نمایش جنبه های مهم تغییرات مجموعهٔ بزرگی از منحنیهای مشابه

ام.سی.جونز جی.ا.رایس
مترجم: آیت الله موسوی[۱]

چکیده

نمایش تعداد زیادی از منحنیها بر روی یک نمودار (به طوری که هر منحنی بر روی سایر منحنیها قرار گیرد یا در قسمتهای زیادی با یکدیگر تلاقی داشته باشند) از نظر ظاهری ناخوشایند و از جنبهٔ اطلاع دهی ضعیف است. در این مقاله با استفاده از روش تحلیل مؤلفه های اصلی، تعداد محدودی از منحنیها طوری انتخاب می شوند که نحوهٔ تغییرات موجود بین همهٔ منحنیها را تا حد زیادی بیان می کنند.

۱. طرح مسأله

شکل ۱ ظاهر زیبایی ندارد. در این شکل مجموعه ای از ۱۰۰ منحنی مشابه با یک روش ثابت و با استفاده از ۱۰۰ مجموعه دادهٔ شبیه سازی شده، روی یک نمودار رسم شده اند. روشن است که برای نمایش جنبه های مهم تغییرات این مجموعه منحنی، شکل ۱ از چند جهت با شکست مواجه می شود. رفتار هر یک ازمنحنیها در اکثر قسمتهای نمودار، در انبوهٔ حاصل از سیاهی ایجاد شده توسط سایر منحنیها گم می شود. هرچه تعداد این منحنیها افزایش یابد، نمودار حاصل به جای این که اطلاع دهی بیشتری داشته باشد، گیج کننده تر می شود. در بخش ۲ روش ساده أی برای نمایش قسمت عمدهٔ اطلاعات پنهان شده در شکل ۱ ارائه می شود. برای انجام این کار، از روش تحلیل مؤلفه های اصلی استاندارد استفاده می کنیم. نتیجهٔ این کار، مجموعه أی از دو یا سه شکل است که هر یک تعداد محدودی، مثلاً ۳، منحنی را در بردارند. لازم به تذکر است که در این روش فقط زیرمجموعه ای از منحنیهای اولیه نمایش داده می شوند، نه اینکه اطلاعات موجود در منحنیهای اولیه ترکیب شوند تا توابع جدیدی به دست آیند. در عوض تلاش می شود تا منحنیهایی را انتخاب کنیم که جنبه های مهم تغییرات مجموعهٔ منحنیها را به بهترین وجه نمایش دهند. برای تشریح این روش، داده های شکل ۱ را به کار می بریم. در حقیقت هر یک از این ۱۰۰ منحنی، برآورد یک تابع چگالی به صورت زیر است: [۱]

$$\hat{f}(x) = ۰/۰۲\sum_{i=۱}^{۵۰} h^{-۱}\phi\{h^{-۱}(x - X_i)\}$$

که در آن $X_۱, X_۲, \ldots, X_{۵۰}$ نمونه ای از توزیع نرمال استاندارد با تابع چگالی ϕ است. پارامتر h را پهنای پنجره یا پهنای نوار[۲] می نامند که یک پارامتر هموار کننده است و میزان هموار کردن داده ها را کنترل می کند.اخیراً اکثر کارهای انجام شده در این زمینه بر نحوهٔ انتخاب h متمرکز بوده است. این مثال از جمله مواردی است که به کار بردن روش این مقاله برای آن می تواند مفید باشد.

کاربرد دیگر این روش در داده های حاصل از

---

[۱] دانشگاه پیام نور، بروجرد

[۲] Band Width

[Downloaded from andisheyeamari.irstat.ir on 2025-07-14]

انـدازه گیری های مکرر[۳] یک صفت (بر روی انسان، حیوان و ...) در زمانهای مختلف یا تحت شرایط متفـاوت است. در این مورد نیز مثال داده های سطح ازون شرح داده خواهد شد. منحنیهای رشد نـیز از این قبیل اند.

۲. تحلیل مؤلفه های اصلی

از بین ۱۰۰ منحنی شکل ۱ باید تعداد محـدودی از منحنیـها را طوری انتخاب کنیم که جنبه های مهم تغییرات موجود در مجموعـهٔ منحنیها را به بهترین وجه منعکس کننـد. یکـی از روشـهای مؤثـر در انجام این کار، استفاده از تحلیـل مؤلفـه هـای اصلـی اسـت. در ایـن حالت می توانیم هر یک از N تـابع را بـه صـورت بـرداری شـامل k مقدار در نظر بگیریم. در منحنیهای رشد و طرح اندازه هـای تکـرار شده نیز هر یک از N موجود (انسان، حیـــوان، قطعه و ...) تحت k وضعیت مختلف یا در k زمان متفاوت، مـــورد انـدازه گیـری قـرار می گیرند. لذا یک ماتریس N×k از داده هـا داریـم کـه مـی توانیـم تحلیل مؤلفه های اصلی استاندارد را نسبت بـه مـاتریس کوواریـانس مربوط به آنها به کار ببریم [۲]. فرض کنید

$$X'_{N\times k}=\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{Nk} \end{bmatrix}$$

اگر $S_{k\times k}$ مـــاتریس کـــوواریانس X' باشد و $e_1,\ldots,e_k$ بردارهای ویژهٔ ماتریس S باشند و

$$X=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}$$

آنگاه مؤلفه های اصلی به ترتیب عبارت اند از:

$Y_1=e'_1X$ مؤلفهٔ اصلی اول

$Y_2=e'_2X$ مؤلفهٔ اصلی دوم

$\vdots$

$Y_k=e'_kX$ مؤلفهٔ اصلی k ام

هر یک از این مـــــؤلفه ها دارای N مشاهده است کـه آنـها را مـــرتب کرده و با $P_{(N)},\ldots,P_{(2)},P_{(1)}$ نشان می دهیـــــم. $P_{(1)}$ مینیمم، $P_{(\frac{N+1}{2})}$ میانـه، $P_{(N)}$ ماکسیمم ایـن مشـاهدات انـد. حـال منحنیهای متناظر با این سه مشاهده را از بین منحنیها انتخاب و رسـم می کنیم. معمـولاً مؤلفـه هـای اصلـی اول و دوم، بیشـترین تغییـرات موجـود در داده هـا را توضیـح مـی دهنـد. شـکلهای ۲ و ۳، نتیجـهٔ کاربرد این روش، دربارهٔ منحنیهای شکل ۱ هستند. شـکل ۲ مربـوط به مؤلفهٔ اصـــلی اول است که ۴۰/۶ درصد از کـل تغییـرات را بیـان می کند. منحنـــی خـــط چیـــن کشیده تر مربوط به $P_{(1)}$، یـک چگالی با پراکندگی کم اسـت ، منحنی خـط چیـن عریضـتر، یـک چگالی با پراکندگی زیاد و منحنی پر رنگتر مربوط به میانه است کـه پراکندگی متوسطی دارد.

شکل ۳ مربوط به مؤلفـهٔ اصلـی دوم اسـت و نـوع متفـاوتی از تغییر منحنیها را نمایش می دهـد. مؤلفـهٔ اصلـی دوم ۳۶/۴ درصـد از کل تغییرات را بیان می کند. این میزان تغییر بـا تغییراتی کـه توسـط مؤلفهٔ اول بیان مـی شـود، همبستگی نـدارد. در شـکل ۳ منحنیـهای مربوط بـه مینیمـم و ماکسـیمم، مشاهدات مؤلفـهٔ اصلـی دوم (یعنـی منحنیهای خـط چیـن) در جـهات مخـالف یکدیـگر، چولـه هسـتند. منحنی مربوط به میانه، چولگی کمتری دارد (هرچنـد کـاملاً متقـارن نیست). این دو مؤلفه مجموعاً ۷۷ درصـــد از کـل تغییـرات را بیـان می کنند، لـذا مؤلفـهٔ اصلـی سـوم و بـالاتر از آن از اهمیـت کمـتری برخوردارند و همچنین تغییراتی کـه توسـط آنـها بیـان مـی شـود بـه آسانی قابل تفسیر نیست.

در شکل ۲ قطعاً و در شکل ۳ تـا حـــدی قـابل توجـه، تفسـیر جنبه های تغییر نمایش داده شده، از خـود شـکلها واضـح اسـت. امـا همیشه این طور نیست. در چنین حـــالتهایی نمـــودارهای کـــمکی

---

[۳] Repeated Measures

[Downloaded from andisheyeamari.irstat.ir on 2025-07-14]

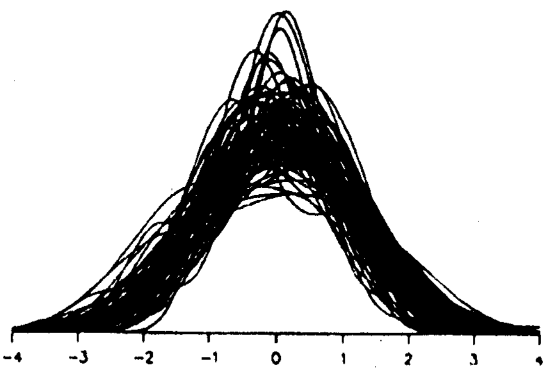
مؤلفه های اصلی، ممکن است مفید باشند. از این رو مؤلفه های اصلی متناظر با شکلهای ۲ و ۳، در شکل ٤ نمایش داده شده اند. برای تأیید ادعای ما در رابطه با پراکندگی و چولگی منحنیها، به تجمع اکثر منحنیهای مؤلفهٔ اصلی اول در مرکز، نسبت به دو انتهای منحنیها، در شکل ٤ توجه کنید. قسمت عمیق مرکزی را با قله های کوچک دو طرف مقایسه کنید. به طور مشابه دومین مؤلفهٔ اصلی تغییرات قابل ملاحظهٔ برابر، ولی در خلاف جهت یکدیگر در شانه های توزیع را، نمایش می دهد. لذا در اینجا منحنیها از منحنی متوسط (تقریباً متقارن) تا منحنیهایی که به شانه های راست و چپ چوله اند، تغییر می کنند. نمودارهای کمکی دیگری نیز مانند نمودار جعبه ای مشاهدات مؤلفه های اصلی، قابل استفاده اند.

۳. مثال دوم: داده های لایه های ازون

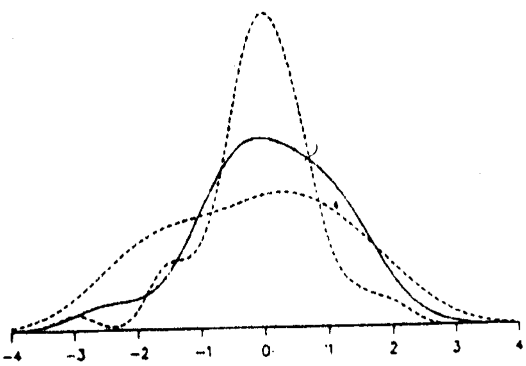
داده های مثال دوم مربوط به حداکثر لایهٔ ازون است که برای $N = ٤٧$، هفته از سال و در هر هفته $k = ٧$ مقدار روزانه در منطقهٔ معینی ثبت شده است [۳]. داده های خام در شکل ۵ نمایش داده شده اند و همانگونه که ملاحظه می شود تفسیر این شکل خیلی مشکل است. چهار مؤلفهٔ اصلی اول این منحنیها به ترتیب ۶۱/۹، ۱۸/۳، ٧/٤ و ۵/٧ درصد از کل تغییرات را بیان می کنند. برای هر یک از مؤلفه های اصلی اول و دوم، منحنیهای مربوط به $P_{(۱)}$، $P_{(۲٤)}$ و $P_{(٤٧)}$ در شکلهای ۶ و ٧ رسم شده اند. خود مؤلفه های اصلی اول و دوم نیز در شکل ۸ نمایش داده شده اند. در اینجا رسم مؤلفهٔ اصلی اول، در مقابل یک متغیر کمکی که در این مثال، شمارهٔ هفته در نظر گرفته شده، یک روند تغییرات را نشان می دهد (شکل ۹). در واقع قسمت عمیق نمودار مربوط به افزایش لایه های ازون در ماههای تابستان است. مؤلفهٔ اصلی دوم یک تغییر خطی در ازون را در طول هفته نشان می دهد. در برخی هفته ها لایهٔ ازون، افزایش، در برخی دیگر کاهش داشته و در اغلب هفته ها لایهٔ ازون تقریباً ثابت بوده است.

توانایی این روش ساده در معنی دارتر کردن نمودارهای نامفهومی مانند شکلهای ۱ و ۵، شگفت انگیز است، و نشان می دهد که استفاده از روش تحلیل مؤلفه های اصلی که در اکثر نرم افزارهای رایانه ای در دسترس است، فواید عملی بسیاری دارد.

اینکه چه منحنیهایی باید در هر مسأله انتخاب شوند تا بیشترین تغییرات را بیان کنند، هنوز نیاز به کار بیشتری دارد. استفاده از مینیمم، ماکسیمم و میانه تنها یک پیشنهاد است که در این مقاله مورد استفاده قرار گرفت.

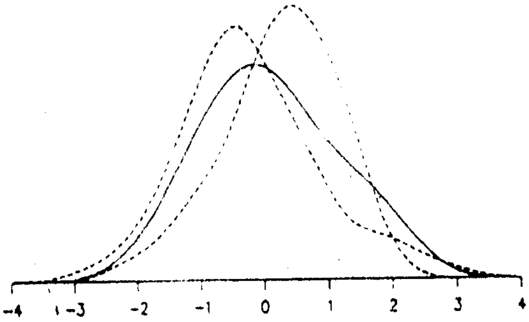


شکل ۱: مجموعه ای از برآوردهای تابع چگالی بر اساس ۱۰۰ نمونهٔ مستقل به اندازهٔ ۵۰ از توزیع نرمال استاندارد [٤]

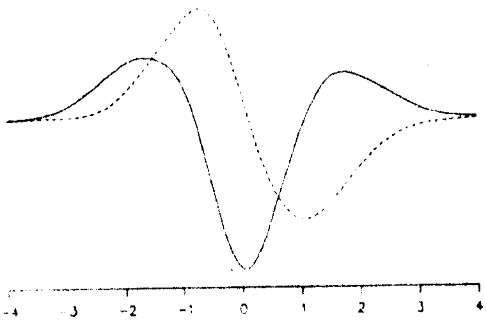


شکل ۲: منحنیهای حاصل از مؤلفهٔ اصلی اول برای شکل ۱. منحنی پررنگتر میانه و منحنیهای خط چین، مینیمم و ماکسیمم هستند.

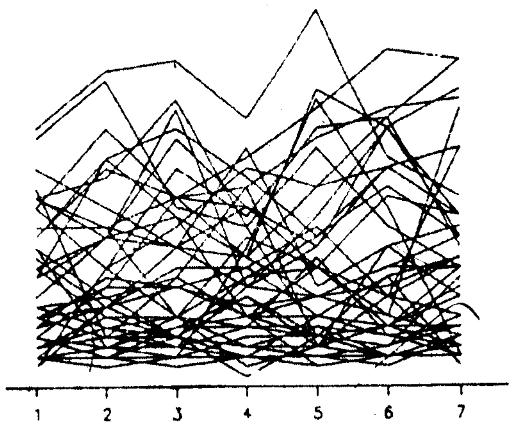
[Downloaded from andisheyeamari.irstat.ir on 2025-07-14]



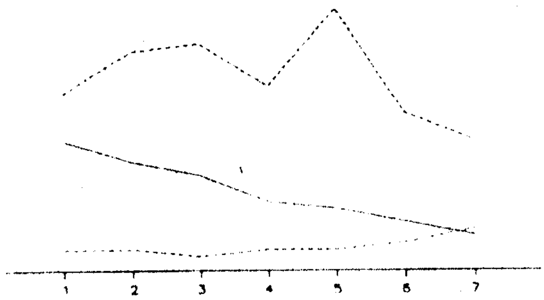
شکل ۳: منحنیهای حاصل از مؤلفهٔ اصــلی دوم برای داده های شکل ۱



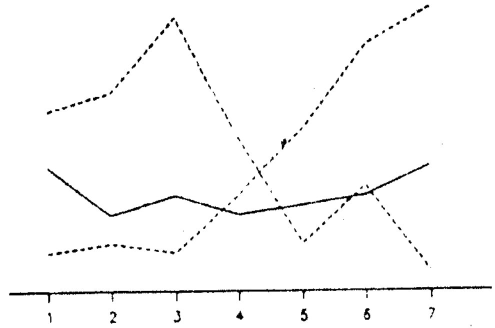
شکل ٤: مؤلفه های اصلی اول و دوم ـ مؤلفهٔ اصلی اول (منحنی پررنگتر) و مؤلفهٔ اصلی دوم (منحنی خط چین)



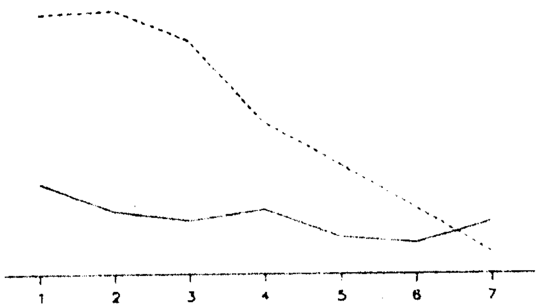
شکل ۵: داده های ازون



شکل ٦: منحنیهای حاصل از مؤلفه های اصلی اول برای داده های ازون ـ منحنی پررنگتر، میانه و منحنیهای خط چین مینیمم و ماکسیمم اند.



شکل ۷: منحنیهای حاصل از مؤلفهٔ اصلی دوم برای داده های ازون



شکل ۸: مؤلفه های اصلی اول و دوم برای داده هـای ازون ـ مؤلفـهٔ اصلی اول (منحنی پررنگتر) و مؤلفهٔ اصلی دوم (منحنی خط چین)

[Downloaded from andisheyeamari.irstat.ir on 2025-07-14]