



با سلام و احترام،

ادامه فصل سیزدهم: خلاصه آماره ها با استفاده از فراوانی ها

جدول آماری

جدول آماری چند نکته جالب را درباره توزیع فروش بیان می کند. مرکز توزیع را می توان با میانه (یا چارک دوم) **20.25** و نیمه مقادیر داده ها که بین **12.0** و **52.875** می افتد (اولین و سومین چارک) تخمین زد.

منظور از چارک ها نقاطی بر روی مقیاس اندازه گیری هستند که کلیه مشاهدات و نمره ها را به چهار قسمت مساوی تقسیم می کند. چارک اول نقطه ای است که **۲۵** درصد نمرات (معادل یک چهارم) در پایین یا مساوی آن قرار دارند. چارک دوم یا وسط برابر با میانه بوده و از طریق تفاضل چارک اول از چارک سوم به دست می آید. چارک سوم نقطه ای است که **۷۵** درصد نمرات در پایین یا مساوی آن واقع شده اند. هر چه مقدار انحراف چارکی بیشتر باشد، پراکندگی نمرات از میانه بیشتر است.

شاخص چولگی (**Skewness**) که به شاخص انحراف از قرینگی نیز معروف است، میزان تقارن یک توزیع را اندازه گیری می کند. این ضریب بین **-۳** و **+۳** تغییر می کند و در عمل به ندرت به حد خود نزدیک می شود. چولگی داده ها می تواند به سه شکل مثبت، منفی و صفر روی دهد:

۱. **چولگی مثبت:** یعنی توزیع داده ها و منحنی آن به گونه ای مثبت به صورت چوله درآمده است. به عبارت دیگر، نمرات افراد حول و حوش مقادیر پایین متغیر متمرکز هستند.
۲. **چولگی منفی:** یعنی توزیع داده ها و منحنی آن به گونه ای منفی به صورت چوله درآمده است. یعنی نمرات افراد حول و حوش مقادیر بالای متغیر متمرکزند.
۳. **چولگی صفر:** یعنی توزیع داده ها متقارن است.

در کل هر چه مقدار چولگی بزرگتر باشد، (معمولاً بزرگتر از یک) توزیع داده ها تفاوت فاحشی با توزیع متقارن دارد.

شاخص کشیدگی (Kurtosis)

نمایان گر میزان تفاوت توزیع پاسخ گویان از توزیع نرمال می باشد. این آماره، نشان می دهد که آیا مقادیر توزیع پاسخ گویان بزرگتر از میانگین است یا کوچکتر از آن؟ این شاخص یکی از آماره های مناسب برای مقایسه پراکندگی توزیع جامعه با توزیع نرمال است.

توزیع داده ها با توجه به مقدار کشیدگی می تواند به سه شکل زیر باشد:

۱. **کشیدگی مثبت:** یعنی توزیع داده ها از توزیع نرمال بلندتر است. در چنین حالتی، توزیع داده ها حول میانگین متمرکز شده و از پراکندگی کمتری برخوردار است.
۲. **کشیدگی منفی:** یعنی توزیع داده ها از توزیع نرمال کوتاه تر است. در چنین حالتی، توزیع داده ها در اطراف میانگین پخش شده و از پراکندگی زیادی برخوردار است.
۳. **کشیدگی صفر:** یعنی توزیع داده ها متقارن است. یعنی مقدار کشیدگی داده ها با مقدار کشیدگی توزیع نرمال کاملاً مساوی است.

در کل، هر چه مقدار کشیدگی بزرگتر (معمولاً بزرگتر از یک) باشد، در آن صورت نتیجه می گیریم که توزیع داده ها تفاوت فاحشی با توزیع نرمال دارد.

خلاصه کردن داده های تبدیلی (Transformed)

بسیاری از دستورالعمل های آماری برای داده های مقداری موقعی که توزیع مقادیر داده ها از نوع غیر نرمال (non- normal) باشد معتبر نیست (همانند مورد Amount of Last Sale). بعضی وقت ها، یک تبدیل متغیر (Transformation) توزیع مقادیر را به نرمال نزدیک تر می کند.

تبدیل داده ها

۱. برای تبدیل متغیر فروش (Sale) مسیر Transform > Compute را از نوار منو برگزینید.
۲. کلمه Logsale را در کادر Target Variable (متغیر هدف) تایپ کنید.

۳. LN (Sale) را در کادر Numeric Expression وارد نمایید. (در اینجا ما تابع Ln یعنی لگاریتم

طبیعی مدنظرمان می باشد.)

۴. روی Ok کلیک کنید.

تبدیل لگاریتم، یک گزینه مناسب است زیرا Amount of Last Sale فقط مقادیر مثبت را به خود می گیرد و چولگی آن در یک خط راست می باشد.

۵. جعبه محاوره Frequencies را بازخوانی نمایید.

۶. Sale را به عنوان متغیر تحلیل از حالت انتخاب خارج سازید.

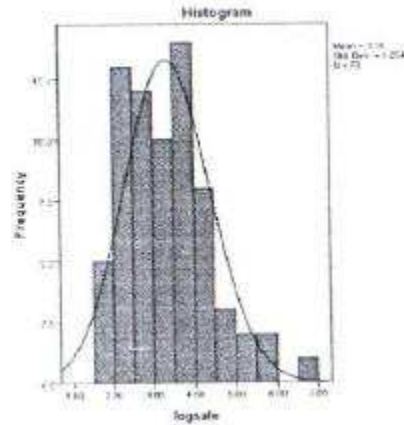
۷. Logsale را به عنوان متغیر تحلیل برگزینید.

۸. روی Ok کلیک کنید.

از جدول آماره پیداست که تبدیل، توزیع را به نزدیکی نرمال آورده است. چولگی (Skewness) و کشیدگی (Kurtosis) خیلی کاهش یافته است، و میانگین و میانه خیلی به هم نزدیک شده اند. هیستوگرام نیز خیلی به منحنی نرمال نزدیک شده است.

logsale		
N	Valid	70
	Missing	0
Mean		3.3373
Median		3.1772
Std. Deviation		1.05361
Skewness		.721
Std. Error of Skewness		.287
Kurtosis		.367
Std. Error of Kurtosis		.666
Minimum		1.79
Maximum		6.65
Percentiles	25	2.4849
	50	3.1772
	75	3.9879





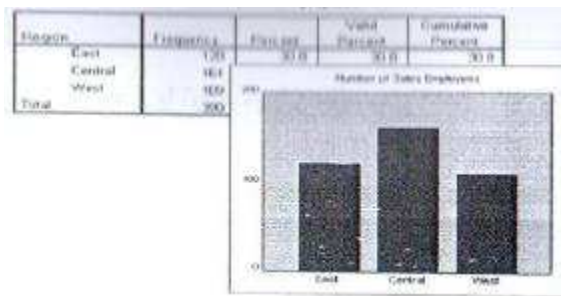
خروجی های جداول فراوانی

جدول فراوانی

در مورد هر مقدار یک متغیر، جدول فراوانی تعداد مرتبه ای را که مقدار رخ می دهد نشان می دهد (a). جدول تعداد و درصد موارد مربوط به مقدار هر متغیر را نشان می دهد (b). این جدول برای خلاصه کردن متغیرهای مقوله ای یعنی متغیرهای با یک تعداد محدود از مقوله های مجزا، سودمند می باشد (c).

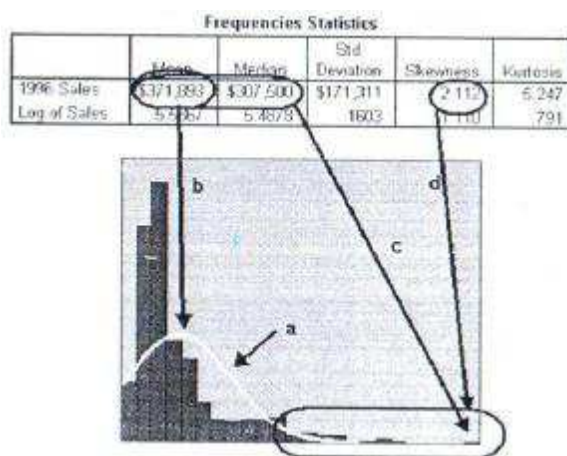
Region	Frequency	Percent	Valid Percent	Cumulative Percent
East	120	30.8	30.8	30.8
Central	161	41.3	41.3	72.1
West	109	27.9	27.9	100.0
Total	290	100.0		

اطلاعات در جدول فراوانی را می توان به صورت گرافیکی در نمودار میله ای نشان داد.



آماره های فراوانی

جدول آماره های فراوانی، آماره های خلاصه را برای متغیرهای عددی پیوسته فراهم می آورد. آماره های خلاصه شامل شاخص های گرایش به مرکز نظیر میانگین و میانه (ستون **Mean** و **Median**) و شاخص های پراکندگی نظیر انحراف معیار (**Std. Deviation**) و همچنین شاخص های توزیع، مثل چولگی (**Skewness**) و کشیدگی (**Kurtosis**) می باشد. این دو شاخص میزان متفاوت بودن توزیع نسبت به توزیع نرمال را نشان می دهد. اطلاعات خلاصه مربوط به داده های عددی پیوسته را می توان در یک هیستوگرام نیز نمایش داد.

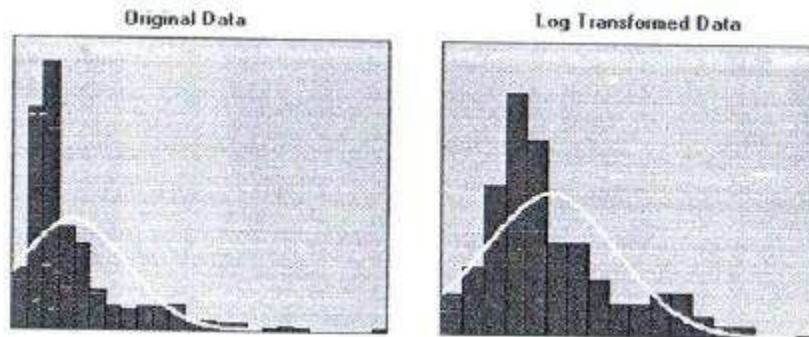


تلفیق منحنی با هیستوگرام نشان می دهد که توزیع چقدر با توزیع نرمال تفاوت دارد (a). نوک منحنی نرمال مقدار میانگین می باشد (b). اگر میانه خیلی با میانگین فرق کند، این ممکن است درصد مقادیر پرت را مشخص نماید (که روی میانگین تأثیر دارد اما روی میانه ندارد) (c).

چولگی میزان نامتقارن بودن توزیع را نشان می دهد. یک مقدار مثبت برای چولگی مشخص می کند که توزیع چوله به صورت مثبت است (d). در حالت کلی، یک مقدار چولگی بزرگتر از چولگی دیگر مشخص می کند که توزیع به طور معنی داری از حالت نرمال (توزیع متقارن) فاصله دارد.

توابع تبدیلی نظیر تبدیل لگاریتم (**Log**) می تواند توزیع چوله را بیشتر نرمال نماید. از آنجا که بسیاری از آزمون های آماری فرض می کنند داده ها به طور نرمال توزیع شده اند، بهتر است که توزیع داده ها را بررسی نمایید و یا توابع تبدیلی مناسب را اعمال کنید یا از آزمون های ناپارامتریک که به داده های توزیعی به صورت نرمال نیاز ندارند استفاده نمایید.

Frequencies Statistics					
	Mean	Median	Std. Deviation	Skewness	Kurtosis
1996 Sales	\$371.893	\$307.500	\$171.311	2.112	5.247
Log of Sales	5.5367	5.4878	.1603	1.110	.791



مبحث را در مرجع کاربردی SPSS 20 (۲۰) دنبال نمایید.



Telegram.me/iepnu

کانال تخصصی مهندسی صنایع دانشگاه پیام نور