

با سلام و احترام،

## ادامه فصل بیستم و چهارم: رگرسیون خطی (Linear Regression)

### استفاده از رگرسیون خطی برای مدل فروش کامیون

یک گروه صنعتی خودروسازی قصد دارد میزان فروش کامیون‌های خود را بررسی نماید. برای این کار بررسی ارتباط بین فروش کامیون‌ها و مشخصات آنها لازم می‌باشد. از رگرسیون خطی برای شناسایی مدل‌هایی که خوب به فروش نمی‌رسند استفاده نمایید.

### آغاز تحلیل

۱. برای اجرای یک رگرسیون خطی، مسیر **Analyze > Regression > Linear** را از منوی اصلی برگزینید.

۲. **Log – transformed Sales** را به عنوان متغیر وابسته برگزینید.

توزیع **Log – transformed Sales** نسبت به **Sales in thousands** به حالت نرمال نزدیک‌تر است و مدل رگرسیون خطی با متغیرهای نرمال بهتر کار می‌کند.

۳. گزینه‌های **Vehicle type** تا **Fuel efficiency** را به عنوان متغیرهای مستقل انتخاب نمایید.

۴. دکمه **Statistics** را کلیک کنید.

۵. گزینه‌های **Part and Partial Correlations** و **Collinearity diagnostics** را فعال نمایید.

۶. دکمه **Continue** را کلیک کنید.

۷. **Ok** را در کادر محاوره **Linear Regression** کلیک کنید.

### بررسی همواری مدل

جدول **ANOVA** یک آماره **F** معنی‌دار را شرح می‌دهد، و نشان می‌دهد که استفاده از مدل بهتر از تخمین زدن میانگین است.

به طور کلی، رگرسیون فروش‌ها را به طور مناسب مدل‌سازی می‌کند. تقریباً نیمی از تغییر در فروش‌ها با مدل بیان می‌شود.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	130.300	10	13.030	13.305	.000 <sup>a</sup>
	Residual	138.082	141	979		
	Total	268.383	151			

## ضرایب

با این که همواری مدل مثبت به نظر می‌رسد، بخش اول جدول ضرایب نشان می‌دهد که پیشگوهای فراوانی در مدل وجود دارند. چندین ضریب بی‌معنی وجود دارد که نشان می‌دهد این متغیرها خیلی به مدل کمک نمی‌کنند (a).

برای تعیین پیشگوهای معنی‌دار نسبتاً مهم، به ضرایب استاندارد شده نگاه کنید. حتی اگر **Price in thousands** یک ضریب کوچک در مقایسه با **Vehicle type** دارد، **Price in thousands** در واقع به مدل کمک بیشتری می‌کند چون یک مقدار مطلق ضریب استاندارد شده بزرگتری دارد (b).

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.017	2.741		-1.101	.273
	Vehicle type	.883	.331	.293	2.670	.008
	Price in thousands	-4.64E-02	.013	-.502	-3.596	.000
	Engine size	.358	.190	.281	1.871	.063
	Horsepower	-2.15E-03	.004	-.092	-.509	.611
	Wheelbase	4.162E-02	.023	.241	1.785	.076
	Width	-2.81E-02	.042	-.073	-.676	.500
	Length	1.460E-02	.014	.148	1.032	.304
	Curb weight	.156	.350	.075	.447	.655
	Fuel capacity	-5.67E-02	.047	-.167	-1.203	.231
	Fuel efficiency	8.122E-02	.040	.262	2.023	.045

در بخش دوم جدول سه نوع همبستگی به شرح زیر وجود دارد:

**Zero-order** (همبستگی مرتبه صفر): این همبستگی که از نظر مقداری برابر ضریب همبستگی پیرسون می‌باشد، میزان همبستگی بین متغیرها را با حذف اثر متغیر کنترل نشان می‌دهد.

**Partial** (همبستگی جزئی یا تفکیکی): به کمک این ضریب همبستگی می‌توان اثر یک یا چند متغیر را کنترل نمود. این ضریب، میزان همبستگی بین یک متغیر مستقل با متغیر وابسته را، پس از حذف میزان همبستگی این دو متغیر با سایر متغیرهای مستقل نشان می‌دهد. هر چه مقدار این همبستگی برای یک متغیر بزرگتر باشد، نقش آن متغیر در مدل بیشتر است.

**Part** (همبستگی نیمه جزئی یا نیمه تفکیکی): این ضریب نیز نوعی کنترل آماری است که به کمک آن می‌توان اثر یک یا چند متغیر را کنترل نمود. این ضریب میزان همبستگی بین یک متغیر مستقل با متغیر وابسته را، پس از حذف اثر خطی سایر متغیرهای مستقل بر متغیر مستقل مورد نظر (و نه متغیر وابسته) نشان می‌دهد. در همبستگی نیمه جزئی، هر چه میزان این همبستگی برای یک متغیر بیشتر باشد، نقش آن متغیر در مدل بیشتر است. علاوه بر اینکه در این نوع همبستگی، چون اثر خطی متغیرهای مستقل بر همدیگر حذف می‌شود، بنابراین میزان آن نسبت به دو همبستگی قبلی (مرتبه صفر و تفکیکی) پایین‌تر است.

قسمت دوم جدول ضرائب نشان می‌دهد که ممکن است با هم خطی چندگانه مشکلی وجود داشته باشد. برای بیشتر پیشگوها، مقادیر همبستگی‌های **Partial** و **Part** به طور ناگهانی از همبستگی مرتبه صفر کاهش می‌یابند.

**(a).**

تلرانس درصد واریانس، در یک پیشگو می‌باشد که نمی‌توان با پیشگوهای دیگر آن را توضیح داد. بنابراین، تلرانس‌های کوچک که **90% - 70%** واریانس را در یک پیشگو نشان می‌دهد می‌تواند با پیشگوهای دیگر معلوم شود **(b)**. وقتی تلرانس به **0** نزدیک شود، هم خطی چندگانه بزرگی وجود دارد و خطای استاندارد ضرائب رگرسیون بزرگ خواهند شد. یک عامل افزایشی واریانس بزرگتر از **2** معمولاً مسئله‌ساز می‌باشد، در حالی که کوچک‌ترین **VIF** در جدول معادل **3.193** است **(c)**.





Model		Correlations			Collinearity Statistics	
		Zero-order	Partial	Part	Tolerance	VIF
1	Vehicle type	.274	.219	.161	.304	3.293
	a Price in thousands	-.552	-.290	-.217	.187	5.337
	Engine size	-.135	.156	.113	.162	6.159
	Horsepower	-.389	-.043	-.031	.112	8.896
	Wheelbase	.292	.149	.108	.200	4.997
	Width	.037	-.057	-.041	.313	3.193
	Length	.215	.087	.062	.179	5.605
	Curb weight	-.041	.038	.027	.131	7.644
	Fuel capacity	-.010	-.101	-.073	.189	5.303
	Fuel efficiency	.121	.168	.112	.217	4.604

## هم خطی بودن

یکی از فرضیات مهم بیشتر آزمون‌ها به خصوص در آزمون‌های مربوط به فرضیه‌های علی این است که نباید بین متغیرها رابطه هم خطی وجود داشته باشد. یعنی هیچ یک از متغیرهای مستقل نباید رابطه خطی با همدیگر داشته باشند. رابطه هم خطی وضعیتی است که نشان می‌دهد یک متغیر مستقل تابعی خطی از سایر متغیرهای مستقل است. اگر هم خطی در یک معادله رگرسیون بالا باشد، بدین معنی است که بین متغیرهای مستقل همبستگی بالایی وجود دارد و در چنین حالتی با وجود بالا بودن مربع  $R$ ، مدل اعتبار بالایی ندارد. به عبارت دیگر، با وجود آنکه مدل خوب به نظر می‌رسد ولی دارای متغیرهای مستقل معنی‌داری نمی‌باشد.

در این مثال چون گزینه **Collinearity diagnostics** در کادر محاوره رگرسیون فعال شده بنابراین در خروجی شاهد مقادیر هم خطی هستیم. دو پارامتر **Tolerance** (تولرانس) و **VIF** (عامل تورم واریانس) در جدول قابل مشاهده می‌باشد. منظور از تولرانس نسبتی از واریانس یک متغیر مستقل است که توسط سایر متغیرهای مستقل تبیین نشده است. مقدار آن بین صفر تا یک می‌باشد و نشان دهنده آن است که متغیرهای مستقل تا چه اندازه رابطه خطی با همدیگر دارند. پس هر چه مقدار تولرانس به یک نزدیکتر باشد میزان هم خطی کمتر است و بالعکس هر چه تولرانس به صفر نزدیکتر باشد نشان از میزان هم خطی بالا می‌دهد و خطای استاندارد ضرایب رگرسیون از تورم بالایی برخوردار است. در نتیجه موقع اجرای رگرسیون با مشکلاتی مواجه هستیم.

پارامتر دیگر **VIF** یا عامل تورم واریانس می‌باشد که حاصل تقسیم یک بر مقدار تولرانس است (معکوس تولرانس). هر چه این پارامتر از ۲ بزرگتر باشد، میزان هم خطی بیشتر است. نتیجه و چگونگی تفسیر عامل تورم واریانس، عکس تولرانس می‌باشد. یعنی هر چه این ضریب زیاد شود، واریانس ضرایب رگرسیونی افزایش می‌یابد و در

نتیجه مدل رگرسیون را برای پیش‌بینی نامناسب جلوه می‌دهد. بنابراین، هر چه مقدار عامل تورم واریانس برای یک متغیر مستقل بیشتر باشد نتیجه می‌گیریم که آن متغیر نقش زیادی در مدل، نسبت به بقیه متغیرها، ندارد.

هم‌خطی بودن تأیید می‌کند که مجموعه‌ای از مشکلات همراه با چند خطی چندگانه وجود دارد. چندین مقدار ویژه (Eigenvalue) نزدیک به 0، نشان می‌دهد که پیشگوها همبستگی داخلی بالایی دارند و اینکه تغییرات کوچک در مقادیر داده‌ها ممکن است به تغییرات بزرگی در برآورد ضرائب منتهی شود. Condition Index به صورت مربع ریشه نسبت‌های بزرگ‌ترین مقدار ویژه به مقدار ویژه متوالی محاسبه می‌شود. مقادیر بزرگتر از 15 نشان می‌دهد که احتمالاً مشکلی در هم‌خطی بودن وجود دارد؛ مقادیر بزرگتر از 30، نشان دهنده یک مجموعه از مشکلات هستند، شش تا از این شاخص‌ها بزرگتر از 30 هستند، در نتیجه مشکلات فراوانی با هم‌خطی بودن همراه است (a).



Model	Dimension	Eigenvalue	Condition Index
1	1	9.920	1.000
	2	.733	3.678
	3	.259	6.193
	4	5.025E-02	a 14.051
	5	1.944E-02	22.589
	6	7.679E-03	35.942
	7	5.061E-03	44.275
	8	2.901E-03	58.480
	9	1.710E-03	76.175
	10	5.803E-04	130.747
	11	4.513E-04	148.267

در ادامه تلاش می‌شود که مشکلات هم‌خطی بودن را با اجرای رگرسیون با استفاده از ZScoreهای متغیرهای وابسته و روش قدم به قدم (stepwise) انتخاب مدل حل گردد.

## اجرای رگرسیون خطی قدم به قدم

در این روش (stepwise)، متغیرها یک به یک وارد مدل می‌شوند. یعنی اول متغیری که بیشترین همبستگی را با متغیر وابسته دارد انتخاب می‌نماید. متغیر بعدی متغیری است که پس از تفکیک متغیر مقدم بر آن، موجب بیشترین افزایش در مقدار ضریب تعیین می‌شود. در این روش، ورود متغیرها به مدل را یک به یک و تا زمانی

انجام می‌دهیم که معنی‌داری متغیر به ۹۵ درصد برسد یعنی سطح خطا ۵٪ درصد گردد، سپس عملیات متوقف می‌شود.

۱. برای اجرای یک رگرسیون خطی قدم به قدم در متغیرهای استاندارد شده، کادر محاوره **Linear Regression** را فراخوانی نمایید.
۲. گزینه‌های **Vehicle type** تا **Fuel efficiency** را به عنوان متغیرهای مستقل از حالت انتخاب خارج سازید.
۳. گزینه **ZScore: type** تا **ZScore: Fuel efficiency** را به عنوان متغیرهای وابسته انتخاب نمایید.
۴. **Stepwise** را به عنوان روش وارد کردن (Method) انتخاب کنید.
۵. گزینه **Model** را به عنوان متغیر عنوان‌گذاری مورد (Case Labels) انتخاب نمایید.
۶. دکمه **Statistics** را کلیک کنید.
۷. گزینه **Part and Partial Correlations** را از حالت انتخاب خارج سازید.
۸. گزینه **Casewise diagnostics** را انتخاب نمایید و ۲ را در کادر متنی تایپ نمایید.
۹. دکمه **Continue** را کلیک کنید.
۱۰. دکمه **Plots** را در کادر محاوره **Linear Regression** کلیک کنید.
۱۱. **\*SDRESID** را به عنوان متغیر **y** و **\*ZPRED** را به عنوان متغیر **X** برگزینید.
۱۲. گزینه **Histogram** را فعال نمایید.
۱۳. دکمه **Continue** را کلیک کنید.
۱۴. دکمه **Save** را در کادر محاوره **Linear Regression** کلیک کنید.
۱۵. گزینه **Standardized** را در مجموعه **Predicted Values** فعال نمایید.
۱۶. گزینه **Standardized** را در مجموعه **Residuals** فعال کنید.
۱۷. گزینه‌های **Cook's** و **Leverage Values** را در مجموعه **Distances** فعال نمایید.
۱۸. دکمه **Continue** را کلیک کنید.
۱۹. **Ok** را در کادر محاوره **Linear Regression** کلیک کنید.

## تشخیص هم‌خطی بودن

هیچ مقدار ویژه‌ای در نزدیکی ۰ وجود ندارد، و همه شاخص‌های شرطی خیلی کوچکتر از ۱۵ هستند. در نتیجه، و مدل ساخته شده با روش‌های قدم به قدم مشکلی با هم خطی بودن ندارد.



## بررسی همواری مدل

مدل جدید قادر است که فروش‌ها را در قیاس با مدل قبلی تبیین نماید. به آماره‌های ضریب تعیینی تعدیل شده (**Adjusted R Square**) که تقریباً یکسان می‌باشند نگاه کنید. یک مدل با پیشگوهای اضافی همیشه ضریب تعیینی (**R-Square**) بزرگتر خواهد داشت. ضریب تعیینی تعدیل شده پیچیدگی مدل را اصلاح می‌کند.



Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.697 <sup>a</sup>	.486	.449	.98960

a. Predictors: (Constant), Fuel efficiency, Length, Price in thousands, Vehicle type, Width, Engine size, Fuel capacity, Wheelbase, Curb weight, Horsepower

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.552 <sup>a</sup>	.304	.300	1.11553
2	.655 <sup>b</sup>	.430	.422	1.01357

a. Predictors: (Constant), Zscore: Price in thousands  
b. Predictors: (Constant), Zscore: Price in thousands, Zscore: Wheelbase

## ضرائب قدم به قدم

الگوریتم قدم به قدم، قیمت و اندازه را (بر حسب فاصله دو محور چرخ کامیون) به عنوان پیشگوها انتخاب می‌کند. فروش‌ها به صورت منفی تحت تأثیر قیمت و به صورت مثبت تحت تأثیر اندازه می‌باشد؛ نتیجه این است که ماشین‌های بزرگ‌تر بهتر به فروش می‌رسند. قیمت ابتدا انتخاب شده است زیرا این پیشگو می‌باشد که همبستگی خیلی قوی با فروش‌ها دارد. سپس باقی‌مانده پیشگوها تحلیل می‌شوند تا تعیین شود که کدام یک برای نتیجه در مرحله بعدی مناسب‌تر است (a).

**Beta In** در صورتی که بعداً به حساب آورده شود مقدار ضریب استاندارد شده برای پیشگو می‌باشد (b). تمام مقادیر معنادار کمتر از **0.05** هستند، بنابراین هر کدام از پیشگوهای باقی‌مانده در صورتی که در مدل به حساب آورده شوند کافی می‌باشند (c).

برای انتخاب بهترین متغیر برای افزودن به مدل، به همبستگی جزئی که همبستگی خطی بین پیشگوی پیشنهادی و متغیر وابسته بعد از حذف تأثیر مدل جاری می‌باشد نگاه کنید. فاصله محوری دو چرخ (**Wheelbase**) بعداً انتخاب شده است چون این بزرگترین همبستگی جزئی را دارد (d).

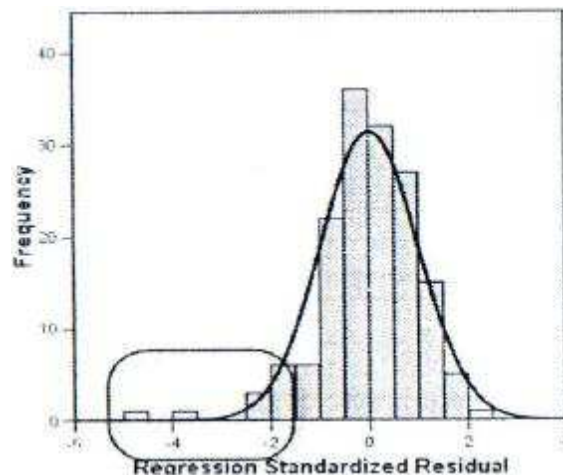
بعد از اضافه کردن فاصله محوری دو چرخ به مدل، هیچکدام از پیشگوه‌های باقی‌مانده معنی‌دار نیستند (e). با این وجود، نوع کامیون فقط مقدار غیر معتبر 0.05 را دارد، بنابراین ممکن است بخواهید بعداً آن را به صورت دستی تحلیل نمایید تا چگونگی تغییر نتایج را ببینید (f). اندازه موتور (Engine Size) در صورتی که به مدل اضافه شود باید ضریب بتای (beta) بزرگتری داشته باشد. این بدان علت است که اندازه موتور یک تیرانس نسبتاً پایین در مقایسه با نوع کامیون دارد که نشان می‌دهد همبستگی محکم‌تری با قیمت و فاصله محور دو چرخ دارد (g).



Model	a	b		c	d	Collinearity Statistics
		Beta In	t	Sig.	Partial Correlation	Tolerance
1	Zscore Type	.251 <sup>a</sup>	3.854	.000	.301	.998
	Zscore Engine size	.342 <sup>a</sup>	4.128	.000	.320	.611
	Zscore Horsepower	.257 <sup>a</sup>	2.062	.041	.167	.293
	Zscore Wheelbase	.356 <sup>a</sup>	5.718	.000	.424	.988
	Zscore Width	.244 <sup>a</sup>	3.517	.001	.277	.892
	Zscore Length	.308 <sup>a</sup>	4.790	.000	.365	.976
	Zscore Curb weight	.346 <sup>a</sup>	4.600	.000	.353	.722
	Zscore Fuel capacity	.266 <sup>a</sup>	3.687	.000	.289	.820
	Zscore Fuel efficiency	-.198 <sup>a</sup>	-2.584	.011	-.203	.758
2	Zscore Type	.129 <sup>b</sup>	1.928	.056	.157	.835
	<b>g</b> Zscore Engine size	.145 <sup>b</sup>	1.576	.117	<b>f</b> .128	.445
	Zscore Horsepower	.028 <sup>b</sup>	.229	.819	.019	.256
	Zscore Width	-.025 <sup>b</sup>	-.275	.784	-.023	.470
	Zscore Length	.027 <sup>b</sup>	.237	.813	.020	.290
	Zscore Curb weight	.105 <sup>b</sup>	1.028	.306	.084	.365
	Zscore Fuel capacity	.002 <sup>b</sup>	.024	.981	.002	.443
	Zscore Fuel efficiency	.014 <sup>b</sup>	.164	.870	.014	.553

## بررسی نرمال بودن عبارت خطا

شکل هیستوگرام کاملاً از شکل منحنی نرمال تبعیت می‌نماید، اما یک یا دو باقی‌مانده (Residual) منفی وجود دارد. برای اطلاعات بیشتر، به قسمت بعدی یعنی «تشخیص موردی» مراجعه نمایید.



## تشخیص موردی

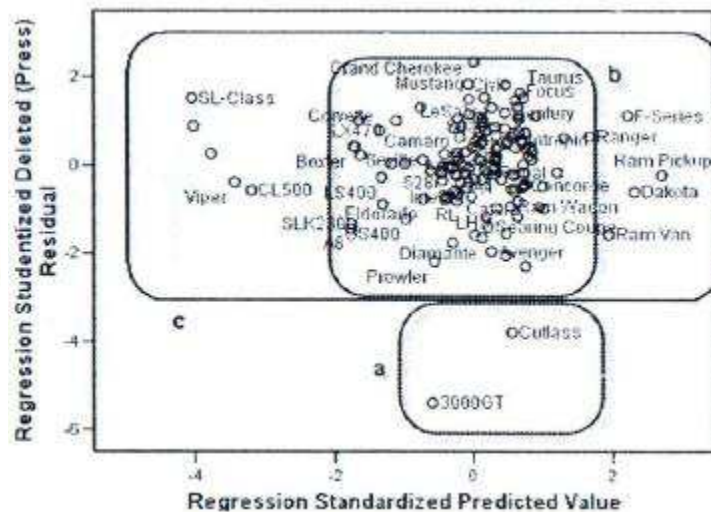


این جدول موارد را با مقادیر باقی مانده خیلی منفی به صورت **3000GT** و **Cutlass** نشان می دهد. این بدان معنی است که نسبت به ماشین های دیگر با همین اندازه و قیمت، این دو مدل در بازار تحت اجرا هستند (a). روشن است که گزینه های **Breeze**، **Prowler** و **SW** نیز با اندازه کمتری اجرا شده اند (b). **Explorer** تنها موردی است که به نظر می رسد اجرا کننده خوبی باشد (c).

Case Number	Model	Std. Residual	Log-transformed sales	Predicted Value	Residual
53	Explorer	2.297	5.62	3.2953	2.3278
84	3000GT	-4.905	-2.21	2.7638	-4.9711
109	Cutlass	-3.610	.11	3.7651	-3.6589
116	Breeze	-2.252	1.66	3.9393	-2.2830
118	Prowler	-2.139	.63	2.7955	-2.1685
132	SW	-2.012	1.65	3.6927	-2.0397

## نمودار پراکنش باقی مانده

نمودار باقی مانده ها با مقادیر پیش بینی شده به وضوح کامیون های تحت اجرایی را نشان می دهد (a). به علاوه، می توانید ببینید که **Breeze**، **Prowler**، **SW** و **Explorer** کاملاً به اکثر موارد نزدیک هستند. این نشان می دهد که اجرای **Breeze**، **Prowler** و **SW** و اجرای فوق العاده **Explorer** می توانند به دلیل شانس و تصادف بوده باشند (b). برخی از موارد در این نمودار در سمت راست و چپ خوشه ای شده اند. مادامی که کامیون ها در این خوشه ها دارای باقی مانده های بزرگ نباشند، فاصله اینها از خوشه کلی ممکن است در تعیین ضرائب رگرسیون این موارد را خیلی تحت تأثیر بگذارد (c).



۱. برای بررسی باقی مانده ها با قیمت، مسیر **Graphs > Legacy Dialogs > Scatter / Dot** را از منوی اصلی برگزینید.

۲. دکمه **Define** را کلیک کنید.

۳. **Standardized Residual** را به عنوان متغیر **y** و **Price in thousands : Z Score** را به عنوان متغیر **x** انتخاب نمایید.

۴. **Model** را به عنوان متغیر عنوان گذاری موارد (**Case Labels by**) برگزینید.

۵. **Ok** را کلیک کنید.

نمودار پراکنش حاصله، نقاط را در سمت راست و دور از هم نشان می دهد (a).

۶. برای شناسایی نقاط، نمودار را با دو بار کلیک کردن آن فعال کنید.

۷. روی آیکن **Data Label Mode** کلیک کنید.

۸. نقاط را برگزینید.

فواصل نمودار پراکنش که نقاط غیر عادی نشان داده شده در نمودار پراکنش باقی مانده ها بر حسب مقادیر پیش بینی شده می باشند کامیون های گران قیمت هستند. توزیع قیمت به سمت راست چولگی دارد، بنابراین می توانیم از لگاریتم تبدیل استاندارد شده قیمت ها در تحلیل های بعدی استفاده نماییم.

۹. برای بررسی باقی مانده ها با فاصله محوری چرخ، کادر محاوره **Simple Scatterplot** را فراخوانی نمایید.

۱۰. به جای **Price in thousands : Z Score** از **Wheelbase : Z Score** به عنوان متغیر **x** استفاده نمایید.

۱۱. **Ok** را کلیک کنید.

۱۲. نمودار پراکنش نتیجه، نقاط را در سمت راست به دور از هم نشان می دهد.

۱۳. برای شناسایی نقاط، نمودار را با دو بار کلیک کردن فعال نمایید.

۱۴. آیکن **Data Label Mode** را کلیک کنید.

۱۵. نقاط را برگزینید.

نقاط موجود در سمت راست خوشه کلی در نمودار با نقاط موجود در سمت راست کلی در نمودار پراکنش باقی مانده ها بر حسب مقادیر پیش بینی شده مرتبط هستند. همانند توزیع قیمت، می توانید از لگاریتم تبدیل استاندارد شده فاصله محورهای تحلیل های بعدی استفاده نمایید.

## شناسایی نقاط مؤثر

۱. برای بررسی فاصله **Cook** نسبت به مقدار مؤثر مرکزی، کادر محاوره **Simple Scatterplot** را فراخوانی نمایید.

۲. به جای **Standardized Residual** گزینه **Cook's Distance** را به عنوان متغیر **y** انتخاب نمایید.

۳. **Centered Leverage Value** را به جای **Z Score: Wheelbase** به عنوان متغیر **x** برگزینید.

۴. **Ok** را کلیک کنید.

نمودار پراکنش حاصله چندین نقطه غیر عادی را نشان می‌دهد.

۵. برای شناسایی نقاط، نمودار را با دو بار کلیک کردن فعال نمایید.

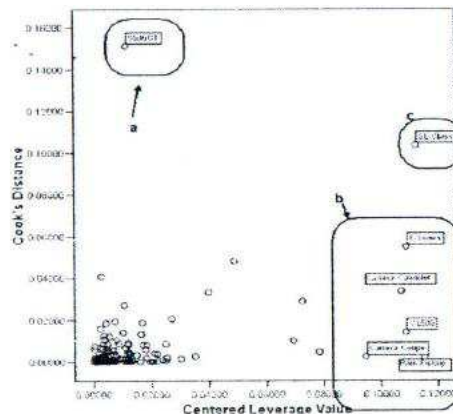
۶. آیکن **Data Label Mode** را کلیک کنید.

۷. نقاط را برگزینید.

نقطه با بزرگترین فاصله **Cook** نقطه **3000GT** می‌باشد. این دارای مقدار مؤثر بزرگ نمی‌باشد، بنابراین با افزودن مقداری تغییرپذیری به برآوردهای رگرسیون، نقطه **3000GT** روی شیب معادله رگرسیون تأثیر ندارد (a). به طور مشابهی، خیلی از موارد با مقادیر مؤثر بالا دارای فواصل **Cook** بزرگی نیستند، بنابراین احتمالاً تأثیر

زیادی روی مدل ندارند (b).

مشکل‌دارترین مورد **SL- Class** می‌باشد، که هم خیلی مؤثر است و هم فاصله **Cook** زیادی دارد. این کامیون یک فاصله محوری و برجسب قیمت زیاد دارد، بنابراین تأثیر آن با مشکلات مشاهده شده در نمودارهای پراکنش باقی‌مانده در ارتباط است (c).



ادامه فصل ۲۴ را در مرجع کاربردی **SPSS 20** (39) دنبال نمایید.

