Nonlinear Regression

این نوع از رگرسیون، روشی برای یافتن مدلی غیر خطی در جهت یافتن رابطه میان متغیر وابسته و مجموعه ای از متغیرهای مستقل است.بر خلاف شیوه قدیمی رگرسیون خطی که (محاسبه مدل خطی را محدود می کرد)رگرسیون غیر خطی می تواند روابط مدل را بصورت اختیاری و ارادی میان متغیر های مستقل و غیر مستقل بررسی و اندازه گیری کند.

استفاده ازآن بر اساس الگوریتم تخمین و براورد کامل شده است.قابل توجه این مطلب که این روش برای مدلهای چند فرمولی(مختلط) که می تواند توسط شیوه قدیمی رگرسیون خطی محاسبه شود انجام نمی گیرد.

مثال:

آیا می توان جمعیت را براساس زمان پیش بینی کرد؟یک نمودار پراکندگی نشان می دهد که رابطه ای قوی میان جمعیت و زمان به نظر می رسد،اما رابطه از نوع غیر خطی است،بنابراین نیاز به استفاده از روش تخمین غیر خطی است.ازطریق تنظیم یک معادله مناسب در رابطه با مدل رشد منطقی جمعیت،می توانیم برآورد خوبی از این مدل داشته باشیم که به ما اجازه می دهد جمعیت را برای زمان مورد نظر خود پیش بینی کنیم.

پارامترهای رگرسیون غیر خطی:

پارامترها درواقع قسمتی از مدل هستند که رگرسیون غیر خطی جهت رسیدن به آیتم های مورد نظر برآورده می کند.این پارامترها می توانند اعدادی با ارزشهای ثابت ریاضی،ضریب یا عامل مشترک افزاینده،تعریف کننده،و یا ارزشهای مورد استفاده شده در ساختار ارزیابی باشند.تمام پارامترهایی که تعریف میکنید با ارزش اصلی آنها برروی پارامتر لیست، parameter list در maindialog box نمایش داده می شود.

در ابتدا باید نامی برای هر پارامتر مشخص شود.این نام باید اسم یک متغیر قابل قبول باشد و باید در بیان و متن مدل در maildialog box وجود داشته باشد.

Starting value به شما اجازه می دهد برای پارامتر، یک ارزش اولیه بصورت بهتر و تا حد امکان نزدیک به راه حل و پاسخ مورد انتظار تعیین کنید . نکته قابل ذکر اینکه ،ارزشهای اولیه کم و یا با مقدار پایین می تواند در نتیجه منجر به از بین رفتن راه حلهای منطقی شود و دستیابی به راه حل را بصورت فیزیکی نیز غیر ممکن سازد..

استفاده از starting value

در قسمتmaindialog box می توانید این تنظیمات را برای دستیابی به ارزشهای اولیه و اصلی پارامترها از ارزشهای پیشین و اولیه آنها داشته باشید.این امکان به شما اجازه می دهد زمانیکه تشابه الگوریتم بسیار کم و پایین است به کار خود ادامه بدهید.

توجه:

درصورتیکه شما مدل خود را تغییر دادید حتما تمام موارد انتخابی را از حالت انتخاب خارج کنید زیرا این تغییرات در dialog box باقی می ماند.

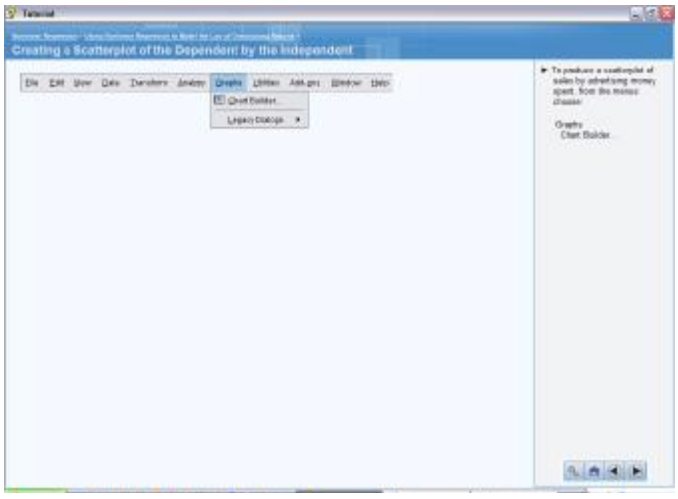
*

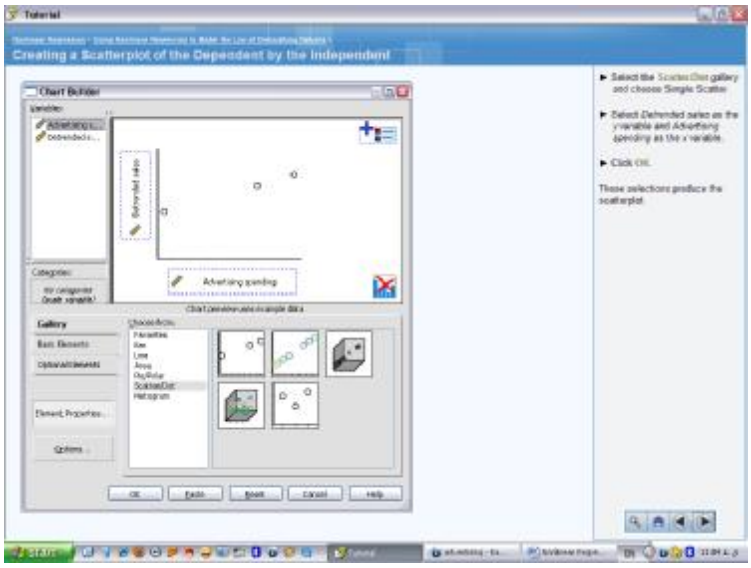
مثال:

یک خرده فروش میخواهد روابط میان پول خرج شده(هزینه)آگهی های بازرگانی و نتایج فروش را مورد بررسی و مطالعه قرار دهد.در آخرارقام فروش پیشین و هزینه (خرج کرد تبلیغات)را به هم مرتبط می کنند.این اطلاعات و داده ها را می توان در قسمت advert.sav مشاهده نمود.این فایل داده همانند آنالیزهای پیشین مدل خطی و Quadratic (وابسته)از طریق شیوه تخمین منحنی نمودار curve estimation و مدل Quadratic و یافته های آن عمل می کند،لذا نسبت به مدل خطی در این حالت برتری دارد،اگرچه خرده فروش علاقه مند است که مدل Quadratic شاید نتواند مدل مناسبی باشد و دلیلش اینست که فرض این مدل براینست که افزایش تبلیغات در نهایت منجر به کاهش میزان فروش می شود به همین دلیل معتقد است استفاده از مدل رگرسیون غیر خطی می تواند مناسب تر باشد .

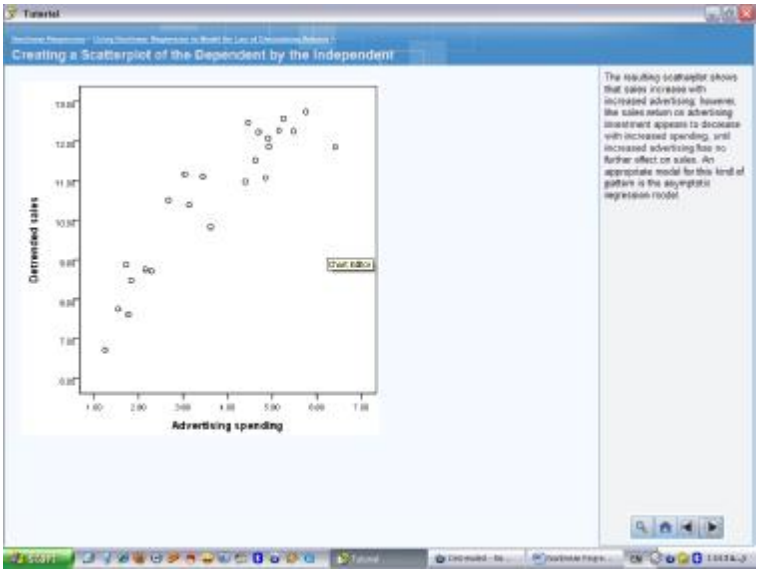
نحوه بوجود آوردن نمودار پراکندگی(Scatterplot) تابع وابسته بصورت مستقل(جدا)

برای ایجاد نمودار پراکندگی فروش از طریق هزینه صرف شده برای آگهی های بازرگانی از مسیر زیر Graphs و Chart Builder... را انتخاب می کنیم.





پس از باز شدن پنجره بالا به قسمت gallery رفته و از میان گزینه ها Simple Scatter را انتخاب میکنیم.فروش Detrended را بعنوان متغیری در محور y و هزینه صرف شده برای تبلیغات را بعنوان متغیری دیگر درمحور x وارد میکنیم.پس از زدن کلید ok نمودار پراکندگی نمایش داده می شود.



نتایج نمودار پراکندگی حاکی از آنست که میزان فروش با افزایش تبلیغات بالا می رود، همچنین میزان فروش بستگی به سرمایه گذاری در تبلیغات دارد که مشخص کننده اینست که با افزایش مخارج ،فروش کاهش می یابد تاجاییکه افزایش تبلیغات هیچ تاثیر مثبتی بر روی فروش ندارد.دراین حالت مدلی که برای این قبیل طرحها مناسب باشد ،مدل رگرسیون asymptotic است.

مدل رگرسیون asymptotic برابر با y=b1+b2eb3x می باشد که برای آنها شروطی تعریف شده که عبارتست از b1>0, b2<0 و b3<0 باشد،که نتایج مدل Mistcherlich را برای قوانین بازده نزولی ارائه می کند.ابتداعا این مدل به سرعت افزایشی همراه با کاهش مقدار x خواهد داشت اما بعد از آن میزان سود کاهش می یابد و در آخر بصورت تدریجی در زیر ارزش b1 متوقف می شود .

(نحوه) انتخاب ارزشها

در روش رگرسیون غیر خطی نیاز است که ارزشهای ابتداعی و اولیه را برای مقادیر پارامترها در مدل در دست داشته باشید،در ابتدا به نظر می رسد وحشت یاترسی در پس زمینه در مورد استفاده از مدل است،اما پس از آشنایی با آن آسانتر می شود.

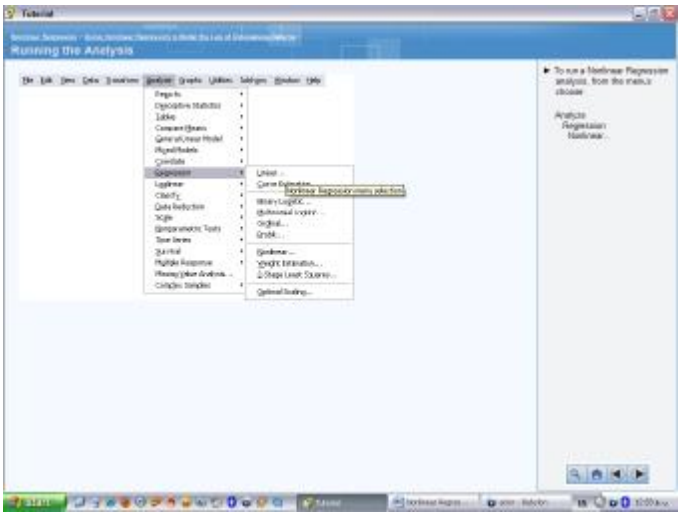
B1 معرف حد بالاتر میزان asymptotic برای فروش است. با مشاهده نمودار ،حتی ارزش فروشها در سطح بالا به میزان کمتر از 13 تنزل می یابد که یک شروع منطقی برای ارزش است.

B2 نشاندهنده تفاوت میان ارزش y زمانیکه x=0 و میزان آن بالاتر از asymptotic است میباشد. ارزش ابتدایی منطقی حداقل میزان y منهای b1 است.با مشاهده نمودار می توان اینطور گفت 6-=13-7.

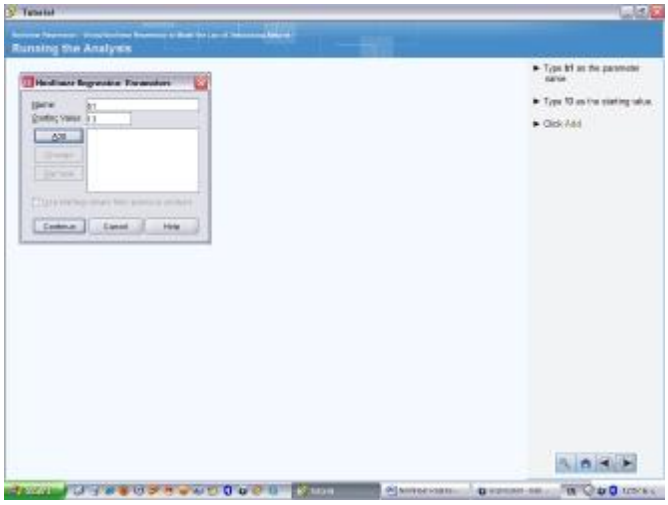
امکان غیر پارا متری بودن مدل از طریق جایگزینی b2با,(b4-b1) میباشد، و آن زمانی است کهb4 معرف فروش است و هیچ هزینه ای خرج تبلیغات نمی شود.ارزش b4 قابلیت تفسیر بیشترو راحت تری نسبت به b2 دارد،اما در پیشرفت و یا تغییر میزان احتمال ما تغییری ایجاد نخواهد کرد.

بطور کلیb3 در ابتدا دارای شیب مثبتی بین دو" جدایی گزینی خوب" از نقاط بر رو ی نمودا میباشد.با مشاهده نمودار نقاط کمی در پیرامون x=2 ،y=8 و پیرامون x=5 و y=12 مشاهده میشود.(شیب)این نقاط برابر با 1.33=(5-2)/(12-8) میباشد بنابراین در ابتدا احتمال وقوع یک ناهنجاری برای b3 که برابر با 1.33- است را خواهیم داشت.

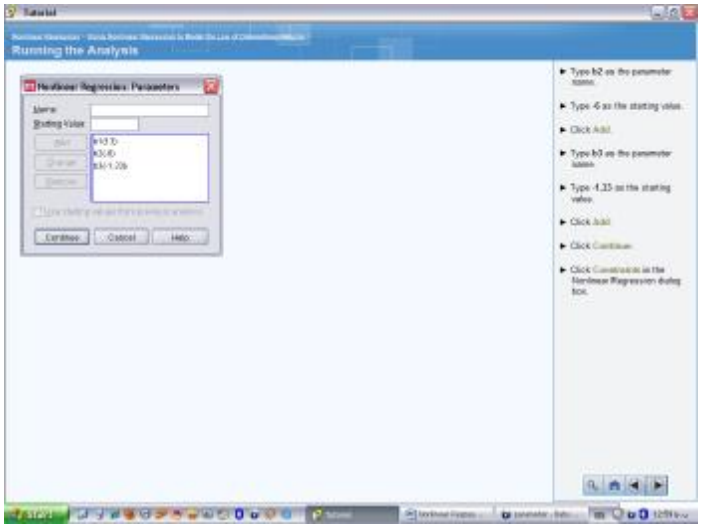
مسیرزیر را برای باز کردن این مدل طی می کنیم:



فروش Detrended را در قسمت متغیر وابسته وارد می کنیم و در قسمت پایین آن عبارت b1 + b2*exp(b3*advert) را بعنوان حالت مدل (نام)تایپ می کنیم، بر روی تیک Parameters.... کلیک می کنیم تا پنجره زیر نمایان شود.

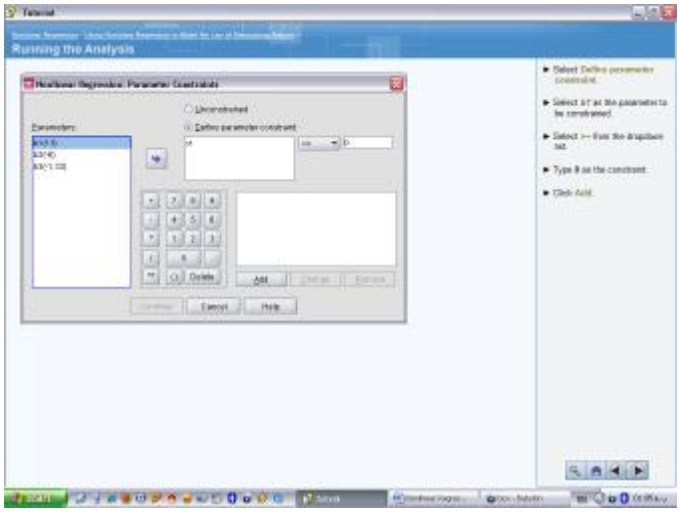


در قسمتname نام پارامتر مورد نظر را وارد می کنیم،در قسمتstarting value عدد 13 را بعنوان ارزش ابتدایی وارد می کنیم سپس کلید add را میزنیم.(منظور از عدد 13 میزان ماکسیموم و یا آستانه میزان فروش بود که در در scster plot نیز نمایش داده شد.



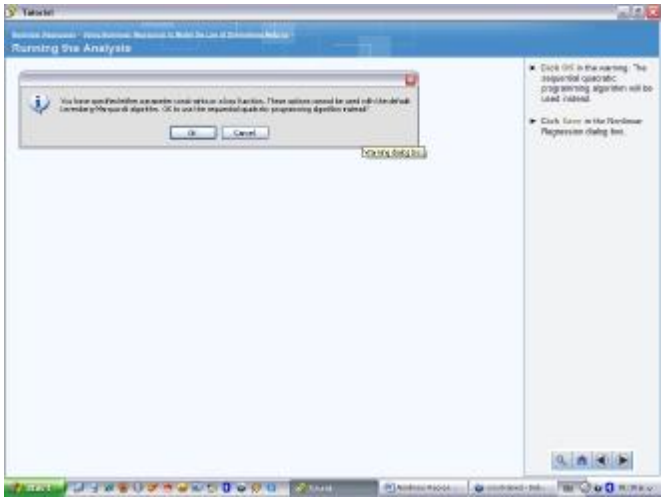
b2 را نیز بعنوان نام متغیر وارد میکنیم،عدد 6- را بعنوان ارزش ابتدایی پارامتر 6 وارد کرده و add میکنیم.همچنین b3 را نیز بعنوان نام پارامتر سوم انتخاب و به ارزش ابتدایی آن 1.33- می دهیم.کلید add و پس از آن Continue را میزنیم.

برروی محدودیت ها یا(Constraints) در قسمت رگرسیون غیر خطی کلیک میکنیم.

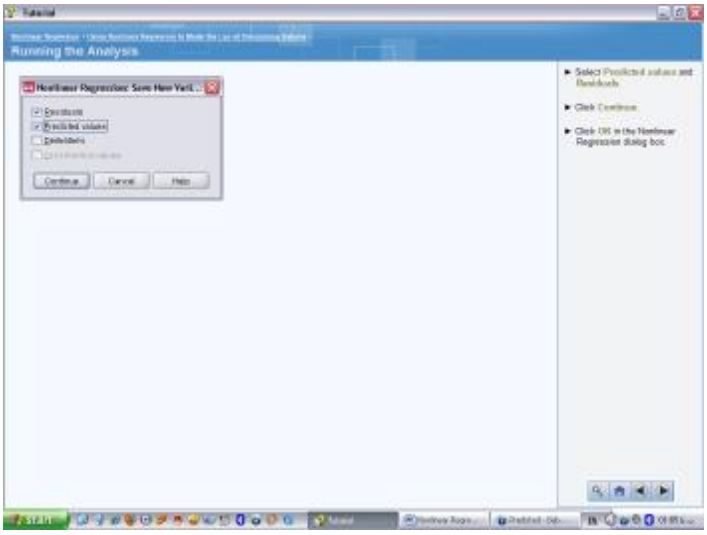


تیک Define parameter constraint را زده و b1 را بعنوان پارامتری تحمیلی یا اجباری وارد میکنیم،تیک =< را زده و عدد 0 را بعنوان محدودیت یا constraint تعریف می کنیم و آنرا add میکنیم.

مراحل فوق را برای b2 و b3 نیز انجام میدهیم و در انتها تیک Continue را می زنیم.

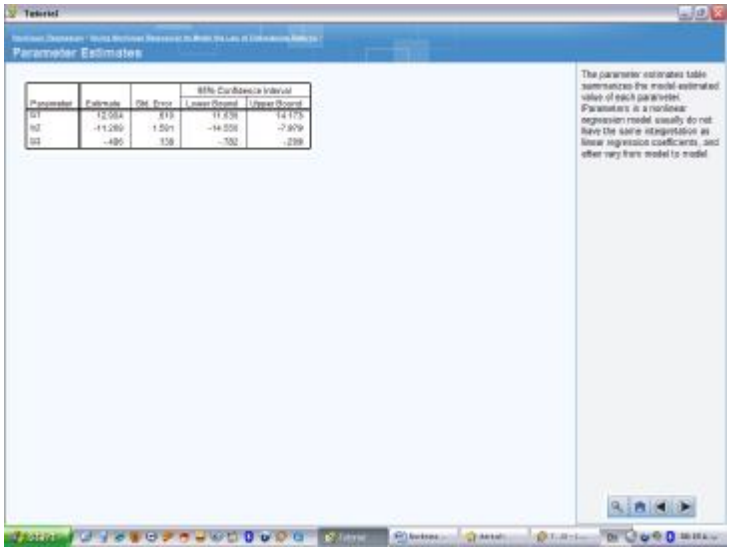


پس از این مرحله پیام اخطاری نمایش داده می شود که okرا انتخاب می کنیم.حال منشور ترتیبی(sequential quadratic) جایگزین الگوریتم و تعریف ما شده است.سپس کلید Save in the Nonlinear Regression dialog box را می زنیم.



تیک ارزش پیشنهادیPredicted values و .Residuals را میزنیم Continue و ok را میزنیم.

تحلیل:

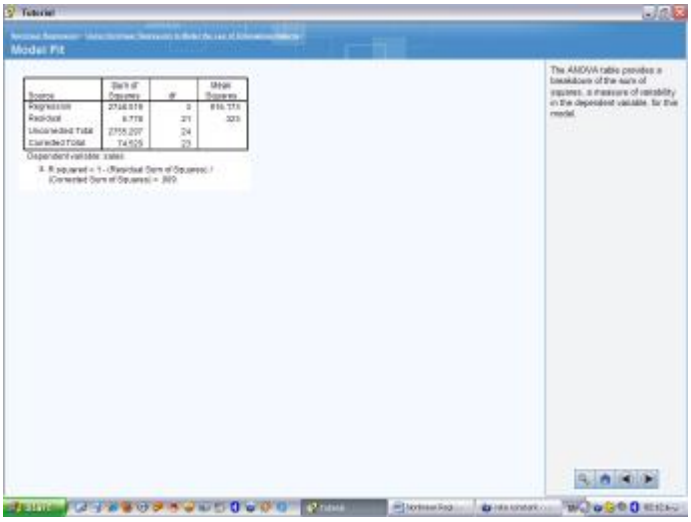


جدول پیش بینی احتمالات بصورت خلاصه نتیجه پیش بینی احتمال مدل را برای هر کدام از پارامترهای وارد شده نشان می دهد.پارامترها در مدل رگرسیون غیر خطی معمولا همسان با ضریب رگرسیون خطی نبوده و مکررا از مدلی به مدل دیگر تغییر می کنند.

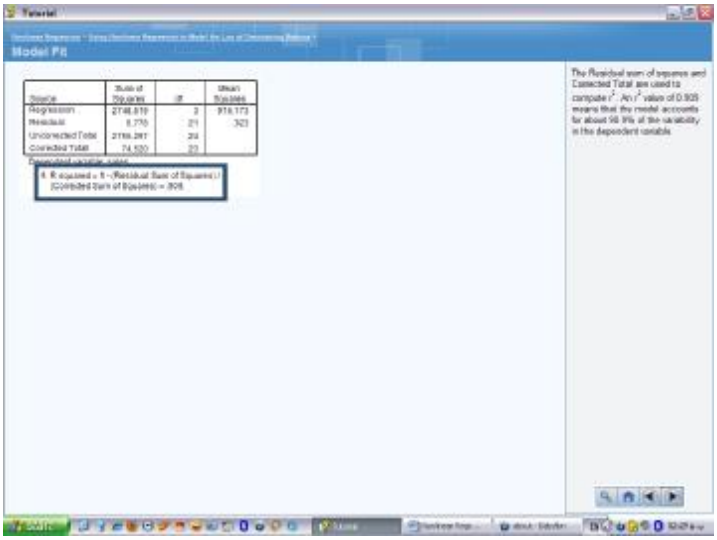
همانطور که قبلا گفته شدb1 معرف حداکثر احتمال فروش است حتی ازمانیکه پول یا هزینه موجود برای تبلیغات نامحدود باشد. خطای استانداردی با رعایت ارزش احتمال پیشنهادی وجود دارد که می توان در نتیجه آن می توان به نتیجه ارزیابی اطمینان کرد.

B2 اختلاف میان حداکثر فروش زمانیکه هیچ پولی برای تبلیغات هزینه نشده باشد است، در آن میزان خطای استاندارد بالا و ضریب اطمینان با گستره بالایی در مقایسه با ارزش احتمالی وجود دارد،بنا بر این خطای تخمینی در اینجا وجود دارد.

B3 میزان نرخ کنترل شده زمانی است که به میزان حداکثر دست می یابیم که اصطلاحا آنرا نرخ ثابت یا rate constant می نامیم.همانند b2 اینجا نیز مقداری خطای تخمینی وجود دارد.

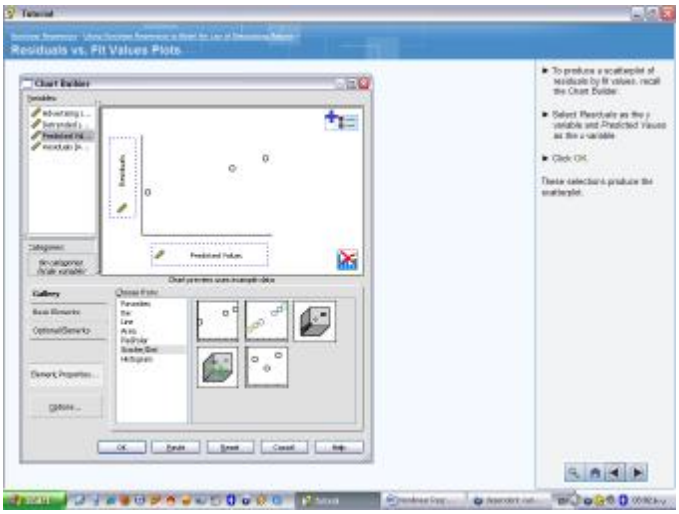


در این مدل جدول ANOVA مجموعه ای از مجذورها را به همراه سنجش تغییر پذیری متغیر تابع(وابسته) نشان می دهد.در سطر مربوط به رگرسیون اطلاعات نوسانات و یا پراکندگی حسابها در مدل نشان داده می شود.سطر مربوط به residual یا"باقیمانده" ، اطلاعات ناشی از نوسان ویا تغییراتی که توسط مدل محاسبه نشدند را نشان می دهد.سطر سوم Uncorrected Total نشاندهنده تغییر پذیری هماهنگ و یکپارچه است زمانیکه مجموع صحیح تنها بر روی تغییر پذیری در حدود فروش میانگین تنظیم شده باشد.

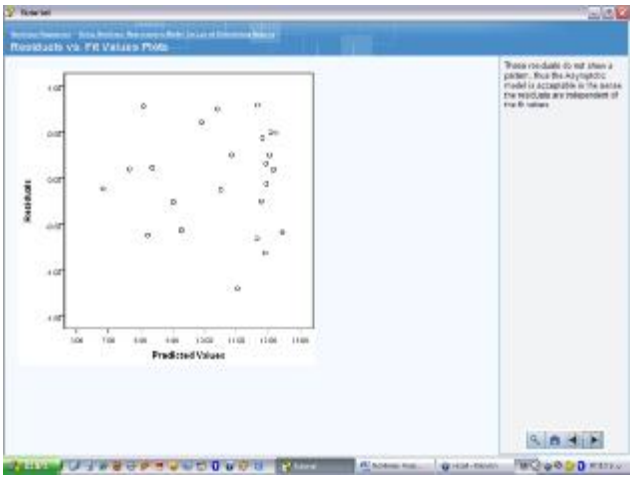


مجموع مجذورات باقیمانده ها(انحرافات)و مجموع تصحیح شده برای محاسبه r2 مورد استفاده قرار گرفته و ارزش r2 که برابر با 0.909 شده است نشان دهنده اینست که حسابرسی در این مدل با صحتی در حدود 90.9% از ضریب تغییر پذیری متغیر وابسته انجام گرفته است.

برای تولید یک نمودار پراکنش residual توسط ارزشهای مناسب قسمت chartbuilder را مجددا باز می کنیم.reduals را بعنوان متغیر y وpredicted value یا متغیرهای پیش بینی را بعنوان متغیر x وارد کرده سپس ok می کنیم



آیتم های انتخاب شده نمودار پراکنش را نشان می دهد.



residual ها ی بالا یک طرح و الگوی مشخصی را نشان نمی دهند،بنابراین مدل یعنی مدل Asymptotic برای تفهیم residual که وابسته به ارزشهاست مورد قبول قرار می گیرد..

خلاصه بحث:

در تحلیل قبلی روش cure estimationمورد استفاده قرار گرفت، مشاهده شد که مدل quadratic با استفاده از روش رگرسیون غیر خطی یک مدل رگرسیون asymptotic را برای ایجاد رابطه میان فروش و تبلیغات

ایجاد نمود.زمان بیشتری برای تفهیم و تصریح یک رگرسیون غیر خطی لازم است،اما نتایج مدل نشاندهنده انتخاب مدلی مناسب و درخور و بسیار خوب با میزان residual خوب و پارامترهای قابل تفسیر و قابل قبولی است.