



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

یک رویکرد دو مرحله ای برای تبدیل متغیر های پیوسته به متغیر های

نرمال: اهمیت ها و توصیه ها برای تحقیقات IS

چکیده

این مقاله به توصیف یک رویکرد دو مرحله ای برای تبدیل متغیر های پیوسته با توزیع غیر نرمال به متغیر های با توزیع نرمال می پردازد. مرحله 1 شامل تبدیل متغیر به یک رتبه درصدی (صدک) می باشد که منجر به احتمالات با توزیع یکنواخت می شود. مرحله 2 از تبدیل نرمال معکوس برای نتایج اولین مرحله برای تشکیل یک متغیر متشکل از امتیازات Z توزیع نرمال استفاده می کند. این رویکرد در منابع و مطالعات به غیر از منابع آماری کم تر شناخته شده است و به ندرت در علوم اجتماعی مورد استفاده قرار گرفته است و هنوز در هیچ مطالعه IS مورد استفاده قرار نگرفته است. این مقاله نشان می دهد که چگونه بایستی رویکرد را در اکسل، SPSS، SAS پیاده کرده و سپس به توضیح اهمیت ها و توصیه هایی برای تحقیقات IS می پردازد.

کلمات کلیدی: مدل سازی ریاضی، رشته، بین رشته ای، تئوری مرجع، اهمیت تئوری

1-مقدمه

از دیرباز، تبدیلات داده ها (برای مثال، توان ولگاریتم) از طریق بهبود نرمال بودن و با استفاده از رویکرد آزمون و خطا انجام می شده اند. متأسفانه، به ندرت محققان به نرمالیت آماری از طریق تست های تشخیصی پذیرفته شده دست پیدا می کردند (تست های پذیرفته شده شامل کلموگروف-اسمیرنوف، P-P پلات، چولگی، کشیدگی بوده اند). این تحقیق به بررسی یک رویکرد ساده ولی قوی می پردازد که موسوم به رویکرد دو مرحله ای بوده و برای تبدیل بسیاری از متغیر های پیوسته با توزیع غیر نرمال به نرمالیت آماری (متغیر های با توزیع نرمال) استفاده شده است (این نشان می دهد که تست های تشخیصی مناسب برای نرمالیت وجود دارند). تبدیل پیشنهادی میتواند به چولگی، کشیدگی و یا آزمون نرمال بودن قابل قبول در بسیاری از شرایط دست پیدا کرده و موجب نرمال بودن در موارد دیگر می شود. به جز دو محدودیتی که

در ادامه گفته شده است، این رویکرد نرمال بودن توزیع متغیر حاصله را بهینه سازی می کند. روش دومرحله ای یک استاندارد ایده ال را برای تبدیل متغیر ها به نرمال بودن و یک دیدگاه جدید را در خصوص تحقیقات MIS در اختیار می گذارد.

در مطالعات مربوط به اثرات غیر نرمال بودن بر روی آزمون های همبستگی، تحقیقات قبلی از داده های شبیه سازی استفاده کرده اند (برای مثال، فیگلن 2009)، در حالی که روش دو مرحله ای پیشنهاد شده امکان استفاده از متغیر های مشاهده شده را خواهد داد. برای مثال، پارادوکس بهره وری، اصطلاحی است که ناتوانی سیستم های پیچیده را در کشف روابط بین طیف وسیعی از معیار های سرمایه گذاری فناوری اطلاعات و بهره وری سازمانی نشان می دهد. در همین رابطه، تعداد زیادی از تحقیقات علمی چند رشته ای برای درک بهتر مسیر های خاص نظیر رابطه بین سرمایه گذاری فناوری اطلاعات و عملکرد مالی ارایه شده اند (برینفلسون و هیت 2003). علی رغم تلاش های زیاد در رشته های مختلف، توجه کمی به مفهوم پارادوکس شده است و نکته جالب این است که مطالعات مربوط به این موضوع به ندرت ابعاد توزیعی داده ها را ذکر کرده اند. مطالعات شبیه سازی، که از داده های بدون تئوری برای مطالعه اهمیت و اثرات نرمال بودن استفاده می کنند، نمی توانند مستقیماً یک مسیر پارادوکس را توسعه دهند. بر عکس، روش دومرحله ای پتانسیل تبدیل متغیر های مشاهده شده را به نرمال بودن آماری و درک اثرات پایین دست را بر روی یافته های مطالعه نظیر اندازه اثر اصلی تبدیل می کند.

در میان ده ها توزیع عمومی موجود، توزیع نرمال دارای کاربرد بسیاری در تحقیقات کیفی می باشد. بسیاری از روش های آماری (رگرسیون چند گانه، تحلیل عاملی) مورد استفاده در تحقیقات کمی، به نرمال بودن حساس می باشند. برای مثال، وجود نرمال بودن، موجب بهبود تشخیص تفاوت های بین گروهی در هر دو مدل سازی معادلات ساختاری مبتنی بر مولفه و کواریانس شده است (کارشی و کامپیو 2009). بهبود نرمال بودن موجب کاهش ناهم واریانسی در pp پلات شده (هیر و همکاران 2010) و به این ترتیب موجب افزایش سطح همبستگی آماری بین دو متغیر شده است.

رویکرد پیشنهادی به بررسی چهار خلاء و شکاف در تحقیقات IS می پردازد. اولاً، همان طور که در بخش بعدی نشان داده شده است، مسئله نرمال بودن در مطالعات مورد بررسی قرار گرفته است. دوماً، رویکرد تبدیل دومرحله ای در این جا در همه تحقیقات IS تا کنون استفاده نشده است. در نتیجه محققان از این روش کم تر آگاه بوده اند. برای اولین بار، این راهنمای آموزشی، به بررسی IS پرداخته است. سوماً، روند تغییرات اخیر در زمینه رایانش، سنجش از دور و رایانش ابری موجب شده اند تا داده های زیادی برای بسیاری از سازمان ها و اعضای جامعه اطلاعات ارایه کند. افزایش دسترسی به داده ها موجب افزایش وابستگی جامعه به تحلیل این داده ها می شود. برای مثال، داده کاوی در حال افزایش بوده و این مسئله منجر به افزایش اهمیت آزمون های سببی و علی شده است. چهارماً، به دلیل دسترسی به دیتابیس های الکترونیکی جامع و کم هزینه، محققان به مسئله کاهش داده ها علاقه مند بوده اند. در نتیجه، مطالعات ساخت شاخص ترکیبی که متکی به نتایج تست های همبستگی بین متغیر های گروه بندی شده منطقی بوده اند، مهم تر است. اگرچه رویکرد دو مرحله ای در مطالعات استفاده کننده از داده های پیوسته اهمیت دارد، باین حال برای افرادی که از انجمن تحقیقات IS بین رشته ای استفاده می کنند سودمند است.

هدف این راهنمای آموزشی بررسی و شفاف سازی یک رویکرد تبدیلی است که به پیشبرد موضوعات مربوط به غیر نرمال بودن داده های پیوسته کمک می کند. اهمیت این مقاله، توصیف روش جدید و پتانسیل آن برای ارایه دیدگاه های جدید در تحقیقات IS می باشد. اگرچه هر مرحله به طور مجزا در علوم اجتماعی استفاده شده است، اصلیت این مقاله، مربوط به توصیف دو مرحله برای تبدیل متغیر های مشاهده شده به نرمال بودن است. به خصوص، الگوریتم معرفی شده در تحقیقات توصیف نشده است. از این روی، مقاله به عنوان یک ابزاری است که محققان IS از آن برای تعیین این که آیا قادر به بهبود دانش نظری و میزان پیشرفت علمی در رشته IS میباشند یا خیر استفاده می کنند.

این مقاله یک پیش زمینه ای از مفاهیم اساسی را ارایه کرده و به توضیح الگوریتم منطقی مورد استفاده محققان پرداخته و نیز به بررسی کاربرد آن در سه نرم افزار پرداخته و مثال هایی در رابطه با کاربرد آن بر

روی داده های مشاهده شده ارایه می کند. سپس این مقاله در مورد اهمیت آن برای رابطه IS، کاربرد روش و توصیه هایی برای محققان بحث کرده و در نهایت مروری بر نتیجه گیری نهایی دارد.

2- پیش زمینه

این بخش به توضیح زمینه و شرایط مطالعه می پردازد که شامل وضعیت فعلی تحقیقات مربوط به نرمال بودن در IS و سایر رشته های علوم اجتماعی است. اگرچه نرمال بودن یک مسئله مهم در IS و سایر رشته ها است، به طور کلی در مطالعات مربوطه مورد بررسی قرار نگرفته است. یک مرور منابع در مجله منابع کسب و وکار EBSCO Host برای یافتن همه مطالعات مرتبط با کسب و کار که بایستی به مسئله نرمال بودن پردازند انجام شد. همین قالب شامل مقالاتی است که با عبارات " عملکرد مالی شرکت (CFP) و مدل سازی معادله ساختاری SEM " جست و جو شده اند. شاخص های CFP به شدت از حد نرمال متفاوت هستند (بارنز 1982، دیکین 1976) و مطالعاتی که از این شاخص ها استفاده می کنند بایستی به بررسی میزان نرمال بودن در همه موارد پردازند. SEM یک روش تحلیلی حساس به فرض نرمال بودن است و مطالعاتی که از این رویکرد استفاده می کنند نیز بایستی به مسئله نرمال بودن پردازند. نتایج نشان می دهد که نرمال بودن به ندرت در مطالعاتی که ترکیبی از دو روش را استفاده کرده اند ذکر شده اند. از هفتاد و نه مقالاتی که از هر دو عبارت در متن استفاده کرده اند، دوازده مورد از عبارت چولگی، هشت مورد از عبارت کشیدگی و پانزده مورد از عبارت نرمال بودن استفاده کرده اند. این تحلیل نشان می دهد که محققان در مقالاتی که باید به بررسی مسئله نرمال بودن پرداخته اند و بسیاری از آن ها با داده های با توزیع غیر نرمال کار میکنند. به علاوه، عدم گزارش تبدیلات متغیر در تحلیل در مقالات نهایی را گزارش نکرده اند. (ماستی 1998).

ارزیابی منابع IS و مدیریت راهبردی نشان دهنده بار نرمال بودن فعلی است که پیش روی محققان کمی است. برای مثال، سعید و همکاران (2005) رابطه بین قابلیت تجارت الکترونیک و cfp را تست کرده اند. آن ها دو متغیر CFP را با استفاده از تبدیل باکس- کاکس برای رسیدگی به مسئله ناهم واریانسی در تست های همبستگی مدل (ارزش افزوده اقتصادی و Q توبین) تبدیل کردند. محققان هیچ گونه تشخیص نرمال بودن

قبل و پس از تبدیل را گزارش نکردند. در نتیجه، اطلاعات کمی در خصوص شیوه بهبود ناهم واریانسی در تست های همبستگی مدل توسط این تغییرات در توزیع وجود داشته است و مطالعات کمی به بررسی اندازه اثر و تبدیل متغیر ها به سمت نرمال بودن آماری پرداخته اند. یک مثال دیگر، کای و وانگ (2009) می باشد که به آزمون رابطه بین روابط ذی نفعان و CFP پرداخته است و با نرخ بازده و Q توبین نشان داده شده است. در نهایت، تنریوردی (2005) همبستگی بین ابعاد مدیریت دانش به واسطه فناوری اطلاعات و CFP را بررسی کردند (نرخ بازده دارایی و Q توبین). علی رغم استفاده از SEM، ترانریوردی (2005) تلاشی برای تبدیل CFP و نه تشخیص نرمال بودن نکرده است. اگرچه نتایج آزمون همبستگی معنی دار بوده است، تلاشی برای بهبود یافته ها از طریق افزایش نرمال بودن صورت نگرفته است. هر یک از سه مثال فوق شامل نمونه های قوی و الگو می باشند. مثال ها نشان دهنده این هستند که نرمال بودن با چه شیوه هایی در مطالعات بررسی شده اند. آن ها وضعیت مواجهه محققان با عدم قطعیت را در خصوص شیوه تهدید نتایج، قدرت آماری و اطمینان پذیری و پایایی مطالعات کمی توسط محققان را نشان داده اند.

3- الگوریتم تبدیل

این بخش اطلاعاتی در مورد منشا، منطق و محدودیت های الگوریتم تبدیل دو مرحله ای ارائه می کند. سپس بخش پایانی در مورد تفاوت آن با سایر روش های رایج در تحقیقات IS بحث می کند.

منشاء

محققان و متخصصان رشته های مختلف از دیرباز از تابع توزیع برای تولید متغیر های تصادفی برای مطالعات شبیه سازی (یان و چن 2009) و ارزیابی برآزش توزیعی متغیر های اصلی استفاده کرده اند (فیشمن 2003). متاسفانه، اصلی ترین داده های پیوسته از پدیده های جهان واقعی دارای توزیع دلخواه می باشند. یعنی، داده ها از نظر آماری منطبق بر یکی از توزیعات اصلی ای است که توسط تابع توزیع تجمعی ایجاد شده است (نرمال، کای اسکوتر و پارتو، F). وقتی که تابع توزیع تجمعی معکوس شود (تابع توزیع تجمعی معکوس)، آن ها را می توان برای محاسبه اعداد تصادفی منطق با توزیع عمومی استفاده کرد. برای مثال، برای تولید متغیر های تصادفی با توزیع نرمال، محققان شبیه سازی، احتمالات تصادفی (از 0 تا 1) را

تولید می کنند که با استفاده از CDF نرمال معکوس تبدیل می شوند. بر این اساس، رویکرد پیشنهادی درک شده است ولی ریاضی مسائل جانبی مختلف و اهمیت ها، بین تولید متغیر نرمال تصادفی و تبدیل دو مرحله ای نسبت به نرمال بودن متفاوت بوده است.

یک توزیع عمومی مهم، توزیع یکنواخت است که با متغیر های تولید شده توسط تابع اعداد تصادفی همراه است. توزیع یکنواخت قبل از تبدیل موفق به توزیع پیوسته عمومی فرض می شود. همان طور در زیر نشان داده شده است، تبدیل متغیر های مشاهده شده به سمت یکنواختی شامل استفاده از تابع توزیع معکوس نیست.

منطق

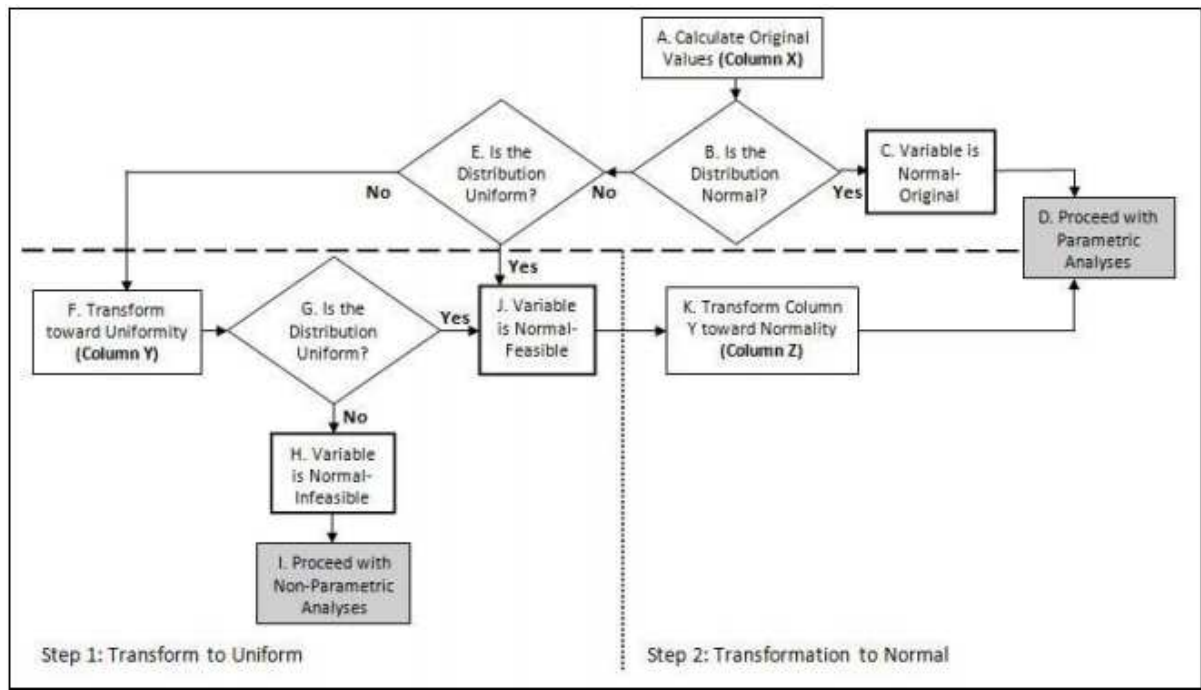
شکل 1 الگوریتم دو مرحله ای (روش A-K) را برای تبدیل متغیر های پیوسته با توزیع اختیاری به توزیع نرمال نشان می دهد. این رویکرد همه متغیر ها با استفاده از سه مقوله را طبقه بندی می کند

1- نرمال - اصلی: یک مورد بسیار نادر که در آن مقادیر اصلی و اولیه از نظر آماری نرمال هستند (یعنی منطبق بر کثرت تشخیص های نرمال بودن است)

2- نرمال - عملیک موردی که در آن متغیر اصلی غیر نرمال است ولی به نرمال قابل تبدیل است

3- نرمال - غیر ممکن: موردی که در آن متغیر اصلی غیر نرمال بوده و غیر قابل تبدیل به نرمال است.

این طرح طبقه بندی به این دلیل است که محاسبه روش دو مرحله ای منجر به این می شود تا متغیر پیوسته تبدیل شده به سطح نرمال برسد. این روش از نظر دو ویژگی توزیعی توصیف شده در بخش محدودیت ها، محدودیت دارد. در بسیاری از موارد، این رویکرد موجب می شود تا نتایج تبدیل به سمت نرمال بودن آماری میل کنند (یعنی متغیر منطبق بر فراوانی تست های تشخیص نرمال مناسب است) و یک استاندارد عالی برای IS و علوم دیگر است. محققان از این الگوریتم استفاده کرده و از روش دو مرحله ای برای اهداف کشف بهره برده اند (برای مثال، شناسایی ویژگیهای توزیعی که موجب محدود شدن کارایی آن می شوند).



شکل 1 : یک الگوریتم دومرحله ای برای تبدیل متغیر های پیوسته به متغیر های نرمال

مرحله 1: تبدیل به یکنواختی

اولین مرحله شامل تبدیل متغیر اصلی به یکنواختی آماری با محاسبه رتبه درصدی هر امتیاز است. در تلاش

برای کمک به قابلیت تفسیر، سه مدل از متغیر ها استفاده می شوند

1- ستون X: متغیر اصلی

2- ستون Y: نتیجه تبدیل ستون X به احتمالات یکنواخت (مرحله 1)

3- ستون Z: نتیجه تبدیل ستون Y به مقادیر با توزیع نرمال (مرحله 2)

رویکرد با محاسبه مقادیر اصلی و اولیه برای متغیر پیوسته شروع می شود (با روش A در شکل 1) که اشاره

به ستون X دارد. اگر تست های تشخیصی نشان دهند که ستون X نرمال است (روش B) ، به صورت نرمال-

اصلی (روش C) طبقه بندی شده و محقق تست های آماری پارامتری را انجام می دهد (روش D). اگر تست

نشان دهد که ستون X نرمال نیست محقق باید از تست های تشخیصی برای یکنواختی استفاده کند. در

صورتی که ستون X نرمال یکنواخت باشد، به مرحله 2 برای تبدیل به مقدار نرمال وارد می شود. اگر

یکنواخت نباشد، ستون X به یکنواخت با استفاده از تابع رتبه صدک محاسبه می شود (روش F، که منجر به

ستون Y می شود). یک فرمول اولیه برای درصد رتبه، که منجر به مقادیر از 0 تا 1 می شود به صورت زیر است

رتبه درصدی:

رتبه X_i : رتبه مقدار

N = اندازه نمونه

برای مثال، شرکتیک نرخ سود سالانه 1.3 را گزارش می کند که در میان 100 شرکت رتبه هفتم را دارد. رتبه صدک $93 = (7/100) \times 100$ است و به این معنی است که 93 درصد مشاهدات نمونه دارای نرخ سود کم تر از شرکت اصلی است. دستیابی به یکنواختی آماری پیش نیاز رسیدن به نرمال بودن آماری با استفاده از روش دو مرحله ای است. از این روی اگر ستون Y یکنواختی آماری رانشان دهد، متغیر اصلی به صورت نرمال - عملی طبقه بندی می شود و محقق باید روش های آماری غیر پارامتری را در نظر بگیرد (روش 1)

مرحله 1 یک گام مهم می باشد زیرا دست یابی به یکنواختی آماری قبل از نرمالیت آماری در مرحله 2 لازم است. برخی شرایط امکان دست یابی به یکنواختی آماری را نمی دهند که بر طبق هنجارهای علوم اجتماعی یک استاندارد بسیار بالاست. اگر مرحله 1 قادر به دست یابی به یکنواختی آماری نباشد محقق دارای چند گزینه است

1- تحمل کم تر از استاندارد یکنواخت آماری و وارد شدن به مرحله 2 (اگرچه به استاندارد نرمال بودن آماری نمی رسد)

2- استفاده از رویکرد تبدیل از مون و خطا برای بهینه سازی یکنواختی و وارد شدن به مرحله 2

3- در نظر گرفتن همبستگی و رابطه با ابعاد چند کارکردی و تفکیک نمونه ها. یک بخش از نمونه عملکرد بهتری از بخش های دیگر دارد. مرحله 1 را با بخش های قابل اصلاح اصلاح کنید.

4- تنها در صورت امکان، شیوه از بین رفتن مقادیر صفر در نظر گرفته شده و مرحله 1 تکرار می شود

مرحله 2: تبدیل به نرمال بودن (از یکنواختی)

هر متغیر منطبق با یکنواختی آماری ، نرمال - امکان پذیر است. احتمالات یکنواخت (ستون Y) را می توان به مقدار نرمال (روش K) با تابع توزیع نرمال معکوس نشان داده شده در 2 تبدیل کرد

$$p = \mu + \sqrt{2} \sigma \operatorname{erf}^{-1}(-1 + 2\operatorname{Pr})$$

که

امتیاز Z حاصل از مرحله 2	P
میانگین P (توصیه برای امتیاز های Z استاندارد صفر است)	μ
انحراف معیار (توصیه برای امتیاز Z استاندارد 1 است)	σ
تابع خطای معکوس	erf^{-1}
احتمال اینکه ناشی از مرحله 1 است	Pr

برای مثال، یک محقق می خواهد تا نتایج مرحله 1 (ستون Y) رابه یک متغیر مطابق با توزیع نرمال تبدیل کند (ستون Z). سه پارامتر برای تابع معکوس نرمال لازم هستند 1- احتمال 2- مقدار مورد انتظار متغیر حاصله (ستون Z) و 3- انحراف معیار مورد انتظار ستون Z. محققان تنها مقادیر Q و سیگما را برای نزدیک شدن به مقدار نرمال در نظر می گیرند. برای تولید امتیازات Z، محقق از Pr = نام متغیر، $\mu = 0$ و $\sigma = 1$ به صورت پارامتر استفاده می کند یک مقدار 0.25 به امتیاز Z-1.96 تبدیل می شود.

همان طور که در بالا گفته شد، هر یک از این دو مرحله در منابع آماری مختلف و ابزار های آماری تحلیلی توضیح داده شده است. علوم اجتماعی از هر دو مرحله برای تبدیل متغیر های مشاهده شده چشم پوشی کرده اند.

محدودیت ها:

دو محدودیت کارایی در خصوص روش دو مرحله ای وجود دارد. روش دو مرحله ای با در نظر گرفتن این دو محدودیت ، تبدیلاتی را با ویژگی های نرمال بودن کامل تولید کرده است. محدودیت ها به صورت زیر هستند 1- وجود تعداد کم سطوح 2- وجود شیوه های موثر

تعداد سطوح

رویکرد تا جایی موفق است که متغیر اصلی پیوسته باشد (به این معنی که بتوان احتمال تجمعی معنی دار را بدست آورد). از این روی، نظم میان سطوح لازم است و انواع داده های اسمی را نمی توان به طور منطقی به توزیع نرمال با رویکرد تبدیل کرد. انواع داده های فاصله ای و ترتیبی با تعداد زیادی از سطوح موفق بوده اند و داده های نسبت قابل اصلاح به تبدیلات موفق با استفاده از رویکرد است. به عنوان مثال، رویکرد اثر بسیار کمی بر روی متغیر های بیان شده با پاسخ لیکرت 5 و 7 نقطه ای دارد. برای بهبود ویژگی نرمال بودن این مطالعات، دانشمندان علوم اجتماعی بایستی قادر به طراحی مقیاس ها با سطوح بیشتر باشند. تحقیقات بیشتر در زمینه کاربرد این مقیاس ها آن قدر مهم است که نرمال بودن یک فرض مهم در روش های آماری است.

اثر حالت ها و شیوه ها

هنگام استفاده از فنون، محققان بایستی از فراوانی و اثر ارزش ها و مقادیر حالت آگاه باشند. متغیر ها از نظر موقعیت و فرکانس نسبی حالت متنوع هستند. در همه زمینه های تحقیقاتی، محققان باید حذف نمونه های بی معنی را در مطالعات علی معلولی در نظر بگیرند (شادیش و همکاران (2002)). هم چنین محققان بایستی اطمینان حاصل کنند که تنها مقادیر مربوط به تئوری تست شده در تحلیل حفظ میشوند. اگر حالت ها منجر به تغییر نتایج شوند، محققان بایستی به بررسی و جایگزینی مقادیر حالت و مقادیر مفقود بپردازند. یک رابطه کلی بین تعداد سطوح و اثر حالت وجود دارد. بر اساس تعریف، یک متغیر با تعداد کم سطوح دارای حالت های موثر بالا است. یک متغیر با تعداد زیاد سطوح ممکن است فاقد یکی از حالت های موثر باشد. تجربه نشان می دهد که روش دو مرحله ای در تبدیل متغیر ها با هر دو تعداد بالای سطوح و با سطح کم شیوه های فرکانس بالا موثر تر است. به عنوان مثال، داده های اقتصادی و نهایی نسبت به روش مربوطه قابل اصلاح بوده و این حالی که داده های دودویی (نظیر جنسیت) قابل اصلاح نیست.

تمایزات

روش دو مرحله ای از تولید و استاندارد سازی عدد تصادفی متمایز است. کاربران رویکرد دو مرحله ای به طور صحیح بیان داشته اند که تولید عدد تصادفی از دو مرحله استفاده کرده و دومین مرحله، مشابه با مرحله مورد استفاده در تولید عدد تصادفی است. با این حال، هدف و محاسبه برای مرحله 1، عامل تمایز دو رویکرد است. در طی تولید عدد تصادفی، مرحله اول شامل تولید احتمالات با توزیع یکنواخت است. این عملیات در روش دو مرحله ای رخ می دهد. تولید عدد تصادفی بدون مقدار توصیف می شود، در حالی که روش دوم مرحله ای برای مقادیر مشاهده شده استفاده می شود. بر این اساس، هدف تولید عدد تصادفی، تولید متغیرهای نرمال تصادفی است که به یکنواختی آماری پس از اولین مرحله نزدیک می شوند. این موضوع در خصوص رویکرد دوم مرحله ای صادق نیست و به یکنواختی آماری پس از یک مرحله دست پیدا می کند. محققان باید مد نظر داشته باشند که روش دو مرحله ای متفاوت از استاندارد سازی است، برای هدف محاسبه امتیازات Z برای همه مقادیر در نظر گرفته می شود. هدف استاندارد سازی، مفاهیم 1- قابل مشاهده سازی متغیرها با همه واحدهای اندازه گیری 2- بیان دقیق توزیع اصلی می باشد. رویکرد دوم مرحله ای به استاندارد سازی واحد و واحد امتیاز دست پیدا می کند. از این روی، محققان استفاده کننده از روش دو مرحله ای توصیف می شوند.

کاربرد در نرم افزار آماری

هر دو مرحله لازم برای کامل کردن تبدیل، به آسانی در بسته های نرم افزاری آماری مدرن قابل دسترس است. این بخش، دستوراتی در سه ابزار تحلیلی ارابه می کند: اکسل، SPSS، SAS. هر مثال از متغیر CitationsPerPublication به عنوان مثال استفاده کرده و نتایج تبدیل در ادامه بررسی شده است.

اکسل

برای انجام مرحله 1 در اکسل، از دستور PERCENTRANK استفاده کنید که دارای تعاریف استدلال و نحو زیر است

PERCENTRANK (سری های اولیه، مقدار اصلی)

که

سری های اصلی = ارجاع به متغیر اصلی که قرار است تبدیل شود

مقدار اصلی: سلول، یا مقدار تکین که قرار است تبدیل شود.

با توجه به سومین استدلال، PERCENTRANK تولید احتمالاتی با استفاده از سه رقم می کند. استفاده از این دستور در ستون با مرحله 1 در شکل 2 نشان داده شده است. یک مثال نشان می دهد که منطق برای اجتناب از سه مسئله بالقوه در تبدیل لازم بود.

1- اولین دستور IF برای اجتناب از کاربرد تابع به یک سلول خالی است که در غیر این صورت منجر به بروز خطا می شود

2- دومین دستور IF برای جایگزینی 1 با 0.9999 لازم است. در غیر این صورت خطا در طی دو مرحله دو رخ می دهد

3- هم چنین، سومین دستور IF مقدار 0 را با 0.0001 جایگزین میکند. یک خطا در طی دومین مرحله رخ می دهد

در موارد 2 و 3 فوق جایگزینی بسیار کم تر از حذف است و امکان حفظ همه مقادیر متغیر را می دهد

=IF(C2="", "",
IF(PERCENTRANK(\$C\$2:\$C\$451,C2)=1,0.9999,
IF(PERCENTRANK(\$C\$2:\$C\$451,C2)=0,0.0001,
PERCENTRANK(\$C\$2:\$C\$451,C2))))

Original	Step 1	Step 2
153.30303030303	=IF(C2="", "", IF(PERCENTRANK(\$C\$2:\$C\$451,C2)=1,0.9999, IF(PERCENTRANK(\$C\$2:\$C\$451,C2)=0,0.0001, PERCENTRANK(\$C\$2:\$C\$451,C2))))	=IF(D2="", "", NORMINV(D2,0,1))
211.411764705882	=IF(C3="", "", IF(PERCENTRANK(\$C\$2:\$C\$451,C3)=1,0.9999, IF(PERCENTRANK(\$C\$2:\$C\$451,C3)=0,0.0001, PERCENTRANK(\$C\$2:\$C\$451,C3))))	=IF(D3="", "", NORMINV(D3,0,1))
66.3214285714286	=IF(C4="", "", IF(PERCENTRANK(\$C\$2:\$C\$451,C4)=1,0.9999, IF(PERCENTRANK(\$C\$2:\$C\$451,C4)=0,0.0001, PERCENTRANK(\$C\$2:\$C\$451,C4))))	=IF(D4="", "", NORMINV(D4,0,1))
269.238095238095	=IF(C5="", "", IF(PERCENTRANK(\$C\$2:\$C\$451,C5)=1,0.9999, IF(PERCENTRANK(\$C\$2:\$C\$451,C5)=0,0.0001, PERCENTRANK(\$C\$2:\$C\$451,C5))))	=IF(D5="", "", NORMINV(D5,0,1))
20	=IF(C6="", "", IF(PERCENTRANK(\$C\$2:\$C\$451,C6)=1,0.9999, IF(PERCENTRANK(\$C\$2:\$C\$451,C6)=0,0.0001, PERCENTRANK(\$C\$2:\$C\$451,C6))))	=IF(D6="", "", NORMINV(D6,0,1))
325.142857142857	=IF(C7="", "", IF(PERCENTRANK(\$C\$2:\$C\$451,C7)=1,0.9999, IF(PERCENTRANK(\$C\$2:\$C\$451,C7)=0,0.0001, PERCENTRANK(\$C\$2:\$C\$451,C7))))	=IF(D7="", "", NORMINV(D7,0,1))

=IF(D2="", "", NORMINV(D2,0,1))

شکل 2: استفاده از روش دو مرحله ای در اکسل

برای انجام مرحله 2 در اکسل، از دستور NORMINV() استفاده کنید که دارای عبارت نحوی زیر است

(NORMINV) (نتیجه مرحله 1، میانگین تحمیل شده و انحراف معیار) که

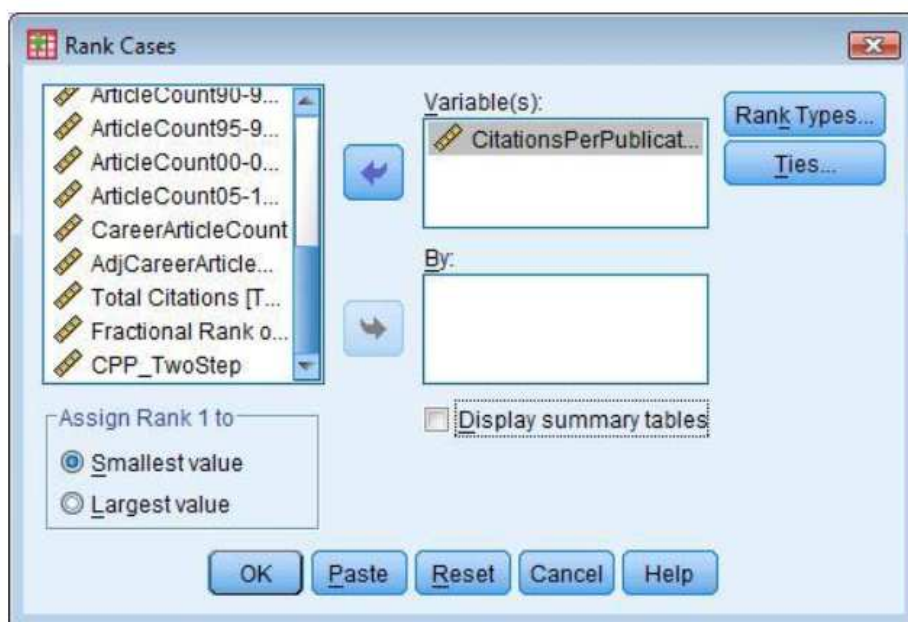
نتیجه مرحله 1: نتیجه مرحله 1 که بایستی در شکل میانگین احتمال = میانگین متغیر حاصل از تبدیل و انحراف معیار باشد

استدلال های میانگین و انحراف معیار از این نظر دلخواه هستند که اثری روی شکل توزیع ندارند. برای استاندارد سازی واحد های امتیاز Z ، مقدار میانگین برابر با 0 و انحراف معیار 0 تنظیم می شود استفاده از این دستور در ستون مرحله 2 در شکل 2 نشان داده شده است. تابع IF برای اجتناب از کاربرد دستور برای هر سلول خالی لازم است که در غیر این صورت منجر به بروز خطا می شود. این زمانی رخ می دهد که نتیجه مرحله 1، یک سلول خالی باشد

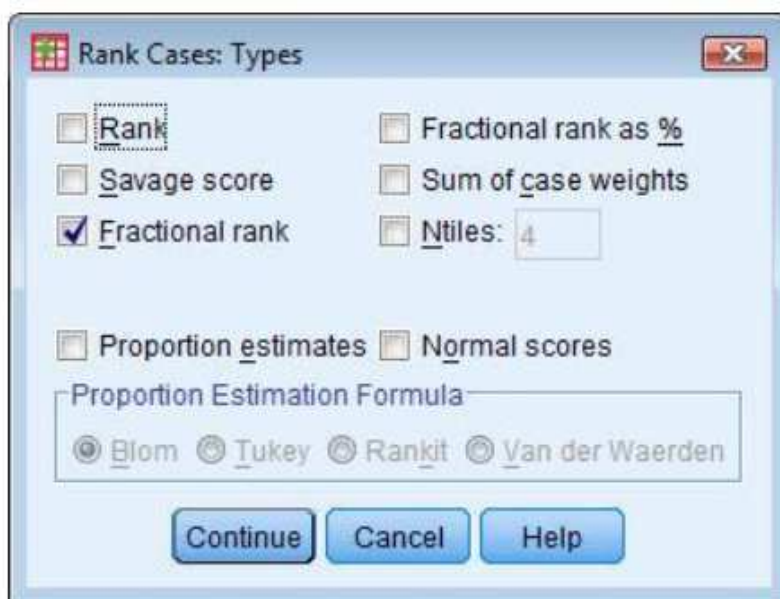
SPSS

مرحله 1 در SPSS با انتخاب Rank cases (نمونه های رتبه ای) از گزینه transform در منوی اصلی انجام می شود. پس از انتخاب متغیر و حرکت به سمت باکس، که در شکل 3 نشان داده شده است. دستور انواع رتبه را انتخاب کنید. در فرم موارد رتبه ای، مقادیر اصلی با نتایج تبدیل مرتبط نمی باشند. همان طور که در شکل 4 نشان داده شده است، دستور رتبه را حذف کنید و رتبه فرکشنال را انتخاب کنید. دوکمه کم ترین مقدار بایستی برای رتبه 1 انتخاب شود. انتخاب دکمه دستور Ok تولید مقادیری برای یک متغیر جدید می کند که در هر دو نمای داده ها و نمای متغیر قابل مشاهده است. کاربران بایستی سه احتیاط را در نظر بگیرند 1- نوع متغیر ورودی بایستی برای عملیات عددی باشند 2- مقادیر اصلی مفقود منجر به مقادیر تبدیل شده مفقود می شود 3- جایگزینی 0 و 1 پس از مرحله 1 بایستی به طور دستی در SPSS انجام شود. کد SPSS برای مرحله 1 در زیر نشان داده شده است

```
GET  
FILE='FileName.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.  
RANK VARIABLES=CitationsPerPublication (A)  
/RFRACTION  
/PRINT=NO  
/TIES=MEAN.
```



شکل 3: پنجره فرم نمونه های رتبه ای در SPSS



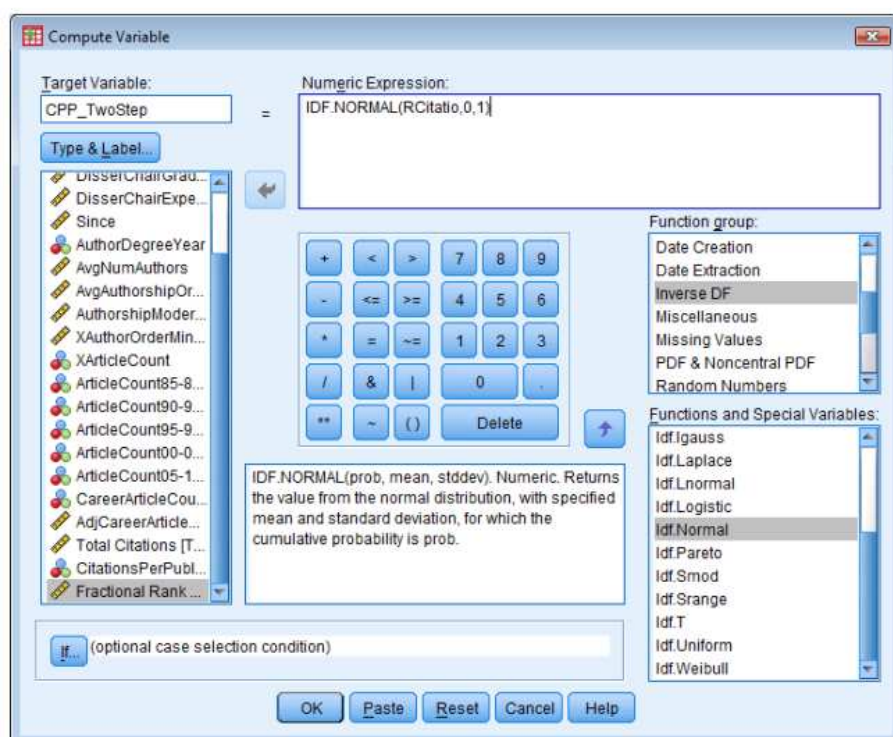
شکل 4: پنجره فرم انواع نمونه رتبه ای در SPSS

مرحله 2 شامل انتخاب TRANSFORM از منوی اصلی و سپس Compute Variable است. از قسمت Compute Variable (شکل 2)، DF معکوس را به صورت Function Group انتخاب کرده و Idf.Normal را از بخش Special Variables و Functions برای ایجاد یک تابع (دستور) انتخاب کنید. اولین پارامتر، حاصل مرحله 1 است و دومین مورد میانگین مطلوب است و سومین مورد انحراف معیار

مطلوب است. برای سهولت تفسیر، ما از 0 به عنوان استدلال اول و 1 به عنوان استدلال سوم استفاده می کنیم. این امتیازات Z استاندارد سازی شده را تولید می کند که به مقدار نرمال تبدیل می شود. در زیر کد

SPSS برای تکمیل مرحله 2 وجود دارد

```
COMPUTE CPP_TwoStep=IDF.NORMAL(RCitatio,0,1).  
EXECUTE.
```



شکل 5: پنجره فرم متغیر محاسباتی در SPSS

SAS

از PROC RANK با گزینه FRACTION برای تکمیل مرحله 1 در SAS استفاده کنید

```
PROC RANK  
DATA=inputfilename  
OUT=outputfilename  
FRACTION  
TIES=MEAN;  
RANKS step1result,  
var step1result;  
RUN;
```


کاربران بایستی به راهنمای SAS برای استفاده از کد برای شرایط موجود مراجعه کنند. برای مثال، گزینه DECENDING را می توان برای معکوس کردن ترتیب مقادیر استفاده کرد (بالاترین مقدار دارای رتبه 1 است).

برای کامل کردن مرحله 2، از دستور PROBIT استفاده کنید که از احتمال یکنواخت به مقدار نرمال تبدیل می شود

```
step2result=PROBIT(step1result);
```

مثال های گویا

شیوه دو مرحله ای در همه انواع متغیر های پیوسته برای کاربرد های متغیر از کشف تا دس یابی به نرمال بودن کامل، سودمند است. برای نشان دادن شیوه استفاده از الگوریتم در شکل 1، مجموعه داده ها از جمله دامنه متنوع ویژگی های متغیر پیروی می کنند. محقق از مجموعه داده جمع اوری شده در زمینه شاخص های علمی استفاده می کند. این می تواند مجموعه های از 450 عضو هیئت علمی در دستور العمل هیئت علمی AIS باشد که از 1985 تا 1990 فارغ التحصیل شده اند. مجموعه داده ها شامل متغیر های مربوط به تجربه های نویسندگی و عملکرد شغلی می باشد و دو متغیر لحاظ شده است: 1- استنادات به ازای هر مقاله 2- استنادات کل. استناد به ازای هر مقاله، تعداد استنادات به مقالات هر نویسنده در مجله تقسیم بر تعداد مقالات می باشند. بر همین اساس، تعریف این متغیر ها، کم اهمیت تر از ویژگی های توزیعاتی است که بر امکان سنجی نرمال بودن اثر دارد. استنادات به ازای هر مقاله به اسانی قابل تبدیل به نرمالیته است و استنادات کل تبدیل به مقدار نرمال نمی شوند. این بخش در نهایت به بررسی اهمیت رسیدگی به مقادیر با فراوانی بالا با استفاده از الگوریتم (در شیوه A از شکل 1) میپردازد.

جدول 1 معیار های مورد استفاده برای ارزیابی نرمال بودن و یکنواختی را برای همه سه نسخه متغیر ها تعریف می کند. توجه کنید که بسته به ابزار نرم افزار، بالاترین و کم ترین مقدار در متغیر به 0 تا 1 در پایان مرحله 1 تبدیل می شود. این مقادیر به صورت ورودی تابع نرمال معکوس مجاز نیست. به طوری که هیچ

یک از این مقادیر در تبدیل از بین می روند و توصیه می شود تا 0 با 0.001 و با 1 با 0.9999 پس از مرحله 1 جایگزین شود. از این روی مرحله 2 طوری کامل می شود به طوری که همه موضوعات حفظ شده و اثرات حداقل بر روی شکل توزیعی و نتایج آزمایشی در نظر گرفته می شود

جدول 1: معیار های آزمایشی و تعاریف

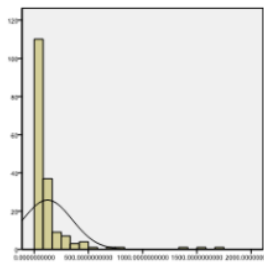
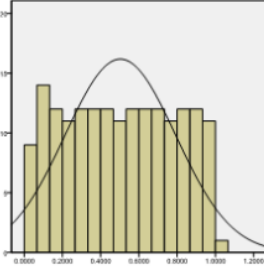
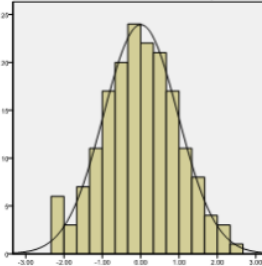
معیار ها	تعریف	شیوه تفسیر
مقدار P چولگی	چولگی درجه و جهت عدم تقارن است. مقدار P چولگی احتمال این است که اماره چولگی کم تر از مقدار مشاهده شده برای آن متغیر باشد	مقدار $P > 0.5$ (مقادیر منفی فرین) یا $0.95 > P$ (مقادیر مثبت) نشان دهنده انحراف معنی دار از 0 است (کشیدگی یا چولگی)
مقدار P کشیدگی	کشیدگی درجه انحراف اندازه دامنه توزیع از حد نرمال است. مقدار P کشیدگی احتمال این است که اماره کشیدگی کم تر از مقدار مشاهده شده باشد	
معنی داری کولوموکروف-اسمیرنوف، تست نرمال بودن	تست هایی برای تعیین این که آیا توزیع متغیر تفاوت معنی داری از توزیع نرمال دارد یاخیر	مقدار P بزرگ تر از 0.5 نشان می دهد که توزیع، توزیع آزمایشی است (نرمال یا یکنواخت)
معنی داری کولوموکروف-اسمیرنوف، تست یکنواختی	تست هایی برای تعیین این که آیا توزیع متغیر تفاوت معنی داری از توزیع یکنواخت دارد یاخیر	

تحلیلی که منجر به نرمال - امکان سنجی می شود

جدول 2 نسخه های استناد به هر مقاله را در زمان ورود به سه مرحله نشان میدهد: 1- متغیر اصلی 2- تبدیل به یکنواختی 3- تبدیل به نرمالیت. محاسبه مقدار اصلی (روش A در شکل 1) منجر به توزیع اصلی شده است (ستون اصلی را در جدول 2 ببینید). تست ها (روش B در شکل 1) نشان داد که متغیر اصلی دارای چولگی و کشیدگی معنی دار و انحراف معنی دار از نرمالیت بر طبق تست K-S است (مقدار

$P=0.000$). چون متغیر از نظر آماری نرمال است، تست K-S برای یکنواختی مشاهده می شود (روش E در شکل 1) و متغیر انحراف معنی داری از یکنواختی دارد. از این روی متغیر به سمت یکنواختی تبدیل شده (روش F، شکل 1) و منجر به توزیعی می شود که در ستون مرحله 1 در جدول 2 نشان داده شده است. تست بعدی (G) نشان داد که متغیر به طور معنی داری یکنواخت است ($P=0.987$) لذا متغیر به صورت نرمال - عملی طبقه بندی شده است. (روش J). متغیر به مقدار نرمال (روش K) بر طبق تست K-S ($P=200$) تبدیل شد و تحلیل پارامتری انجام شد (روش D).

جدول 2: سه نسخه از استنادات به ازای هر مقاله به صورت تبدیل شده با روش دو مرحله ای

نسخه	اصلی	مرحله 1 (رتبه ای فرکشنال)	مرحله 2 (نرمال معکوس)
Histogram			
N	176	176	176
چولگی p-value	1.000	.502*	.580*
کشیدگی p-value	1.000	.001	.183**
نرمالینه K-S - 1	.000	.047	.200***
یکنواختی K-S -	.000	.987****	.000

این مثال بر دو بعد از مد که اثر منفی بر نتایج تبدی میگذارند تاکید دارد. اولاً، درصد نسبتاً بالایی از مقادیر مد به طور منفی بر موفقیت تبدیل اثر دارد. محققان بایستی از مقادیر بالای مد در متغیرهایی که تلاش می کنند تا آن را با استفاده از روش تبدیل کنند آگاه باشند. دوماً، مد ها از میانگین انحراف دارند. هرچه مقدار مد در مقدار اصلی استنادات کل نزدیک تر به میانگین باشد، کارایی بهتر خواهد بود. (به مرحله 1 در جدول 3 مراجعه کنید). با این حال، حتی اگر مقدار مد برابر با میانگین باشد، اثر منفی بر کارایی روش دو مرحله ای دارد.

بحث

بخشهای زیر بر انجام تحقیقات آینده متمرکز بوده و اهمیت رشته IS و توصیه هایی را برای استفاده از روش تبدیل دو مرحله ای پیشنهادی بحث می کند

اهمیت این روش برای موضوعات علمی

انتظار می رود که این روش برای مطالعات IS ادرسه سطح اهمیت داشته باشد: اولاً، این روش موجب تغییر گزینه های روش شناختی، برابند ها و ارزش یافته ها در مطالعات می شود. این روش برای ساده سازی تبدیل به سمت نرمال اهمیت دارد ضمن این که از طیف وسیعی از گزینه ها استفاده می کند. به علاوه، اهمیت رویه های آماری برای تست های همبستگی بستگی به توزیع متغیر ها دارد. به طور کلی، تبدیلات- به نرمال در میان متغیر های مطالعه موجب بهبود قدرت آماری و خطای نوع 1 می شود: اثری که به نوبه خود موجب کاهش اندازه نمونه مورد نیاز می شود. بهبود نرمال بودن موجب افزایش قابلیت کاربرد کران ها و حدود اطمینان شده و نرمال بودن باقی مانده ها را در رگرسیون خطی بهبود می بخشد. در تحلیل عاملی، تبدیلات به نرمال موجب افزایش کیفیت نمونه برداری می شود. بهبود نرمالیتیه در مطالعات موجب تغییر کل واریانس توجیه شده، سادگی، معرف بودن و پایداری ساختار های عاملی میشود.

دوما این روش موجب تغییر نرخ پیشرفت در مسیر های تحقیقاتی میشود. با کاهش ناهم واریانسی در تست های همبستگی، این روش موجب بهبود قدرت اثرات اصلی می شود. در برخی از موارد، این بر درک و دانش نظری اثر دارد. برای مثال، داده های اقتصادی دارای نرمالیتیه ضعیفی در میان مجموعه داده ها می باشند. این روش به ایجاد شاخص های اقتصادی کمک می کند که توسعه آن ها بستگی به یافته های تست همبستگی دارد. در این مثال، تصمیمات مربوط به شاخص های مورد استفاده در شاخص را می توان در نظر گرفت. چون تست های همبستگی می توانند بهبود یابند، احتمال تکرار و نیز ارزش تحقیقاتی که از این روش استفاده می کنند افزایش می یابد.

سوما، روش تبدیل بر موضوعات مختلف در رشته IS اثر می گذارد. روش قابل کاربرد به هر موضوعی است که از داده های پیوسته استفاده کرده و از غیر نرمال بودن رنج می برد. داده های پیوسته را می توان در خروجی پدیده های سازمانی (رویداد های شمارش، تولید، تراکنش های مالی) مشاهده کرد. بسیاری از این داده ها به خصوص در نسبت کارایی (عملکرد مالی) از مقدار نرمال منحرف هستند.

اهمیت ها برای استنباط علی

مرور منابع مختلف نشان می دهد که روش دو مرحله ای قادر به بهبود هر یک از چهار مقوله روایی موثر بر استنباط علی در مطالعات است (کاک 1990). این مقوله های روایی شامل موارد زیر هستند: 1- نتیجه گیری آماری 2- درونی، 3- ساختار 4- بیرونی. به عنوان یک رویکرد تبدیل، روش دو مرحله ای دارای اهمیت زیادی برای بهبود روایی نتیجه گیری آماری استنباط علی است

روایی نتیجه گیری آماری شامل تهدید هایی برای استنباط از فرضیه و تست های اندازه اثر است. چون بزرگی مقادیر و ارزش ها در صورت تبدیل تغییر نمی کند، استنباط با پارامتر، معتبر است. حفظ بزرگی ارزش ها به این مقادیر است که وقتی توزیع جمعیت نرمال باشد، استنباط ها برای متغیر های تبدیل شده با استفاده از روش دو مرحله ای در مقایسه با استنباط های مشابه انجام شده با استفاده از متغیر های با توزیع غیر نرمال صحیح تر است. استفاده از روشهای تبدیل برای کاهش لایه های پرت در نسبت های مالی موجب بهبود روایی نتیجه آماری می شود (واتسون 1990). روش دو مرحله ای نسبت به روش های دیگر مزیت دارد. بدیهی است که غیر نرمال بودن منجر به ناهم واریانسی در تست های رگرسیون شده و ایجاد آریبی در نتایج آماری می کند. این آریبی بایستی تا حد امکان در تحلیل های آماری کاهش یابد و با نزدیک شدن توزیعات متغیر به مقدار نرمال، حداقل می شود. روش دو مرحله ای موجب بهبود پایایی شاخص روایی در مطالعه وان البادا و رابینسون (2007) شده است.

روایی نتیجه گیری آماری بستگی به کاربرد مناسب ابزار های آماری (استراب و همکاران -2004 دارد. از این روی محققان باید قادر به انتخاب روش های آماری مناسب برای بیشینه سازی قدرت آماری باشند و تست های پارامتریک قوی تر از تست های غیر پارامتریک باشند. روش دو مرحله ای، متغیر ها را به شکل توزیعی تبدیل میکند به خصوص زمانی که استنباط آماری مربوط به جمعیت ها موجب بهبود قدرت آماری می شود. به عنوان مثال SEM و PLS دو رویکرد آماری برجسته ای است که در بر گیرنده روش های حساس به نرمالیت است: هر دو روش قادر به پیش بینی یک متغیر وابسته به کشیدگی و چولگی بالا است (کارشی و کامپو 2009).

منابع نشان می دهد که روش دومرحله ای قادر به بهبود بعد روایی داخلی استنباط های عالی است. روایی داخلی مربوط به این است که آیا یافته های استنباطی قابل تعمیم به علیت هستند یا خیر. برای مثال، داده های پرت اثر منفی بر روی آریبی رگرسیون آماری دارد. با حفظ محدود سازی مقادیر، استفاده از روش دومرحله ای منجر به کاهش تهدید برای روایی درونی می شود.

تبدیلات به سمت نرمال بودن با استفاده از روش دو مرحله ای بر روایی ساختار اثر دارند که نشان دهنده میزان عملیاتی شدن شاخص ها از دیدگاه تئوری است. در تحقیقات پیمایشی مدرن، یک مسئله رایج زمانی حادث می شود که روابط مداخله ای بین تعداد کمی از سطوح ساختاری وجود داشته باشد. از این نظر نانالی و برنشتین (1994) استدلال می کند که کاهش آریبی بستگی به افزایش تعداد سطوح اندازه گیری دارد. این مطابق با توصیه طراحی مقیاس های پیمایشی با بیش از 100 سطح است که به آسانی قابل اصلاح برای تبدیلات به سوی نرمال بودن آماری با روش دو مرحله ای است.

در نهایت منابع نشان می دهند که روش دو مرحله ای موجب بهبود روایی خارجی استنباط های علی می شود. روایی خارجی میزان تعمیم رابطه علی به نمونه های دیگر است. داده های پرت، نتیجه اثر متقابل اثرات تاریخی با تیمار ها است. برای مثال، شاخص های حسابداری ناشی از یک رویداد آب و هوایی فاجعه آمیز است. با حرکت این مقادیر از مقدار فرین به سمت مرکز توزیع، روش دو مرحله ای موجب می شود تا محتوی متغیر معتبر تر شود و این منجر به کاهش اثر متقابل تاریخچه- تیمار می شود.

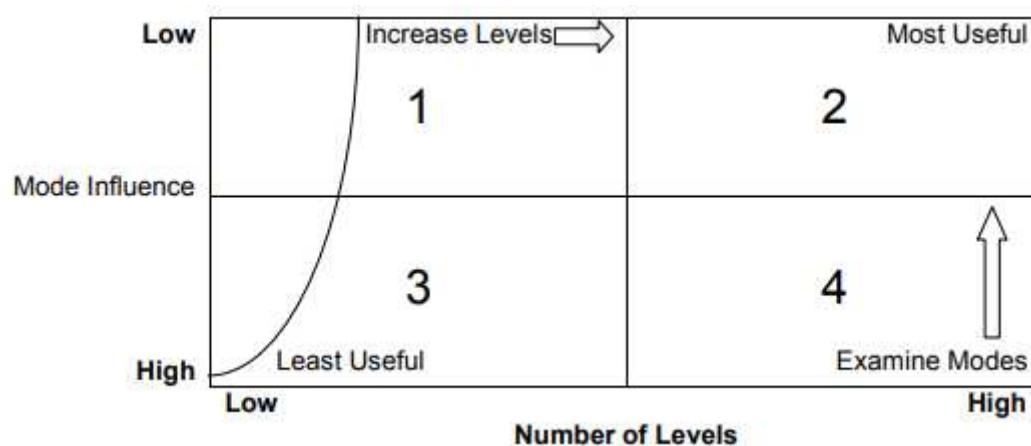
اهمیت برای تحقیقات آینده در خصوص رویکرد دو مرحله ای

به دلیل تغییر فناوری (برای مثال، افزایش استقرار سیستم های داده ها)، داده های پیوسته از اهمیت زیادی در تحقیقات IS برخوردار است. جامعه تحقیقات IS چند رشته ای در زمینه پردازش داده و اطلاعات، بایستی یک نقش برجسته را در حل مسائل نرمال بودن اختیار کند. تحقیقات آینده برای پیشرفت دانش در خصوص نقاط مثبت و منفی تبدیلات به نرمال لازم است. به طور کلی، اثر تبدیل داده های مشاهده شده بر روی تست های همبستگی چیست؟ اگرچه این سوال با شبیه سازی های مونت کارلو پاسخ داده نمی شود،

روش دو مرحله ای امکان این تحقیق را می دهد. به طور کلی، تبدیل دو مرحله ای امکان تحقیقاتی را در خصوص کارایی نرمال بودن در زمان استفاده در پدیده های جهان واقعی می دهد.

کاربرد رویکرد

شکل 6 چارچوب مورد استفاده محققان را برای طراحی شاخص هایی برای روش دومرحله ای نشان می دهد. متغیر ها در این نقشه مفهومی از دو محدودیت رویکرد ساخته شده اند (به جز قوس های ترسیم شده در باکس 1 و 3 که نشان می دهد متغیر با تعداد کم سطوح دارای اثر کمی بوده و در سمت چپ قوس قرار نمی گیرد)



شکل 6: چارچوب کاربردی برای روش دومرحله ای

باکس 1

در یک مورد مربوط به دو سطح (متغیر های دو دویی نظیر جنسیت یا مشارکت در برنامه تیماری دو سویه) روش دو مرحله ای هیچ سودمندی ای در تبدیل به سمت نرمالیته ندارد. متغیر های با سطوح پایین (مقیاس لیکرت 5 تا 7 نقطه ای) دارای اثر مد بالایی می باشند زیرا همه سطوح به عنوان مد در این موارد عمل می کنند. برای خروج از باکس 1، محققان باید قادر به تعیین ابزار ها (برای مثال مقیاس های پرسش نامه ادراکی یا حسگر های از راه دور) با چندین سطوح هستند.

باکس 2

متغیرها در سمت راست بالای ماتریس (تعداد زیاد سطوح و مقدار پایین اثر مد) اصلاح پذیر هستند. تصویر این شرایط ایده ال در جدول 2 نشان داده شده است. متغیرهای با اثر پایین مد و تعداد زیاد سطوح به نرمالیتیه آماری می رسند به خصوص زمانی که با استفاده از روش دو مرحله ای تبدیل شوند. محققان باید قادر به طراحی ابزارهایی با سطوح بسیار زیاد برای تبدیل متغیرها به سطوح نرمالیتیه بی سابقه و اثرات پایین دست باشند.

باکس 3

متغیرها در سمت چپ پایین ماتریس (تعداد کم سطوح و اثر مد بالا) به طور موفق به سمت نرمالیتیه با استفاده از روش دو مرحله ای تبدیل نمی شود. در این حالت، متغیرهای دودویی دارای مد های موثر بوده و با استفاده از رویکرد بهبود نمی یابند. متغیرهای شمارشی دارای تعداد کمی از سطوح و مقدار موثر صفر است. هر دو بعد موجب کاهش بهبود نرمال بودن می شود در حالی که محققان اغلب پیشرفت به سمت نرمال بودن را مشاهده می کنند

باکس 4

در صورت تعداد بالای سطوح و اثر مد بالا، محققان تغییرات توزیعی را مشاهده می کنند که در جدول 3 نشان داده شده است. برای حل این مسئله، محققان باید مقادیر مد را حذف کنند. وقتی که اثر مد کاهش یافت این متغیرها به صورت متغیر باکس 2 طبقه بندی می شود. یک مثال از این طبقه بندی مجدد، استناد کل بوده و اثرات این تغییرات با مقایسه جدول 3 و 4 کاملاً مشهود است.

توصیه ها

جدول 5 توصیه های مهمی را برای محققان با استفاده از روش دو مرحله ای نشان می دهد. توصیه ها به ابعادی طبقه بندی می شود که برای افراد استفاده کننده از مدل مناسب است

جدول 5: توصیه هایی برای محققانی که از روش دو مرحله ای استفاده میکنند

نگرانی محقق	توصیه
چه زمانی من از تبدیل استفاده می کنم؟	هر متغیر پیوسته با استفاده از روش دو مرحله ای تبدیل می شود. متغیرها با تعداد زیاد سطوح و مود غیر موثر، بیشترین بهبود را نسبت به نرمال بودن نشان

می دهد	
این روش قابل کاربرد به داده های اسمی نیست؟ متغیر های دو دویی تغییری را در توزیع و یا بهبود بر اساس تشخیص نرمال بودن نشان نمیدهد	چه زمانی نباید استفاده کرد؟
به شکل 1 و توضیح مربوط به منطق رویکرد دو مرحله ای مراجعه کنید توابع برای تبدیل متغیر ها با استفاده از رتبه های درصدی و توابع معکوس نرمال در هر بسته نرم افزار اماری قابل دسترس است. از استفاده از توابع برای سلول خالی اجتناب کنید مرحله 1: بسته به ابزار نرم افزاری مورد استفاده، بیشترین و کم ترین مقدار در متغیر به 0 یا 1 تبدیل میشود. اگر این رخ دهد، 1 را با 0.9999 و 0 را با 0.0001 جایگزین کنید مرحله 2: روش دو مرحله ای واحد های امتیاز Z را با استفاده از استدلال $\mu=0$ و $\sigma=1$ در تابع نرمال معکوس تولید می کند	چگونه باید از این رویکرد استفاده کرد؟
در شرایط ایده ال، نتایج روش دو مرحله ای به سمت نرمال بودن بر طبق تست های تشخیصی نزدیک می شود. بهبود روایی استنباط علی بستگی به درجه بهبود نرمال بودن با استفاده از روش دارد	اهمیت نتایج در چیست؟
تبدیل موجب بهبود استنباط های علی از جمله قدرت اماری، تست فرضیه، اندازه اثر و تعمیم پذیری می شود. در نتیجه این موجب کاهش اندازه نمونه مورد نیاز می شود	ارزش استفاده از رویکرد چیست؟
کارایی روش بستگی به دو ویژگی توزیعی دارد مد های موثر: در نسخه متغیر اصلی، اثر منفی نشان داده شده با مد ناشی از موارد زیر است 1- نسبت مقادیری که در نمونه نشان داده می شود 2- فاصله از میانگین. تنها مقادیر مد مربوط به تئوری تست شده بایستی در تحلیل حفظ شود تعداد کم سطوح: متغیر های با تعداد کم سطوح قابل اصلاح هستند. در نتیجه، مقیاس لیکرت 5 و 7 نقطه ای استفاده نشده و با مقیاس های با تعداد سطوح زیاد جایگزین می شود	چالش ها کدام ها هستند

نتیجه گیری

تاکید زیادی بر اهمیت نرمال بودن تک و چند متغیره در کلاس های آماری، مطالعات و همایش ها شده است. با این حال مطالعات در زمینه IS و سایر علوم اجتماعی به ندرت آماره های توصیفی را در خصوص

ویژگی های توزیعی داده ها گزارش کرده اند. مقدار نرمال بودن یک متغیر مورد مطالعه اثر زیادی روی تست های آماری ای دارد که بر نتایج و دانش مربوط به موضوع اثر دارد (برگر 2000).

این مقاله دیدگاه جدید را در خصوص نرمال بودن با ارایه ابزاری برای نزدیک شدن به نرمالیت آماری در متغیر های اصلاح پذیر با روش دو مرحله ای ارایه می کند. رویکرد هر متغیر را با تعداد زیادی از سطوح و اثر مد نسبت به نرمالیت آماری تبدیل می کند. بر اساس مرور منابع، محققان IS باید به نتایج امری بیشتری در تستهای همبستگی برسند. یعنی محققان باید یافته های مهم تر، اندازه اثر بیشتر و تهدید کم تر برای استنباط های علی را تجربه کنند و از این روی نتایج قابل اطمینان ناشی از استفاده از روش دو مرحله ای را بدست آورند. محققان باید از دو محدودیت رویکرد آگاه باشند (تعداد کم سطوح و اثر بالای مد). شیوه ها باید بر این اساس طراحی شده و توصیه می شود که ابزار های نظر سنجی دقیق برای بهره گیری از روش دو مرحله ای استفاده شوند. تحقیقات تجربی در خصوص کارایی روش دو مرحله ای اهمیت آن ها برای اثرات پایین دست بر روی مطالعات در آینده باید انجام شوند.



این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی