند.

در یک فرآیند ثابت به شکل $x_t = b + \varepsilon_t$ که $\{\varepsilon_t\}$ متغیرهای تصادفی غیرهمبسته با میانگین صفر و واریانس σ_ε^2 می باشند. میانگین متحرک معادل تجزیه و تحلیل n مشاهده اخیر است. b یک پارامتر نامعلوم است که باید برآورد شود. اگر تمامی مشاهدات در برآورد b از اهمیت یکسانی برخوردار باشند، در این صورت میانگین نمونه ای سری برآورد مناسبی برای b خواهد بود.

$$\hat{b} = \frac{1}{n} \sum_{t=1}^n x_t$$

میانگین متحرک ساده n دوره ای (n دوره ای) به شکل زیر تعریف می شود:

$$m_t = \frac{x_t + x_{t-1} + x_{t-2} + \dots + x_{t-n+1}}{n}$$

همچنین می توان نوشت:

$$m_t = m_{t-1} + \frac{x_t - x_{t-n}}{n}$$

n دامنه میانگین متحرک است. دامنه تعداد مشاهدات گذشته که از آنها میانگین گرفته می شود. همانطور که ملاحظه می شود m_t در واقع میانگین n مشاهده اخیر است. در هر دوره قدیمی ترین مشاهده حذف شده و جدیدترین اضافه می شود.

به عنوان مثال چنانچه از میانگین متحرک سه مقداری استفاده کنیم دنباله مشاهدات واقعی با دنباله زیر جایگزین خواهد شد.

$$\frac{x_1+x_2+x_3}{3}, \frac{x_2+x_3+x_4}{3}, \frac{x_3+x_4+x_5}{3}, \frac{x_4+x_5+x_6}{3}, \dots$$

سری جدید نوسان کمتری خواهد داشت. اما در عوض برخی از مشاهدات از دست خواهند رفت.

در انتهای دوره t ، پیش بینی برای دوره $t+q$ آینده بصورت $\hat{x}_{t+q}(t) = m_t$ خواهد بود.

✓ خواص روش میانگین متحرک ساده بستگی به دامنه آن دارد. دامنه بزرگ باعث می شود که میانگین متحرک در برابر تغییرات در پارامتر b به کندی واکنش نشان دهند و دامنه کوچک نتیجه معکوس می دهد.

اگر فرایند سری های زمانی دارای ارزش $b = b_1$ باشد و ناگهان به ارزش جدید $b = b_2$ منتقل شود، میانگین متحرک n دوره ای، وقت لازم را خواهد داشت تا پیش بینی هایی تولید کند که با مقدار جدید b سازگار باشند.

اگر خطاهای تصادفی متغیرهای مستقل باشند، واریانس m_t بصورت $\frac{\sigma_\varepsilon^2}{n}$ می باشد. در نتیجه برای n های کوچک، واریانس نسبتاً بزرگ خواهد بود.

بنابراین زمانی که فرآیند واقعا ثابت است تمایل داریم از n بزرگ استفاده کنیم تا بتوان مقدار b را دقیق تر تخمین زد و موقعی که فرآیند در حال تغییر است تمایل داریم از n های کوچک استفاده کنیم به نحوی که در مقابل تغییرات b جواب دهنده باشد. در یک فرآیند دارای روند خطی به صورت $x_t = b_1 + b_2 t + \varepsilon_t$ نیز می توان از روش میانگین های متحرک برای هموار کردن داده ها استفاده کرد. در این حالت از میانگین متحرک دوگانه استفاده می شود.

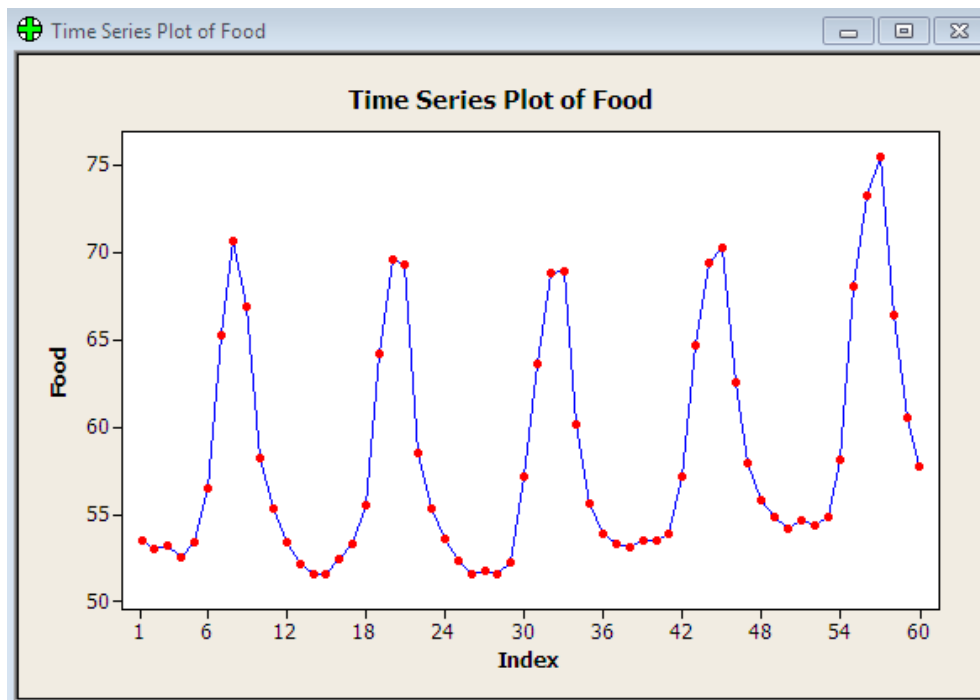
روش های میانگین متحرک که تاکنون بررسی کرده ایم در رده هموار کننده های داده های خطی یا فیلترهای خطی قرار دارند. یعنی آنها در مشاهده، x_t را با یک میانگین (احتمالا وزنی) از نقاطی که از لحاظ زمان نزدیک آن هستند جایگزین می کنند.

در ادامه برای بیان چگونگی انجام فرایند در نرم افزار Minitab مثالی بیان شده است.

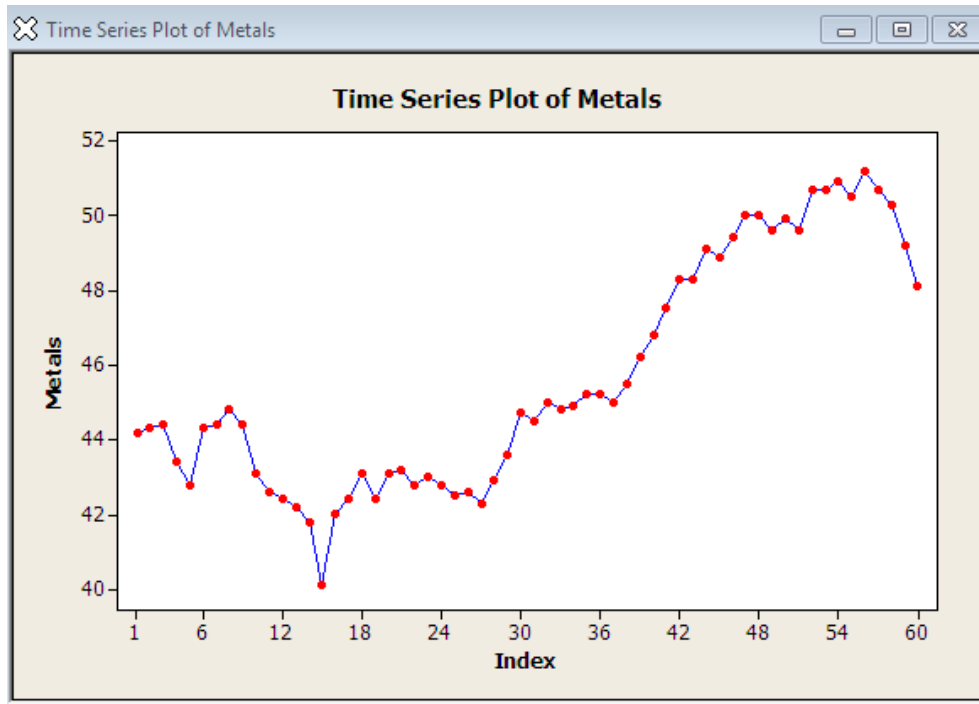
مثال:

فایل EMPLOY.MTW را در نظر بگیرید. این فایل شامل سه متغیر Trade , Food و Metal است. متغیر Trade دارای روند و عامل فصلی می باشد. برای مشاهده ی روند هر یک از متغیرهای نمودار سری زمانی هر کدام را رسم می کنیم.

ok > تمام متغیرها را به کادر روبرو منتقل کنید. > Graph > Time Series Plots > Simple > ok



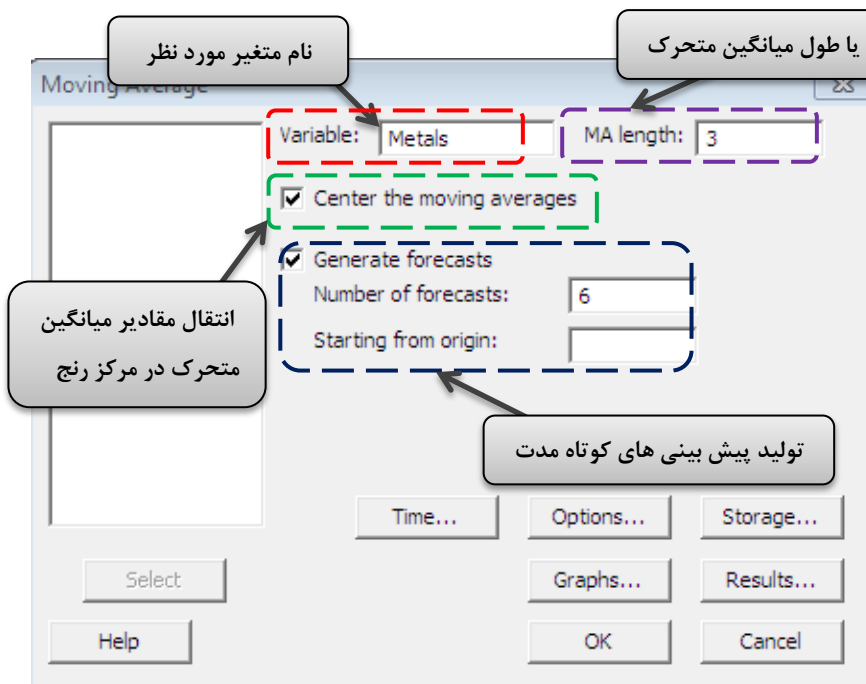
تابع سری زمانی food



تابع سری زمانی metals

همان طور که در نمودار های بالا مشاهده می شود متغیر Food دارای عامل فصلی واضحی می باشد. اما متغیر Metals فاقد روند یا الگوی فصلی مشخصی می باشد.

به طور کلی از روش میانگین متحرک برای تولید پیش بینی های کوتاه مدت در صورتی که داده ها فاقد روند یا الگوی فصلی مشخصی باشند استفاده می شود. بنابراین می توان از این روش برای پیش بینی های کوتاه مدت سری Metals استفاده کرد.



در Minitab برای اجرای این روش از منوی Time Series گزینه Moving Average را انتخاب می کنیم. مسیر زیر را طی نمایید.

Stat > Time Series > Moving Average

با انتخاب این گزینه پنجره روبرو باز می شود

در ادامه به شرح هر یک از گزینه های قرار داده شده در این پنجره را می پردازیم.

گزینه Variable

در کادر مقابل این گزینه باید نام متغیر مورد نظر را که در اینجا متغیر Metals می باشد، وارد کرد.

گزینه MA Length

در این قسمت باید دامنه یا طول میانگین متحرک را مشخص کرد. برای یک سری زمانی غیر فصلی معمولاً از یک میانگین متحرک کوتاه استفاده می شود. ما در این مثال طول میانگین متحرک را ۳ در نظر می گیریم. یک میانگین متحرک طولانی تر (با دامنه بزرگتر) قادر است Noise بیشتری را از مشاهدات حذف کند. اما در عوض حساسیت آن نسبت به تغییرات سری کمتر خواهد بود.

در مورد سری های فصلی معمولاً طول میانگین متحرک را برابر با طول یک دوره سالیانه در نظر می گیرند.

گزینه Center The Moving Averages

این گزینه مقادیر میانگین متحرک را در مرکز رنج قرار می دهد. در حالت پیش فرض این مقادیر در انتهای رنج قرار می گیرند.

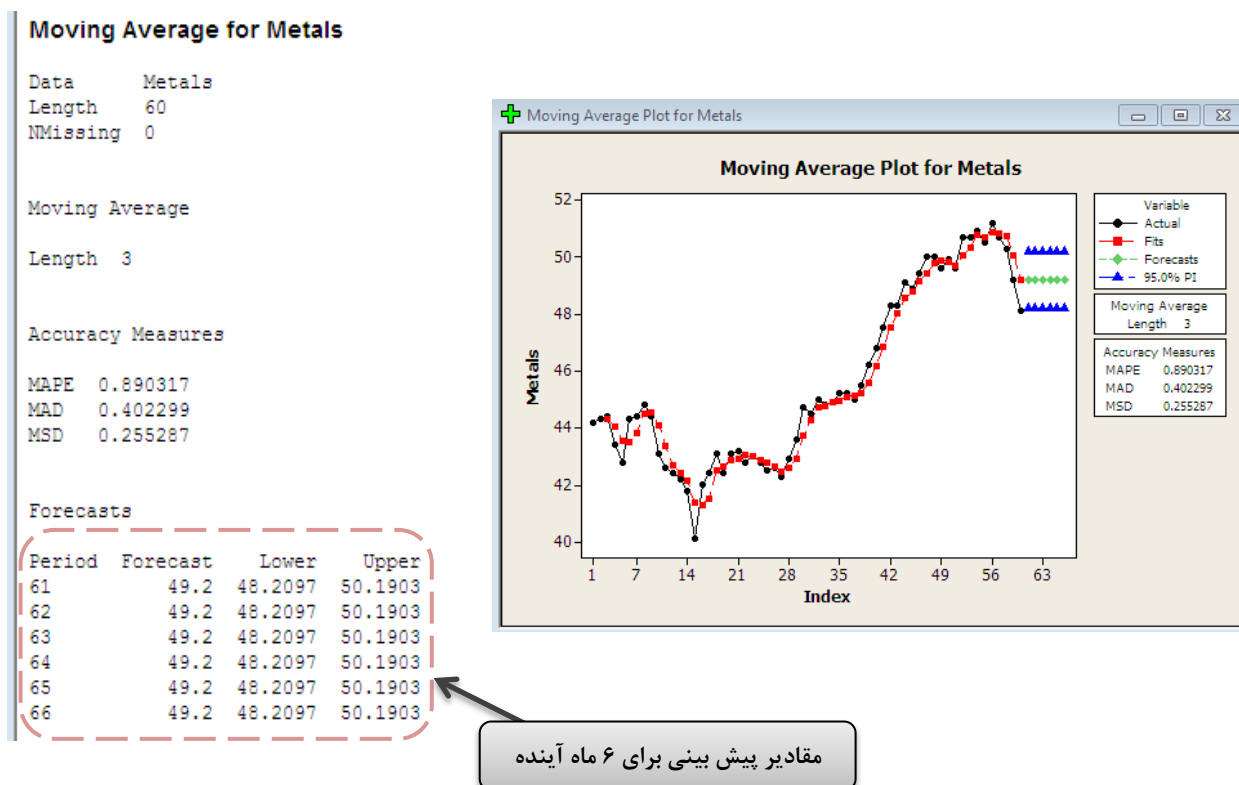
اگر دامنه میانگین متحرک یک عدد فرد مانند ۳ باشد، اولین مقدار محاسبه شده در دوره ی دوم قرار می گیرد و دومین مقدار در دوره سوم و الی آخر. در این حالت مقدار اولین و آخرین دوره Missing خواهد بود.

اگر دامنه میانگین متحرک یک عدد زوج مانند ۴ باشد، Minitab میانگین متحرک اولین چهار مقدار را محاسبه کرده و آن را MA1 می نامد. سپس میانگین متحرک برای دومین چهار مقدار را محاسبه کرده و آن را MA2 می نامد. حال در ستون سوم از MA1 و MA2 میانگین گرفته و آن را در دوره ی سوم قرار می دهد و همین روال را برای سایر مشاهدات ادامه می دهد. در این حالت مقدار میانگین متحرک برای اولین دو دوره و آخرین دو دوره Missing خواهد بود.

گزینه Generate Forecasts

معمولا برای تولید پیش بینی های کوتاه مدت از روش های هموار سازی مانند میانگین متحرک استفاده می شود.

در اینجا ما با مارک دار کردن این گزینه میخوایم مقدار سری را برای ۶ ماه آینده پیش بینی نمایم. بنابراین در کادر مقابل Number Of Forecasts عدد ۶ را وارد می کنیم. با تکمیل پنجره فوق در پنجره Session به صورت زیر خواهد بود.



همان طور که ملاحظه می شود مقادیر پیش بینی برای ۶ ماه آینده و حدود اطمینان ۹۵ درصد برای این پیش بینی ها در پنجره Session و در نمودار مربوطه داده شده است.

منبع:

- تجزیه و تحلیل سری های زمانی با نرم افزار MINITAB14، مصطفی خرمی- دکتر ابوالقاسم بزرگ نیا، انتشارات سخن گستر، ۱۳۸۶