**تحلیل گرافیکی داده های چند متغیره به کمک نمودار ستاره ای**

مهدی جباری نوقابی[[1]](#footnote-2)،وحیده محتشمی برزادران[[2]](#footnote-3)

چکیده:

در این مقاله ابتدا به معرفی نمودار ستاره ای پرداخته شده و سپس طریقه رسم این نمودار را به کمک نرم افزار R شرح می دهیم. در چارچوب یک مثال و با کمک گرفتن از این نرم افزار یک سری داده چندمتغیره مورد بررسی و تحلیل قرار می گیرد.

**1. مقدمه:**

امروزه بر هر کسی پوشیده نیست که روش های آمار توصیفی مقدار زیادی از اطلاعات را به شیوه ای قابل فهم خلاصه می کند. یکی از روش های آمار توصیفی، روش نموداری است. در این مقاله به بررسی نمودار ستاره ای که یکی از روش های نموداری برای تحلیل داده های چندمتغیره است، پرداخته می شود.

نمودار ستاره ای اولین بار توسط چامبرز[[3]](#footnote-4) در سال 1983 معرفی شد. برای آشنایی بیشتر به، چامبرز و همکاران[[4]](#footnote-5) (1983)، جانسون و ویچرن[[5]](#footnote-6)(2007) ارجاع داده می شود.

در این مقاله ابتدا به معرفی نمودار ستاره ای پرداخته و سپس طریقه رسم این نمودار را به کمک نرم افزار R شرح می دهیم. در نهایت یک مثال کاربردی بیان و به کمک نرم افزار نمودار ستاره ای رسم می گردد.

**2. *نمودار ستاره ای:***

نمودار ستاره ای تکنیکی برای تحلیل گرافیکی داده ها به منظور بررسی رفتار همبستگی همه متغیرها در مجموعه ای از داده های چند متغیره است.

نمودار ستاره ای به شکل چند طرحی ساخته شده است و هر صفحه متشکل از تعداد زیادی ستاره می باشد. هر ستاره یک مشاهده چندمتغیری را نمایش می دهد. هر ستاره شامل رشته ای از پره های متساوی الزاویه است (رادی[[6]](#footnote-7)) هر پره نماینده متغیری در مجموعه داده های چندمتغیره است. به عبارتی، هر متغیر به کمک پاره خطی که از نقطه مرکزی انتشار یافته، نمایش داده می شود.

 تعداد متغیرها، نشان دهنده تعداد پاره خط هایی که از مرکز نمودار رسم شده، می باشد و طول هر پاره خط با مقدار مربوط به هر متغیر متناسب است. ستاره ها را می توان مطابق با شباهت های شان دسته بندی کرد. نمودار ستاره ای مناسب ترین روش توصیفی برای نمایش داده های مثبت است ولی برای نمایش داده های منفی با استفاده از نمودار باید داده ها را به گونه ای منطقی تبدیل کنیم. استاندارد کردن مشاهدات در رسم نمودار ستاره ای بسیار مفید است.

**3. چگونگی رسم نمودار:**

دستور کلی رسم نمودار ستاره ای با استفاده از نرم افزارR، به شرح زیر است.

**stars**(x, full = TRUE,scale=TRUE,radius = TRUE, labels = dimnames(x)[[1]], locations = NULL, nrow = NULL, ncol = NULL, len = 1,key.loc = NULL, key.labels = dimnames(x)[[2]],xlim = NULL, ylim = NULL, flip.labels = NULL, draw.segments = FALSE,col.segments = 1:n.seg, col.stars = NA, col.lines = NA, axes = FALSE, frame.plot = axes,main = NULL, sub = NULL, xlab = "", ylab = "",cex = 0.8, lwd = 0.25, ltpar("lty"), add = FALSE, plot = TRUE, ...)

عبارتند از: stars برخی شناسه های فرمان

**:x** ماتریس یا frame data داده های چندمتغیره است که در این جا برای هر سطر این ماتریس یک نمودار ستاره ای رسم خواهد شد.

 **تذکر**:در رسم این نمودار داده گمشده مجاز است و نرم افزار مقدار کمی آن را صفر منظور خواهد کرد.

 **:full** اگر این زیر دستور TRUE(یاT ) باشد نمودار روی یک دایره کامل و اگرFALSE (یاF ) باشد، بر روی نیم دایره بالایی رسم خواهد شد.

**:scale** اگر مساوی TRUE(یاT ) باشد، به این معنی است که داده ها در هر ستون مستقل می باشند و می توان برای هر متغیر یک نمودار ستاره ای مستقل رسم کرد ولی اگر FALSE(یاF ) باشد، به این معنی است که داده ها مستقل اندازه گیری نشده اند و نمودار همه متغیر ها را بر روی یک نمودار رسم خواهد کرد.

**Radius:** اگر مساوی TRUE(یاT ) باشد، محورهای مختصات را رسم می کند و اگر FALSE (یاF ) باشد محورهای مختصات را رسم نمی کند.

**labels:** لیستی است که به کمک آن می توان نام هر متغیر را در پایین ستاره مربوط به آن نوشت.

**locations:** ماتریسی است با دو ستون که می توان به کمک این دستور محل قرار گرفتن هر ستاره را مشخص کرد. نرم افزار به طور پیش فرض همه نمودار ها را بر روی یک صفحه مستطیلی رسم می کند.

**ncol وnrow :** به کمک این دستورها می توان تعداد ستون ها و سطر هایی که می خواهیم صفحه نمایش را به کمک آن ها دسته بندی کنیم، مشخص نماییم. به طور پیش فرض تعداد سطرها و ستون ها با هم برابر است.

**len:** طول میله ها را با این دستور می توان تغییر داد.

**key.loc:** برداری است برای تعیین محل قرار گرفتن کلید راهنما جهت نمایش محل قرار گرفتن شاخص ها.

**key.labels:** برداری از متغیر های رشته ای است که به کمک آن مشخصه ها را برچسب گذاری می کنیم.

**xlim:** برداری است که می توان به کمک آن دامنه تغییرات محور x ها را متناسب با نمودار تنظیم نمود.

**ylim:** برداری است که می توان به کمک آن دامنه تغییرات محورy ها را متناسب با نمودار تنظیم نمود.

**flip.labels:** اگر مقدار آن NULLباشد، نام همه متغیرها را در یک ردیف می نویسد، ولی با تغییر آن می توان نام متغیر ها را به صورت ستونی( بالا و پایین) نوشت.

**draw.segments:** اگر مقدار این دستور TRUE (یاT ) باشد، هر شاخص را با یک رنگ نشان خواهد داد و اگر FALSE(یا F) باشد شاخص ها رنگی خواهند بود.

**col.segments:** شاملبرداری است از رنگ ها که می توان برای هر شاخص، یک رنگ خاص در نظر گرفت.

**col.stars:** برداری از رنگ ها است که می توان به کمک آن به هر ستاره یک رنگ خاص اختصاص داد.

**تذکر:**اگر draw.segments=TRUE باشد، نرم افزار مقدار شناسه دستور col.stars را نادیده می گیرد.

**col.lines:** برداری از رنگ هامی باشد که می توان به کمک آن رنگ خطوط دور ستاره ها را تغییر داد.

**تذکر**:اگر draw.segments=TRUE باشد، مقدار دستور col.lines در نرم افزار مورد نظر قرار نمی گیرد.

**axes:** اگر مقدار آن برابر TRUE(یا T) باشد، محور های مختصات را بر روی صفحه نمایش رسم خواهد کرد.

**frame.plot:** مقدار این دستور اگر برابر TRUE (یاT ) باشد، یک قاب اطراف نمودار رسم خواهد کرد.

**main:** به کمک این می توان عناوین اصلی بالای نمودار ستاره ای را تعیین نمود.

**sub:** به کمک این دستور زیر عنوانی که در زیر نمودار نوشته می شود، تعیین می گردد.

**xlab:** این دستور برای تعیینبرچسبمحور طول ها روی نمودار می باشد.

**ylab:** این دستور برای نام گذاری و تعیین برچسب برای محور عرض ها می باشد.

**cex:** در صورتی که بخواهیم نام متغیرها در داخل نمودار ستاره ای نوشته شود، می توان از این زیر دستور استفاده کرد.

**lwd:** مقدار این زیر دستور عددی است که تعیین می کند، ضخامت خطوط اطراف ستاره ها چقدر باشد.

**lty:** مقدار این زیر دستور نیز عددی است که به کمک آن نوع خط مورد استفاده در رسم نمودار مشخص می شود.

**add:** اگر مقدار آن TRUE (یاT ) باشد بعدا می توانیم یک نمودار دیگر روی آن بیندازیم.

 **plot:** اگرمقدار این دستور FALSE(یاF ) باشد در صفحه ی نمایش هیچ نموداری رسم نخواهد شد.

4. **مثال :** در یک تحقیق،20 نوع آمینو اسید را براساس 19 شاخص، مورد بررسی قرار داده ایم. که داده های جمع آوری شده به شرح زیر می باشند. واضح است که این 20 نوع آمینواسید به همدیگر دو به دو وابسته اند. در این مثال می خواهیم به کمک نمودار ستاره ای آن ها را مورد بررسی قرار دهیم. (این مجموعه داده شامل 19 متغیر و 20 مشاهده است)

stars(x,labels=dimnames(x)[[1]],key.labels=dimnames(x)[[2]],draw.segments=TRUE,col.segments=c("red","yellow","orange","blue","green","pink","purple","black","brown","white","brown","burly wood","gold","deeppink","lightcyan","lavender","hotpink4","lightgoldenrod"))

شاخص های مورد بررسی به همراه نام فراخوانی آن ها در نرم افزار R که داخل پرانتز نوشته شده است به شرح زیر می باشد.

شاخص 1: Arthrobacter.sp..ON14(x1)

شاخص 2:Psychrolactophilus.strain.F2 (x2)

شاخص 3: (x3)Arthrobacter.sp..20B

شاخص 4: Arthrobacter.sp..sB(x4)

شاخص5: Streptococcus.thermophilus(x5)

شاخص6 :Pseudoalteromonas.sp..22b(x6)

شاخص7: x7)Uncultured.bacterium.zd410)

شاخص8: Planococcus.sp..L4(x8)

شاخص9:Strain..SOS.Orange .(x9)

شاخص10: Lactobacillus.acidophilus(x10)

شاخص11: Alcalilactibacillus.ikkense(x11)

شاخص12: T..maritime.Bgalc(x12)

شاخص13: Deinococcus.geothermalis(x13)

شاخص14:T..maritima.BgaA (x14)

شاخص15:T..maritime.BgalB (x15)

شاخص16: Thermus.thermophilus(x16)

شاخص17:Thermus.sp..IB.21.bgaA (x17)

شاخص18:Bacillus.subtilis.strain.G1(x18)

شاخص19:Paecilomyces.aerugineus (x19)

 نمودار ستاره ای به کمک نرم افزار R در شکل 1 رسم شده است.





 *شکل1*

*ب*ه نظر می رسد که آمینواسید شماره 10 بیش تر از سایر آمینواسیدها، تحت تاثیر شاخص ها قرار دارد، هر چند که آمینواسید دهم نسبت به شاخص های x7,x4,x3 کمتر تحت تاثیر قرار می گیرد. همچنین آمینواسید دوم تقریبا تحت تاثیر هیچ یک از شاخص ها قرار ندارد و آمینواسیدهای

7،11،19 کمتر از سایر آمینواسیدها تحت تاثیر شاخص ها قرار می گیرند. به طور کلی آمینواسیدهای 1،3،4،6،18 تقریبا تحت تاثیر شاخص ها قرار دارند و آمینواسیدهای 16،15،13،17 کمتر از آمینواسیدهای1،3،4،6،18تحت تاثیر شاخص­ها قرار می­گیرند. آمینواسیدهای 14،12،9،8،5،20 کمتر از آمینواسیدهای 13،15،16،17 تحت تاثیر شاخص­ها می­باشند...........................................اگر به اشتباه متغیرها را به صورت یک متغیره بررسی می­کردیم، این نتیجه حاصل نمی­شد. توجه داریم که هر آمینواسید به 19 شاخص بستگی دارد.

منابع:

1. <http://stat.ethz.ch/R-manual/R-patched/library/graphics/html/stars.html>

**2**.-جانسون ،ریچارد.آ؛دین دبلیو.ویچرن، ((تحلیل آماری چند متغیری کاربردی))، ترجمه:حسینعلی نیرومند، دانشگاه فردوسی مشهد، مشهد، چاپ سوم بهار 1386.

3.Chambers, J. M. Cleveland, W. S.et al. (1983) Graghical methods For Data Analysis, Wadsworth and Brooks /cole

4. Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). The New S Language. Wadsworth & Brooks/Cole.

1. استادیار دانشگاه فردوسی مشهد [↑](#footnote-ref-2)
2. دانشجوی کارشناسی ارشد آمار ریاضی،دانشگاه فردوسی مشهد [↑](#footnote-ref-3)
3. chambers [↑](#footnote-ref-4)
4. Chambers and et.al [↑](#footnote-ref-5)
5. Johnson and wichern [↑](#footnote-ref-6)
6. radii [↑](#footnote-ref-7)