

تبدیل داده‌ها

گاهی داده‌های خامی که برای تحلیل داریم مناسب گروهی از آزمون‌های آماری نیستند و برای این که بتوانیم از این دسته آزمون‌های آماری استفاده کنیم و همچنین دقت تحلیل را بالا ببریم باید در داده‌های خام تغییراتی ایجاد کنیم. یکی از این تغییرات، تبدیل داده‌ها نام دارد. تبدیل داده‌ها، روش‌هایی ریاضی است که برای تعدیل متغیرهایی به کار می‌رود که از مفروضه‌های آماری نرمال بودن، خطی بودن و یکسانی پراکندگی پیروی نمی‌کنند یا الگوهای با داده‌های پرت غیرمعمول دارند (میزر و دیگران، ۱۳۹۱: ۱۱۱).

در مجموع زمانی که پیش شرط‌های آزمون‌های چندمتغیره برقرار نباشد، باید داده‌های به دست آمده را تبدیل کنیم تا امکان استفاده از برخی آزمون‌های مدنظر (عموما پارامتریک) فراهم شود.

در ابتدا باید میزان تخطی و تفاوت داده‌ها از پیش‌فرض‌های ذکر شده را تعیین کرد و در صورتی که پیش‌فرض‌ها یا پیش‌شرط‌های آماری به دست آمده دارای تفاوت قابل اعتنایی با مقدار معیار باشند از روش تبدیل داده‌ها استفاده کرد. تبدیل داده‌ها با هدف تعدیل متغیرها از جنبه علمی روشی پذیرفته شده است. البته زمانی که اختلاف داده‌ها با پیش‌فرض‌های آماری اندک باشد و به طور تقریبی مفروضات آماری برقرار باشد می‌توان از تبدیل داده‌ها صرف نظر کرد.

باید توجه داشت که تبدیل داده‌ها تا اندازه‌ای مانند شمشیر دولبه است. حسن این روش این است که می‌تواند دقت معنی‌داری تحلیل‌های آماری را افزایش دهد و عیب آن این است که ممکن است تفسیر داده‌ها را دشوارتر کند. در نتیجه باید از روش تبدیل داده‌ها به شیوه‌ای مدبرانه استفاده کرد.

دشواری کردن تفسیر داده‌ها بدین معناست که وقتی داده‌ها را تبدیل می‌کنیم، مقدار حداقل و حداکثر و شیوه توزیع متغیر و تمامی شاخص‌های میانگین و انحراف استاندارد تغییر می‌کند و با حالت معمول و عادی تفاوت پیدا می‌کند. مثلا اگر سن افراد که به صورت کمی (نسبی) سنجیده شده است را به توان دو برسانیم شاخص‌های آماری سن افراد تغییر می‌کند و با سن‌های غیر عادی مثل ۲۵۰، ۳۰۰ و غیره مواجه می‌شویم. یا وقتی متغیری مانند اعتماد اجتماعی داریم و با ۱۰ سوال این متغیر را سنجیدیم و دامنه میانگین این متغیر بین ۱ تا ۵ باشد، لگاریتم گرفتن از این متغیر دامنه نمرات را تغییر می‌دهد و توضیح و تفسیر متغیر را با مشکل مواجه می‌کند. یکی از راه‌های رفع این مشکل این است که هنگام گزارش یافته‌های توصیفی و شاخص‌های آماری (مانند میانگین، انحراف استاندارد و مقدار حداقل و حداکثر)؛ یافته‌ها و شاخص‌های آماری را هم به صورت عادی (قبل از تبدیل داده‌ها) و هم بعد از تبدیل داده‌ها گزارش کنیم.

انواع روش‌های تبدیل داده

روش‌های مختلفی برای تبدیل داده‌ها وجود دارد که برخی از روش‌های متداول عبارتند از: ریشه دوم^۱، لگاریتم^۲، وارون^۳ و مجذور کردن^۴. بین آمارشناسان در مورد استفاده از انواع روش‌های تبدیل داده‌ها در شرایط خاص اتفاق نظر وجود

¹ Square root

² Logarithm or Log

³ Invers

⁴ Square

ندارد. با وجود این، یک راهبرد اساسی برای به کار بردن روش‌های تبدیل این است که برحسب جدی بودن تخطی از پیش-فرض‌های آماری، راهبردهای پیشرفته‌تر (سطح بالاتر) به کار بسته شود. برای مثال، تاباچینک و فیدل (۲۰۰۱) و مرتلر و ااناتا (۲۰۰۱) برای پیشرفت تبدیل داده‌ها از ریشه دوم (برای اصلاح تخطی متوسط)، لگاریتم (برای تخطی اساسی‌تر) و سپس ریشه دوم و وارون (برای رسیدگی به تخطی شدید) جانبداری می‌کنند. مجذور کردن یک متغیر در رابطه دو متغیری غیرخطی می‌تواند مشکل غیرخطی بودن را به طور مؤثر کاهش بدهد (میزر و دیگران، ۱۳۹۱:۱۱۳). به طور کلی زمانی که داده‌ها نرمال نیستند از روش‌های ریشه دوم، لگاریتم و وارونه کردن برای تبدیل داده‌ها استفاده می‌شود.

جدول ۳-۴- انواع روش‌های تبدیل داده‌ها برحسب نوع انحراف از پیش فرض‌ها

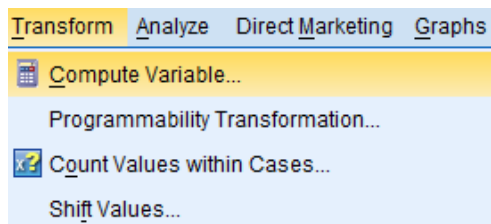
نوع مشکل در پیش فرض‌ها	نوع تبدیل پیشنهادی
نقض پیش‌فرض‌ها به طور کلی	ریشه دوم (برای اصلاح تخطی متوسط) لگاریتم (برای تخطی اساسی‌تر) ریشه دوم و وارون (برای رسیدگی به تخطی شدید)
غیرخطی بودن رابطه	مجذور کردن (به توان دو رساندن) ریشه دوم
چولگی متوسط (مثبت یا منفی)	لگاریتم (Log 10)
چولگی شدید (مثبت یا منفی)	

مثال



نتایج آزمون‌های نرمال بودن نشان داد که یکی از متغیرهای مثال کتاب از توزیع نرمال برخوردار نیست. بدین جهت که قصد استفاده از آزمون‌های پارامتریک را داریم نیازمند این هستیم که توزیع این متغیر را به توزیع نرمال نزدیک کنیم. در نتیجه از تبدیل داده‌ها بهره می‌گیریم.

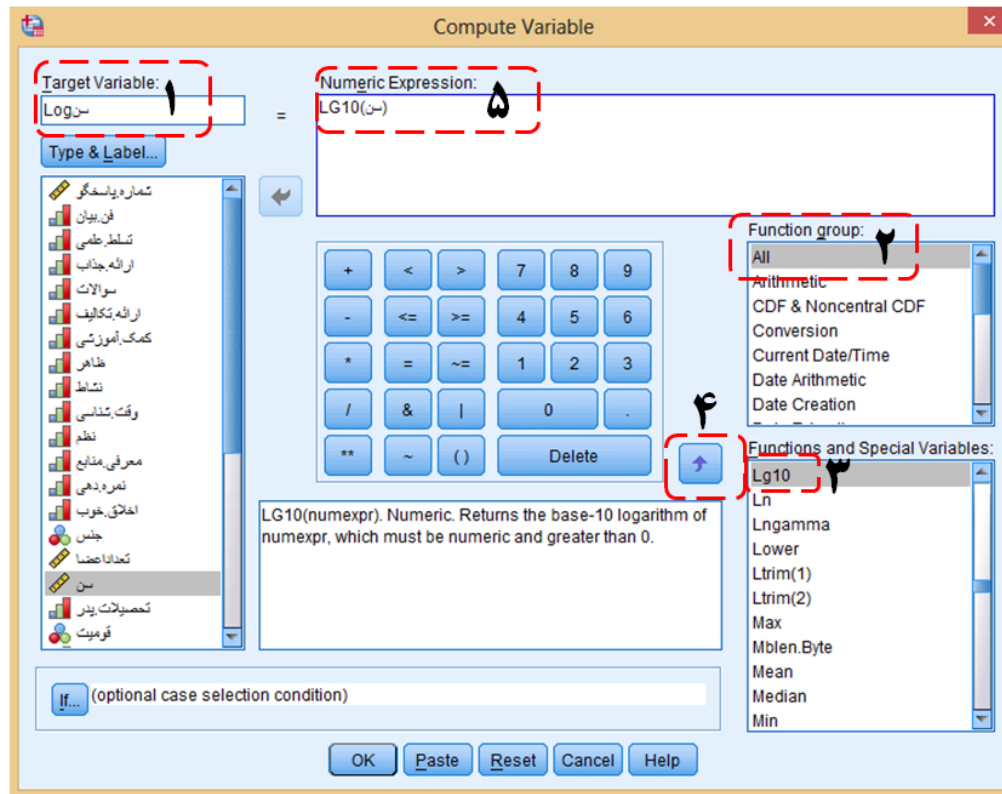
اجرا:

تبدیل داده‌ها در برنامه SPSS از طریق دستور Compute انجام می‌شود. این دستور در منوی Transform در دسترس است.



۱- در پنجره ایجادشده (Compute Variable) و در کادر Target Variable نامی برای متغیری که می‌خواهیم ایجاد کنیم انتخاب می‌کنیم (نام انتخاب شده باید با نام متغیرهای موجود تفاوت داشته باشد).

۲- برای گرفتن لگاریتم، در کادر Function group بر روی گزینه All و در کادر پایین آن (Function and Special Variables) بر روی گزینه Lg10 کلیک می‌کنیم و با استفاده از دستور  آن را وارد کادر بیان عددی (Numeric Expression) می‌کنیم. متغیر مورد نظر را از کادر متغیرها در سمت چپ انتخاب می‌کنیم و با دستور  در داخل پرانتز کادر Numeric Expression قرار می‌دهیم و در انتها گزینه OK را می‌زنیم. دستورات مطرح شده در شکل بعد نشان داده شده است.



عبارت داخل کادر بیان عددی باید بدین صورت باشد.

Numeric Expression:
LG10(سن)

چگونگی انجام سایر روش‌های تبدیل

نکته: برای مجذور کردن یا به توان دو رساندن متغیر کافیست در کادر Numeric Expression متغیر مد نظر را وارد کرده و سپس علامت * (ضرب) را از کادر ماشین حساب در کادر قرار دهیم و دوباره همان متغیر را وارد کادر کنیم.

Numeric Expression:
سن * سن

نکته: برای وارون یا معکوس کردن متغیر، عدد ۱ را در کادر Numeric Expression وارد کنیم و علامت / (تقسیم) را از کادر ماشین حساب انتخاب کرده و سپس متغیر مد نظر را وارد کادر کنیم.

Numeric Expression:

سن / 1

نکته: برای به دست آوردن ریشه دوم (جذر گرفتن) کافیست که گزینه SQRT را از کادر Function and... انتخاب کنیم و مانند مورد لگاریتم عمل کنیم.

Numeric Expression:

SQRT(سن)

نتیج:

بعد از انجام تبدیل متغیر، متغیری جدید در صفحه داده‌ها ایجاد می‌شود که در تحلیل داده‌ها از آن استفاده می‌کنیم.

متغیرهای تبدیل شده با روش‌های گوناگون	معکوس	ریشه دوم	توان دوم	لگاریتم	متغیر قبل از انجام تبدیل داده
	.05	4.69	484.00	1.34	سن
	.04	5.10	676.00	1.41	22.0
	.03	5.39	841.00	1.46	26.0
	.03	5.57	961.00	1.49	29.0
	.03	5.66	1024.00	1.51	31.0
	.04	5.00	625.00	1.40	32.0
	.04	4.90	576.00	1.38	25.0
	.04	5.10	676.00	1.41	24.0
	.04				26.0

☑ نکته: زمانی که شیوه تعریف داده‌ها طوری باشد که برخی از موردها با کد ۰ (صفر) نشان داده شده باشند، قبل از تبدیل داده‌ها باید تغییراتی در داده‌های خام ایجاد کرد. بدین صورت که به همه داده‌های خام متغیر مورد نظر عدد ۱ را اضافه می‌کنیم تا حداقل نمره در داده‌ها برابر با ۱ باشد و بتوان به نتایج معنی‌دار و مناسبی از تبدیل داده‌ها دست یافت. برای افزودن عدد ۱ به همه داده‌ها کافیست در دستور Compute و در کادر Numeric Expression ابتدا نام متغیر را وارد کنیم و سپس علامت + (جمع) را وارد کنیم و عدد یک را بزنیم، مانند شکل پایین.

Numeric Expression:

سن + 1

