

آموزش

تبدیل داده ها در SPSS

تدوین: رامین کریمی

www.kharazmi-statistics.ir

خلاصه مقاله	
عنوان	تبدیل داده ها
تعریف	روش های ریاضی برای تعدیل و آماده کردن متغیرها برای تحلیل
هدف	برقراری پیش فرض های آماری: نرمال بودن، خطی بودن رابطه و یکسانی پراکندگی
انواع روش ها	ریشه دوم، لگاریتم، وارون و توان دوم و...
روش اجرا	دستور Compute برنامه SPSS
مزایا	برقراری پیش فرض ها و پیش شرط های آماری آزمون ها، در نتیجه افزایش دقت معنی داری تحلیل های چندمتغیره
معایب	دشواری تر شدن تفسیر داده ها

تبدیل داده‌ها چیست و چرا انجام می‌شود؟

گاهی داده‌های خامی که برای تحلیل داریم مناسب گروهی از آزمون‌های آماری نیستند و برای این که بتوانیم از آزمون‌های آماری استفاده کنیم و همچنین دقت تحلیل را بالا ببریم می‌بایست 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 الگوهایی با داده‌های پرت غیرمعمول دارند (میزر، گامست و گارینو، ۱۳۹۱: ۱۱۱). در مجموع زمانی که پیش شرط‌های آزمون‌های چندمتغیره برقرار نباشد، باید داده‌های بدست آمده را تبدیل کنیم تا امکان استفاده از برخی آزمون‌های مدنظر فراهم شود.

در ابتدا باید میزان تخطی و تفاوت داده‌ها از مفروضه‌های ذکر شده را تعیین کرد و در صورتی که پیش فرض‌ها یا پیش شرط‌های آماری دارای تفاوت قابل‌اعتنایی با مقدار یا وضعیت مورد نظر باشند از روش تبدیل داده‌ها استفاده کرد. تبدیل داده‌ها با هدف تعدیل متغیرها از جنبه علمی روشی پذیرفته شده است و پژوهشگران و آماردانان بسیاری از آن

استفاده می کنند. البته زمانی که اختلاف داده ها با پیش فرض های آماری اندک باشد و به طور تقریبی مفروضات آماری برقرار باشد می توان از تبدیل داده ها صرف نظر کرد.

البته باید توجه داشت که تبدیل داده ها تا اندازه ای مانند شمشیر دولبه است. حسن این روش این است که می تواند دقت معنی داری تحلیل های آماری را افزایش دهد و عیب آن این است که ممکن است تفسیر داده ها را دشوارتر کند. در نتیجه باید از روش تبدیل داده ها به شیوه ای مدبرانه استفاده کرد. دشوار کردن تفسیر داده ها بدین معناست که وقتی داده ها را تبدیل می کنیم مقدار حداقل و حداکثر و شیوه توزیع متغیر و تمامی شاخص های میانگین و انحراف استاندارد تغییر می کند و با حالت معمول و عادی تفاوت پیدا می کند. مثلا اگر سن افراد که به صورت کمی (نسبی) سنجیده شده است را به توان دو برسانیم شاخص های آماری سن افراد تغییر می کند و با سن های غیر عادی مثل ۲۵۰ ، ۳۰۰ و غیره مواجه می شویم. 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 می دهد و توضیح و تفسیر متغیر را با مشکل مواجه می کند. یکی از راه های رفع این مشکل این است که هنگام گزارش یافته های توصیفی و هنگام بیان شاخص های آماری متغیرها (مانند میانگین، انحراف استاندارد و مقدار حداقل و حداکثر) یافته ها و شاخص های آماری را هم به صورت عادی (قبل از تبدیل داده ها) و هم بعد از تبدیل داده ها گزارش کنیم.

انواع روش های تبدیل داده

روش های مختلفی از تبدیل داده ها وجود دارد که برخی از روش های متداول عبارتند از: ریشه دوم، لگاریتم، وارون، مجذور X، بازتاب و ریشه دوم، بازتاب و لگاریتم و بازتاب و وارون. بین آمارشناسان در مورد استفاده از انواع روش های تبدیل داده ها در شرایط خاص اتفاق نظر وجود ندارد. با وجود این، یک راهبرد اساسی برای کاربست روش های تبدیل این است که برحسب جدی بودن تخطی از پیش فرض های آماری، راهبردهای پیشرفته تر (سطح بالاتر) به کار بسته شود. برای مثال، تاباچینک و فیدل (۲۰۰۱) و مرتلر و انااتا (۲۰۰۱) برای پیشرفت تبدیل داده ها از ریشه دوم (برای اصلاح تخطی متوسط)، لگاریتم (برای تخطی اساسی تر) و سپس ریشه دوم وارون (برای رسیدگی به تخطی شدید) جانبداری

می کنند. مجذور کردن^۱ یک متغیر در رابطه دو متغیری غیرخطی می تواند مشکل غیرخطی بودن را به طور مؤثر کاهش بدهد (میزر، گامست و گارینو، ۱۳۹۱:۱۱۳).

به طور کلی زمانی که داده ها نرمال نیستند از روش های ریشه دوم^۲ لگاریتم^۳ و وارونه کردن^۴ برای تبدیل داده ها استفاده می شود.

نوع مشکل در پیش فرض ها	نوع تبدیل پیشنهادی
نقض پیش فرض ها به طور کلی	ریشه دوم (برای اصلاح تخطی متوسط) لگاریتم (برای تخطی اساسی تر) ریشه دوم وارون (برای رسیدگی به تخطی شدید)
غیرخطی بودن رابطه	مجذور کردن
چولگی متوسط (مثبت یا منفی)	ریشه دوم
چولگی شدید (مثبت یا منفی)	لگاریتم (Log 10)

¹ Square transformation

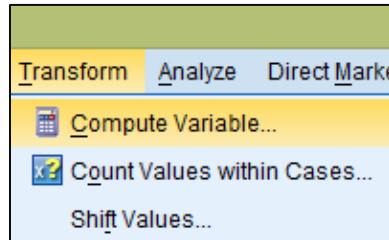
² Square root transformation

³ Logarithmic transformation

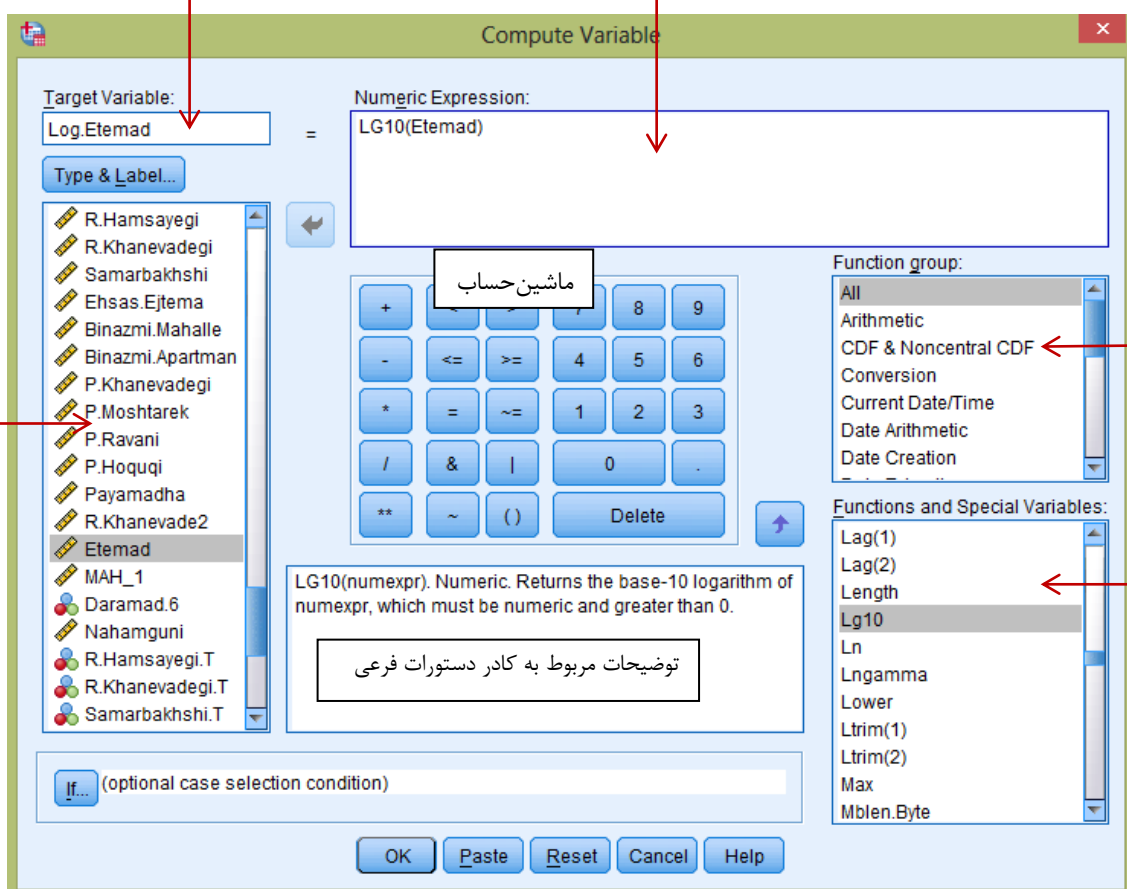
⁴ Inverse transformation

انجام تبدیل داده ها در SPSS

تبدیل داده ها در برنامه SPSS و از طریق دستور Compute انجام می شود. این دستور در منوی Transform قابل دستیابی است.



معرفی بخش های پنجره دستور Compute



The screenshot shows the 'Compute Variable' dialog box in SPSS. The 'Target Variable' is 'Log.Etamad' and the 'Numeric Expression' is 'LG10(Etamad)'. The variable 'Etamad' is selected in the list of variables on the left. The 'Function group' is set to 'All', and 'Lg10' is selected in the 'Functions and Special Variables' list. A calculator keypad is visible in the center, and a text box at the bottom explains the LG10 function.

کادر انتخاب نام متغیر

کادر اعمال دستورات

کادر دستورات اصلی

کادر دستورات فرعی

ماشین حساب

فهرست متغیرهای موجود

توضیحات مربوط به کادر دستورات فرعی


if... (optional case selection condition)

OK Paste Reset Cancel Help

۱- در پنجره ایجاد شده (Compute Variable) و در کادر Target Variable نامی برای متغیری که می

خواهیم بسازیم انتخاب می کنیم (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 می کنیم و با استفاده از دستور  آن را وارد کادر

Numeric Expression می کنیم. متغیر مورد نظر را از کادر متغیرها در سمت چپ انتخاب می کنیم و با

دستور  در داخل پرانتز کادر Numeric Expression قرار می دهیم و در انتها گزینه OK را می زنیم.

Numeric Expression:
LG10(Etemad)

۳- برای مجذور کردن یا به توان دو رساندن متغیر کافیت در کادر Numeric Expression متغیر مد نظر را از

کادر متغیرها انتخاب کرده و بین آن دو علامت * (ضرب) را از کادر ماشین حساب قرار دهیم و دوباره همان

متغیر را وارد کادر کنیم.

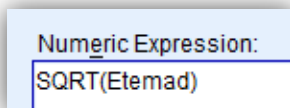
Numeric Expression:
Etemad * Etemad

۴- برای وارون یا معکوس کردن متغیر بایست 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:
1 / Etemad

۵- برای بدست آوردن ریشه دوم (جذر گرفتن) کفایست که گزینه SQRT را از کادر Function and... انتخاب می کنیم و مانند مورد لگاریتم عمل کنیم.



بعد از انجام دستور تبدیل متغیر (Compute)، متغیری جدید در صفحه داده ها ایجاد شده که در تحلیل از آن استفاده می کنیم.

Etemad	Log	Square.root	Inverse	Square
3.20	.51	1.79	.31	10.24
3.00	.48	1.73	.33	9.00
2.20	.34	1.48	.45	4.84
3.00	.48	1.73	.33	9.00
1.60	.20	1.26	.63	2.56
2.20	.34	1.48	.45	4.84
2.80	.45	1.67	.36	7.84
3.80	.58	1.95	.26	14.44
1.80	.26	1.34	.56	3.24
3.60	.56	1.90	.28	12.96
3.40	.53	1.84	.29	11.56
2.80	.45	1.67	.36	7.84
3.40	.53	1.84	.29	11.56
2.00	.30	1.41	.50	4.00
5.00	.70	2.24	.20	25.00

نکته: زمانی که شیوه تعریف داده ها طوری باشد که برخی از موردها با کد ۰ (صفر) نشان داده شده باشند، قبل از تبدیل داده ها بایست تغییراتی در داده خام ایجاد کرد. بدین شکل که به همه داده های خام متغیر مورد نظر عدد ۱ را اضافه می کنیم که حداقل نمره در داده ها برابر با ۱ باشد و بتوان به نتایج معنی دار و مناسبی از تبدیل داده ها دست یافت. برای افزودن عدد ۱ به همه داده ها کفایست 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:

Etemad + 1

منابع

میزر، لاورنس اس و گامست، گلن و گارینو، ا.جی. (۱۳۹۱) پژوهش چندمتغیری کاربردی (طرح و تفسیر)، ترجمه حسن پاشا شریفی و دیگران، تهران: رشد.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn and Bacon.

مرکز خدمات آماری خوارزمی

انجام تحلیل آماری پایان نامه کارشناسی ارشد و دکترا و مقالات ISI

با نرم افزارهای SPSS – LISREL – AMOS – PLS – Eviews و شبکه های عصبی با Matlab

ایمیل: RKarimi777@yahoo.com

سایت: www.kharazmi-statistics.ir

www.SPSS100.ir

رامین کریمی: ۰۹۱۲۷۶۹۴۰۶۶

مؤلف کتاب "راهنمای آسان تحلیل آماری با SPSS"