



با سلام و احترام،

ادامه فصل پانزدهم : آنالیز اکتشافی داده‌ها

خروجی‌های توصیف‌های اکتشافی

آماره‌های توصیفی اکتشافی

جدول **Explore Descriptive** نیز مشابه آماره‌های فراوانی و توصیفی، خلاصه‌های آماره‌های مربوط به متغیرهای عددی پیوسته را فراهم می‌آورد. این جدول نیز شامل شاخص‌های میانه و میانگین، انحراف معیار، چولگی و کشیدگی با همان مشخصات می‌باشد.

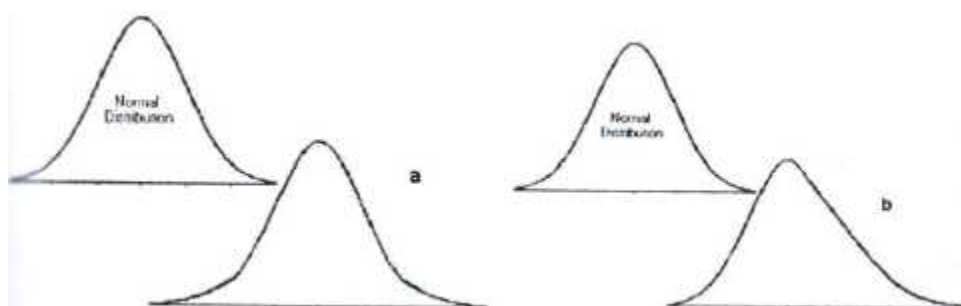
تخمین زننده **M** اکتشافی **M-Estimators**

M-Estimators شاخص‌های قوی از گرایش به مرکز هستند که می‌توانند به عنوان جانشینی برای میانگین و میانه به کار روند. (به این علت گفته شد شاخص قوی، چون به انحراف از حالت نرمال حساسیت ندارند.)

M-Estimators

	Division	
	Consumer Products	Business Products
Huber's M-Estimator	\$292,070.47	\$358,004.73
Tukey's Biweight	\$285,405.94	\$326,021.41
Hampel's M-Estimator	\$290,005.75	\$351,967.79
Andrews' Wave	\$285,292.41	\$325,641.46

وقتی داده‌ها دارای توزیع متقارن با دنباله‌های طولانی می‌باشند (a) یا وقتی داده‌ها دارای مقادیر انتهایی می‌باشند، تخمین زننده **M** برآوردهای گرایش به مرکز مناسب‌تری نسبت به میانگین یا میانه فراهم می‌آورد (b).



صدک اکتشافی

صدک‌ها درصد مواردی را که در زیر مقادیر نمایش داده شده قرار می‌گیرند مشخص می‌کنند. صدک 25م به چارک اول نیز مشهور است. صدک 75م به چارک سوم نیز معروف می‌باشد و صدک 50م همان میانه است.

Percentiles	Weighted Average		Tukey's Hinges	
	Consumer Products	Business Products	Consumer Products	Business Products
5	\$203,400	\$235,200		
10	\$221,400	\$258,300		
25	\$257,400	\$277,500	\$257,400	\$278,250
50	\$289,800	\$339,000	\$289,800	\$339,000
75	\$334,800	\$526,500	\$334,800	\$523,875
90	\$414,000	\$700,000		
95	\$466,200	\$847,500		

مقادیر انتهایی اکتشافی

جدول مقادیر انتهایی، موارد را با پنج مقدار بزرگتر و پنج مقدار کوچک نمایش می‌دهد. ستون **Case Number** تعداد سطر هر مورد را در پنجره **Data Editor** مشخص می‌کند.

	Consumer Products			Business Products	
	Case Number	Value	Case Number	Value	
Highest	1	308	\$697,500	27	\$1,350,000
	2	209	\$601,000	30	\$1,108,250
	3	203	\$652,500	17	\$1,037,500
	4	56	\$648,000	285	\$1,035,000
	5	74	\$612,000	366	\$1,000,000
Lowest	1	311	\$189,000	309	\$213,000
	2	281	\$190,890	219	\$213,000
	3	92	\$194,490	21	\$217,500
	4	337	\$194,490	59	\$219,000
	5	368	\$196,200	117	\$222,000

برای هر مورد با مقادیر زیاد یا کم غیر عادی، می‌توانید یک نمای نزدیک‌تر از همه داده‌های آن مورد در پنجره **Data Editor** داشته باشید.

آزمون‌های اکتشافی حالت نرمال

آماره کولموگروف-اسمیرنوف قضیه‌ای را که می‌گوید داده‌ها به صورت نرمال توزیع شده‌اند بررسی می‌نماید. سطح معنی‌داری کوچک (عموماً کمتر از **0.05**) مشخص می‌کند که توزیع داده‌ها به طور معنی‌داری از توزیع نرمال فاصله دارد.

Division	Kolmogorov-Smirnov		
	Statistic	df	Sig.
Consumer Products	.156	179	.000
Business Products	.192	211	.000

اگر بیشتر از **50** مورد وجود داشته باشد، آزمون شاپیرو-ویلک (**Shapiro-Wilk**) نیز نمایان می‌شود. از آنجا که خیلی از آزمون‌های آماری فرض می‌کنند داده‌ها به صورت نرمال توزیع شده‌اند، فکر خوبی است که توزیع را بررسی نمایید و یا از تبدیل داده‌ها و یا از آزمون‌های ناپارامتری که به توزیع نرمال داده‌ها نیازی ندارند استفاده کنید.

Division	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Consumer Products	.232	22	.003	.741	22	.010
Business Products	.271	28	.000	.810	28	.010

آزمون‌های اکتشافی یکنواختی واریانس

آماره لون (**Levene**) فرضیه یکسانی واریانس متغیر وابسته را برای گروه‌هایی که با متغیرهای عامل مقوله‌ای تعریف شده است بررسی می‌نماید. سطح معنی‌داری کوچک (عموماً کوچکتر از **0.05**) مشخص می‌کند که واریانس به طور معنی‌داری بین گروه‌ها متفاوت است.

	Levene Statistic	df1	df2	Sig.
1996 Sales	6.708	2	387	.001

اگر شما گزینه **Power estimation** را برای یکنواختی آزمون واریانس انتخاب نمایید (در کادر محاوره **Explore: Plots**)، نمودار **Spread** بر حسب **Level** عدد توانی را برای تبدیل داده‌هایتان پیشنهاد می‌نماید.

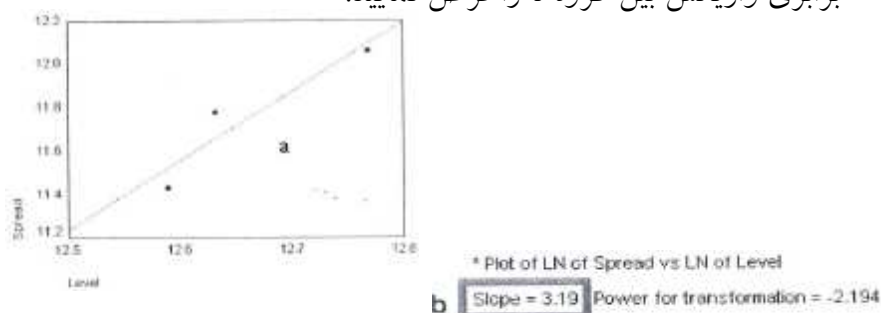
از آنجا که خیلی از آزمون‌های آماری که گروه‌ها را مقایسه می‌کنند فرض برابری واریانس بین گروه‌ها را در نظر می‌گیرند، فکر خوبی است که ابتدا واریانس بررسی شود و یا تبدیلات مناسبی را اعمال نمایید یا از آزمون‌های ناپارامتری که نیازی به برابری واریانس ندارند استفاده کنید.

Power	Transformation
2	Square
1	None
1/2	Square root
0	Logarithm
-1/2	Reciprocal of the square root
-1	Reciprocal

نمودارهای اکتشافی **Spread** بر حسب **Level**

نمودار **Spread** بر حسب **Level** یک ابزار تشخیصی برای ارزیابی برابری واریانس بین گروه‌ها می‌باشد.

اگر یک خط از میان نقاط نمودار عبور داده شود (a) و مقدار شیب نزدیک به صفر نباشد (b) شما نمی‌توانید برابری واریانس بین گروه‌ها را فرض نمایید.



مبحث را در مرجع کاربردی **SPSS 20** (۲۳) دنبال نمایید.

