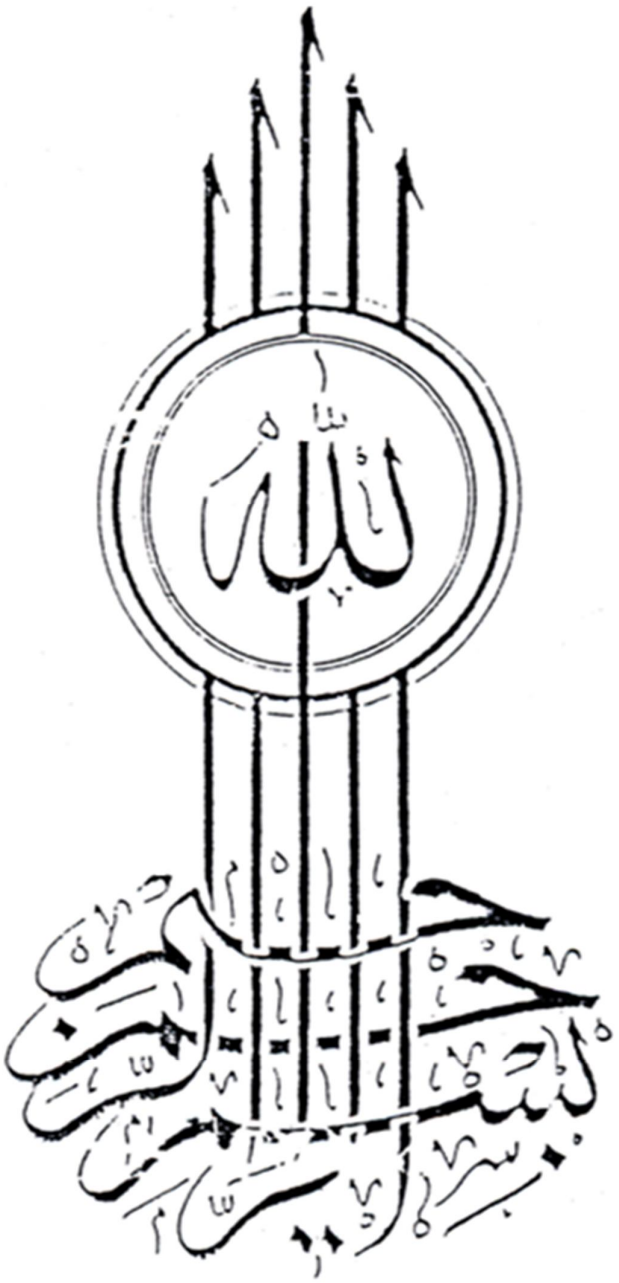
الله
بسم الله الرحمن الرحيم

دانشگاه پیام نور مرکز شیراز
دانشکده علوم گروه آمار

عنوان پایان‌نامه

ماتریس کوواریانس مجانبی میانه اوجا

برای دریافت درجه **کارشناسی ارشد**

رشته آمارریاضی

استاد راهنما:
دکتر نرگس عباسی

استاد مشاور:
دکتر علی‌رضا نعمت‌اللهی

مؤلف:
صفیه آتش‌روز

تیر 1389

چکیده

در بسیاری از تحلیل‌های آماری، برخی از داده‌ها نقش بسزایی در تحلیل‌ها ایجاد می‌کنند و شاید باعث اختلال در نتیجه‌گیری‌ها شوند از جمله‌ی این داده ها، نقاط‌پرت و داده‌های آلوده می‌باشند که در کران‌نفوذپذیری تأثیر گذارند و همچنین برآوردهای پایدار از پارامترهای مکان چندمتغیره و ترکیبی از آنها در مسائل آماری مورد تردید قرار می‌گیرند. در این پایان‌نامه، میانه‌ی پایای تعمیم‌یافته چندمتغیره خطی را که توسط اوجا پیشنهاد شده معرفی و سپس به مقایسه با میانه‌ی فاصله‌ای و میانه در حالت کروی، و بیضوی پرداخته و تأثیر این نقاط پرت و آلوده را روی آنها بررسی می‌کنیم. سپس به معرفی برآوردهای ماتریس کوواریانس میانه پرداخته و با ارائه دو روش اندازه‌گیری و احتمال پوشش تجربی (CP) به مقایسه عددی برآوردگرها می‌پردازیم. اوجا در واقع میانه در حالت چند متغیره را به میانه‌ی فاصله‌ای در حالت یک‌متغیره ساده می‌کند و نشان می‌دهد که به طور مجانبی به کارایی میانه‌ی فاصله‌ای است و دارای خصوصیات کارا وپایدار خوش‌رفتاری است. از این رو در این پایان‌نامه به کارایی مجانبی، تابع نفوذپذیری و اثرات نقاط فروریزش در حالت کلی و برای میانه اوجا، اشاره خواهیم کرد. سپس سه مفهوم اساسی، علامت و رتبه چندمتغیره، ماتریس کوواریانس رتبه و علامت و برآوردهای آن، همچنین پایداری و کارایی این برآوردها، ارائه خواهد شد.

نهایتاً با توجه به مدل‌های چندمتغیره متقارن و راه‌های گوناگون برآورد ماتریس کوواریانس علامت مربوط به این مدل به مقایسه عددی این برآوردها می‌پردازیم و همچنین وابستگی این برآوردگرها را به نقاط فروریزش در نمونه‌های متناهی و همچنین اریبی، آنها را نشان می‌دهیم سپس پیشنهادی برای از بین بردن میزان اریبی ارائه خواهیم کرد.

فهرست مطالب

عنوان **صفحه**

عنوان **صفحه**

عنوان **صفحه**

فهرست نمودارها

فهرست جداول

برنامه‌های کامپیوتری

مقدمه

در بعضی از تحلیل‌های آماری ممکن است با داده‌هایی روبه‌رو شویم که در نتیجه‌گیری‌هایمان تأثیرات بسزایی دارند و باعث اختلال در تحلیل می‌گردند و در این صورت ما را در تصمیم‌گیری دچار تردید و شک می‌کنند، از جمله داده‌های پرت و داده‌های آلوده، بنابراین بایستی به‌دنبال راه حل و پیشنهادی برای حل این موضوع بود. هدف این پایان‌نامه معرفی میانه‌ی پیشنهاد شده توسط اوجا (اوجا1999) و خصوصیات مربوط به آن، ماتریس کوواریانس علامت (هتمنسپرگر1994) می‌باشد، سپس برآوردهایی برای ماتریس کوواریانس و مقایسه عددی آنها (اوجا ، هتمنسپرگر و نادار2003)، تأثیر نقاط پرت و آلوده روی میانه اوجا و برآورد ماتریس کوواریانس میانه اوجا (اوجا، کوئینن و ویسوری 2000)، ارائه خواهد شد. در این میان به مسائلی همچون کارایی مجانبی (نینیما و اوجا1994)، نقاط فروریزش (نینیما و اوجا1994)، کران نفوذپذیری (نینیما و اوجا1994) تعریف توزیع‌های بیضوی (اوجا و کونینن و ویسوری2000)، معرفی برخی از توزیع‌های کروی، مقایسه میانه چندمتغیره در حالت بیضوی با میانه تعمیم‌یافته معرفی شده توسط اوجا (نینیما و اوجا 1984) و همچنین رتبه و علامت میانه فاصله‌ای و میانه اوجا (اوجا و کوئینن و ویسوری 2000)، مطالعه پایداری و کارایی برآوردهای ماتریس کوواریانس (اوجا وکوئینن و ویسوری 2000) خواهیم پرداخت.

در فصل اول به مفاهیم اولیه که در متن این پایان‌نامه به دفعات مورد استفاده قرار گرفته است، اشاره می‌کنیم و در فصل دوم به معرفی میانه اوجا، تابع اصلی یا تابع هدف در حالت سه‌متغیره و kمتغیره و ارائه یک مثال می‌پردازیم. سپس خصوصیات مجانبی میانه تعمیم‌یافته و کارایی مجانبی درحالتی از نرمال چندمتغیره و تابع نفوذپذیری میانه‌ی اوجا و سپس نتایج مربوط به آن می‌پردازیم.

در فصل سوم ماتریس کوواریانس علامت و رتبه‌ای (هتمنسپرگر1999) و مدل‌های بیضی‌گون (اوجا وکوئینن و ویسوری 2000) را شرح داده و با ارائه‌ی برآوردهایی برای ماتریس کوواریانس (اوجا و هتمنسپرگر 2003) به مطالعهٔ کارایی و پایداری آنها می‌پردازیم در این میان چند قضیه و اثبات مربوط به آن ارائه می‌گردد.

در فصل چهارم با تعریف داده‌های آلوده و تأثیرات آنها به بررسی اریبی این برآوردگرها برای نمونه‌های متناهی می‌پردازیم و یافته‌هایمان را از برآورد ماتریس کوواریانس در حالت دو و سه‌متغیره به مقایسه عددی گذاشته و با بیان نتایج، پیشنهادهایی ارائه خواهیم کرد.

فصل اول

مفاهیم

در این فصل برخی از مفاهیم به‌کار گرفته در این رساله آورده شده است و برای درک و تصور بهتر مطالب در برخی از بخش‌ها مثال‌های عددی به همراه نمودار ارائه گردیده است.

توزیع بیضی‌گون و کروی

توزیع‌های کروی[1] و بیضی‌گون[2] به‌ترتیب تعمیم توزیع نرمال استاندارد چندمتغیره‌ی $N_d(0, I_d)$ و توزیع نرمال چندمتغیری $N_d(\mu, \Sigma)$ می‌باشند(رجوع کنید به اندرسن[3] (1993)، لانگ[4] (1979)، آلکین[5] (1992)، یان و بنتلر[6] (2004)، زلنر[7](1976)) می‌توان بسیاری از روش‌های قدیمی تحلیل‌های آماری را مستقیماً یا با اندکی تغییر در مورد جوامع با توزیع بیضی‌گون به‌کار برد.

1-1 معرفی توزیع بیضی‌گون

یک بردار تصادفی با توزیع بیضی‌گون، مانند $\boldsymbol{X}$، به‌وسیله یک سه‌تایی $(\mu, \mathbf{D}, g)$ مشخص می‌شود و به‌صورت $X \sim EC_p(\mu, \boldsymbol{D}, g)$ نمایش داده می‌شود که در آن $E(\boldsymbol{X}) = \mu$ و $Var(\boldsymbol{X}) = \sum = a\boldsymbol{D}$ که a یک مقدار ثابت مثبت و g تابعی است که یک خانواده خاص از توزیع‌های بیضی‌گون را مشخص می‌نماید، تابع چگالی بردار تصادفی بیضی‌گون در صورت وجود به‌صورت زیر نوشته می‌شود:

$$f_{\mathbf{X}}(x) = |\mathbf{D}|^{-\frac{1}{2}} g[(x-\mu)'\boldsymbol{D}^{-1}(x-\mu)]$$

به‌عنوان مثال خانواده نرمال چندمتغیره نامنفرد (دارای ماتریس کوواریانس وارون‌پذیر) به این کلاس تعلق دارد، که در آن $a = 1$ و $g(t) = \exp\left\{-\frac{t}{2}\right\}$ در نظر گرفته می‌شود.

توزیع‌های کروی حالت خاصی از توزیع‌های بیضی‌گون هستند، بدین صورت که اگر در یک توزیع بیضی‌گون $\boldsymbol{\mu} = o$ و $\boldsymbol{D} = \boldsymbol{I}_p$ در نظر گرفته شود، توزیع حاصل توزیع کروی خواهد بود. در نتیجه اگر Y دارای توزیع کروی باشد، آنگاه با انتخاب یک تبدیل مناسب، توزیع $\boldsymbol{X} = \boldsymbol{\mu} + \mathbf{BY}$ بیضی‌گون خواهد بود که در آن $\boldsymbol{D} = BB'$ و $Var(X) \propto BB'$ می‌توان بردار تصادفی بیضی‌گون $\boldsymbol{X}_{p\times 1}$ را به‌صورت $\boldsymbol{X} = R\boldsymbol{BU} + \boldsymbol{\mu}$ نیز نمایش داد. که در آن $\boldsymbol{U}$ یک بردار

---

[1] Spherical Distribution
[2] Empirical Distribution
[3] -Anderson
[4] -lang
[5] -Olkin
[6] -Bentler and yuan
[7] -Zellner

تصادفی است که به‌طور یکنواخت روی سطح کره واحد در R^p توزیع شده است و R یک متغیر تصادفی مثبت و مستقل از U می‌باشد که با استفاده از آن می‌توان به تابع g دست پیدا کرد. به‌عنوان مثال اگر R^2 دارای توزیع کای‌دو با pدرجه آزادی باشد، آنگاه $\boldsymbol{X}$ دارای توزیع نرمال pمتغیره است.

فانگ[8] (1990)با فرض $X \sim EC_p(\boldsymbol{\mu}, \boldsymbol{D}, g)$، و $f(.)$ تابع چگالی متغیر تصادفی R، رابطه میان $g(.)$ و $f(.)$ را به‌صورت زیر بیان کرد.

$$f(r) = \frac{2\,\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} r^{n-1} g(r^2)$$

[8] Fang

1-1-1 معرفی برخی از توزیع‌های کروی

جدول زیر به اختصار برخی از توزیع‌های کروی را معرفی می‌نماید:

نوع	تابع چگالی $f(x)$ یا تابع مشخصه $\Psi(t)$
نوع کوتز	$f(x) = c(x'x)^{N-1}exp(-r(x'x)^s),\ r,s>0, 2N+p>2$
نرمال چند متغیره	$f(x) = c\ exp(-1/2\ x'x)$
پیرسن نوع V	$f(x) = c(1+\frac{x'x}{s})^{-N}$, $s>0$, $N>\frac{p}{2}$
t چند متغیره	$f(x) = c(1+\frac{x'x}{s})^{-\frac{(p+m)}{2}}$, $m>0$, $m\epsilon Z$
کوشی چند متغیره	$f(x) = c(1+\frac{x'x}{s})^{-\frac{(p+1)}{2}}$, $s>0$
پیرسن نوع 2	$f(x) = c\ (1-x'x)^m$, $m>0$
لجستیک	$f(x) = c\ \frac{\exp(-x'x)}{[1+exp(-x'x)]^2}$
بسل چند متغیره $k_\alpha(.)$ که در آن مبین تابع بسل نوع سوم است	$f(x)= c\left(\frac{\|x\|}{\beta}\right)^{\alpha} k_\alpha\left(\frac{\|x\|}{\beta}\right)$, $\alpha>-p/2$, $\beta>0$
ترکیب مقیاس که در آن $G(t)$ یک تابع توزیع پیوسته است.	$f(x) = c\int_0^\infty t^{-\frac{p}{2}} exp\left(-\frac{x'x}{2t}\right) dG(t)$
قوانین پایا	$\Psi(t) = exp\left\{r(t't)^{\frac{\alpha}{2}}\right\}$, $0<\alpha\leq 2$, $r<0$
یکنواخت چند متغیره که در آن ${}_0F_1(.,.)$ یک تابع فوق هندسی تعمیم‌یافته است	$\Psi(t) = {}_0F_1\left(\frac{p}{2}; -\frac{1}{4}\|t\|^2\right)$

جدول(1-1-1): زیر گروههایی از توزیع‌های کروی p بعدی

همانگونه که عنوان شد بردار تصادفی X، با توزیع بیضی‌گون، می‌تواند به‌صورت $X = RBU + \mu$ نمایش داده شود. که در آن B ماتریس تبدیلات، U یک بردار تصادفی است که به‌طور یکنواخت روی سطح کره واحد در R^p توزیع شده و R یک متغیر تصادفی مستقل از U می‌باشد پس برای مشخص شدن یک توزیع بیضی‌گون باید $E(X)$ و $Var(X)$ و توزیع R معلوم باشد. میانگین و واریانس را می‌توان با استفاده از میانگین و واریانس نمونه‌ای داده‌ها یا میانگین و واریانس نمونه‌ای داده‌های پیراسته[9]، که با حذف درصدی از نقاط پرت و دور افتاده به‌دست می‌آید برآورد نمود.

کوک[10] (1998و1994) بیان داشت که انتخاب توزیع R اختیاری است، بدین معنا که با انتخاب‌های مختلف برای توزیع R می‌توان توزیع‌های مختلفی به‌دست آورد، مثلاً برای دستیابی به توزیع نرمال، متغیر R^2 باید دارای توزیع کای‌دو باشد. در عمل برای دستیابی به یک انتخاب

[9] Trimmed Data
[10] - Cook

مناسب برای توزیع R می‌توان از توزیع تجربی شعاع‌های نمونه[11]، که از رابطه زیر به‌دست می‌آیند، استفاده نمود. (در این رساله به‌دلیل عدم‌نیاز به اجرای این مراحل اشاره نخواهد شد).

$$r_i = \left[(\boldsymbol{x}_i - \overline{\boldsymbol{x}})' \widehat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})\right]^{\frac{1}{2}}$$

میانگین پیراسته: تمام مشاهدات کوچکتر از چارک اول و بزرگتر از چارک سوم را حذف، و میانگین داده‌های بین Q_1 و Q_3 را محاسبه می‌کند.

2-1 داده‌های پرت

بارنت[12] (1994) داده‌ای که بیشترین فاصله از بقیه مشاهدات دارد را به عنوان داده پرت معرفی کرد. و گرابس[13] (1965) داده‌پرت را مشاهده‌ی دورافتاده‌ای که انحراف آن از بقیه داده‌ها به‌طور محسوس و برجسته قابل مشاهده است تعریف کرد.

1-2-1 روش تعیین داده‌های پرت

با توجه به بررسی‌های وسیع در زمینه‌های داده‌های‌پرت و روش‌های مورد نظر و بحث‌های مرتبط در این زمینه، چندین روش عمومی برای تعیین داده‌های‌پرت ارائه می‌شود:
استفاده از روش استانداردسازی (Z)، نمودار جعبه‌ای، نمودار پراکنش، آزمون گراب، نمودار هیستوگرام، و همچنین استفاده از چارک‌ها در تعیین داده‌های‌پرت کم و زیاد.

روش استانداردسازی:

مطابق قضیه چبی‌شف همه مشاهدات (تقریباً 99/8 درصد) در مجموعه داده‌ها دارای مقادیر (نمرات Z) کمتر از 3 هستند، که در فاصله $(\bar{X} \pm 3S)$ قرار می‌گیرند که در آن $\bar{X}$ میانگین و S انحراف معیار نمونه می‌باشد. بنابراین مشاهدات دارای مقدار قدرمطلق Z بزرگتر از 3، پرت خواهند بود.

---

[11] *Empirical Distribution of the sample Radii*
[12] - Barnet
[13] - Grubbs

روش رسم نمودار جعبه‌ای:

روش‌های آماری زیادی برای رسم نمودار جعبه‌ای وجود دارند. بعد از رسم نمودار جعبه‌ای برای هر مجموعه داده، مشاهداتی که ما بین دیواره‌های داخلی و خارجی قرار می‌گیرند، مشکوک به پرت هستند. مشاهداتی که بیرون از این دیواره‌های خارجی قرار می‌گیرند، جزو داده‌های پرت هستند. برای داده‌های بزرگ این نمودار با استفاده از برنامه‌های کامپیوتری رسم می‌گردد.

روش رسم نمودار پراکنش:

نمودار پراکنش از شیوه‌های ساده دیگر برای تعیین داده‌های پرت است، بدین صورت که بعد از رسم نمودار پراکنش بین دو متغیر وابسته Y و مسقل X، اگر داده‌ای پرت باشد روی خط برازش شده برای داده قرار نگرفته و در فاصله‌ی دورتری از خط قرار می‌گیرند. لذا خط رگرسیونی رسم شده برای داده‌ها با احتساب مقدار داده‌پرت دارای برازش خوب نبوده و بدون در نظر گرفتن داده‌پرت، برازش مدل دقیق‌تری خواهد شد.

روش آزمون گراب:

این آزمون توسط گراب در سال 1969 بیان گردید. این آزمون دارای مقدار آماره‌ای است که بر اساس اطلاعات موجود در نرم‌افزارهای مربوطه قابل محاسبه می‌باشد، لذا بعد از محاسبه مقدار آماره آزمون گراب، اگر این مقدار، از مقدار استاندارد تک تک داده‌ها بیشتر باشد آن داده از دید آزمون گراب پرت به حساب می‌آید. انجام آزمون گراب با استفاده از برنامه‌های کوییک‌کالک[14] و دیتاپلات[15]، صورت می‌گیرد که آن‌لاین قابل استفاده است.

روش نمودار هیستوگرام:

در داده‌های دارای توزیع **نرمال** اگر نمودار هیستوگرام **رسم** گردد، وجود یک داده پرت می‌تواند الگوی توزیع داده‌ها را تغییر داده و داده‌ها شکل یک توزیع چوله‌دار به خود بگیرند. همچنین وجود یک داده‌پرت باعث می‌گردد که آن داده در فاصله‌ی دورتری از بقیه داده‌ها قرار گرفته و بعد از رسم نمودار هیستوگرام به عنوان یک نقطه دور در طرف چپ و یا راست توزیع قرار بگیرد، و نوع شکل رسم شده را تحت تأثیر قرار دهد.

14- Quick calc
15- Data plot

روش استفاده از چارک‌ها:

اگر فرض کنیم که داده‌ها دارای چارک اول Q_1 و چارک سوم Q_3 با دامنه‌ی میان‌چارکی $IQR = Q_3 - Q_1$ باشند در این صورت مشاهداتی را که در نامساوی‌های $X_i < Q_1 - 1/5IQR$ یا $Q_1 + 1/5IQR < X_i$ قرار دارند جزو داده‌های پرت خفیف و مشاهداتی را که در نامساوی‌های $Q_1 + 3IQR < X_i$ یا $X_i < Q_1 - 3IQR$ قرار می‌گیرند جزو داده‌های پرت شدید هستند.

3-1 داده‌های آلوده

اغلب کسری از داده‌ها با کیفیت پایین، فاسد یا معیوب جمع‌آوری می‌شوند که به این نوع داده‌ها، داده‌های آلوده[16] گویند. گاهی این داده‌ها تأثیر چندانی در استنباط‌هایمان ندارند، اما گاهی نه‌تنها نقش بسزایی دارند بلکه تحلیل‌ها و استنباط‌هایمان را دچار اختلال می‌کنند بنابراین باید روش و ابزاری برای محدود کردن این تأثیرات در نظر گرفت. (برای درک تفاوت داده‌آلوده و داده‌پرت فرض کنید یک نمونه 20تایی از داده‌های دو بعدی نرمال داشته باشید، چنانچه 2 یا 3 داده به جای توزیع نرمال دارای توزیع دیگری باشند و در این داده‌ها قرار بگیرند حکم داده‌آلوده دارند. اما داده‌پرت داده‌ای است که خیلی دورتر و یا به عبارتی خیلی پرت‌تر از داده‌های داده شده قرار می‌گیرد).

در این رساله تأثیر داده‌های پرت و داده‌های آلوده روی میانه‌های فاصله‌ای، بیضی‌گون و اوجا را مورد بررسی قرار داده و ابزارها و یا پیشنهادهایی برای حل این نابسامانی‌ها ارائه خواهیم کرد.

4-1 نقاط فروریزش

برای نمونه‌ی مرتب شده‌ی $X_{(1)} < \cdots < X_{(n)}$ با اندازه‌ی n فرض کنید T_n آماره‌ی مربوط به این نمونه باشد. T_n به ازای $0 \le b \le 1$، دارای نقطه‌ی فروریزش[17] b، است اگر برای هر $\varepsilon > 0$، داشته باشیم:

$$\lim_{x_{((1-b)n)} \to \infty} T_n = \infty, \qquad \lim_{x_{((1-(b-\varepsilon))n)} \to \infty} T_n < \infty$$

توجه داشته باشید که این شرایط در مورد نقاط نمونه‌ای $x_{(1)}, \ldots, x_{(n)}$ است، نه متغیرهای $X_{(1)} < \cdots < X_{(n)}$.

---

16 -Contamination datas

17 -Breakdown points

الف) اگر $T_n = \bar{x}_n$ میانگین نمونه‌ای باشد، می توان نشان داد که $b = 0$ است.

برای این منظور فقط شرایط اول را در نظر می‌گیریم. بنابراین برای $b = 0$، شرط اول به $x_{(n)}$ وابسته است و داریم:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} x_{(i)}}{n} = \frac{\sum_{i=1}^{n-1} x_{(i)}}{n} + \frac{x_{(n)}}{n}$$

هنگامی که $x_{(n)} \to \infty$، اولین خط کسری(نسبت بیان شده در بالا) ثابت باقی می‌ماند و دومین نسبت به سمت ∞ میل می‌کند، بنابراین برای $b = 0$، $\lim T_n = \infty$ است. در نتیجه مقدار نقطه فروریزش برای میانگین نمونه‌ای صفر است.

ب) اگر $T_n = M_n$ میانه‌ی نمونه‌ای باشد، آنگاه $b = \frac{1}{2}$ است.

بنابراین برای $n=2m$ (nهای زوج) داریم:

$$M_n = \frac{1}{2}\left(x_{(m)} + x_{(m+1)}\right)$$

برای $b = \frac{1}{2}$ اولین شرط عبارت است از:

$$x_{\left(\frac{n}{2}\right)} = x_{(m)}$$

$$, x_{(m)} \to \infty$$

$$\Rightarrow x_{(m+1)} \to \infty$$

بنابراین

$$\mathrm{M_n} \to \infty$$

$$\Rightarrow \lim T_n = \infty$$

حال اگر n فرد باشد. داریم

$$M_{\mathrm{n}} = x_{\left((n+1)/2\right)}$$

برای $b = \frac{1}{2}$، از آنجا که $\frac{(n+1)}{2} > \frac{(n)}{2}$ است، بنابراین

$$M_{\mathrm{n}} = x_{\left((n+1)/2\right)} \to \infty$$

برای بررسی کردن شرط دوم $\varepsilon > 0$ را ثابت قرار داده، در نتیجه

$$\left(1 - (b - \varepsilon)\right)n = \left(\frac{n}{2}\right) + \varepsilon n$$

حال برای $n > \frac{1}{\varepsilon}$ داریم:

$$\left(\frac{n}{2}\right) + \varepsilon n > \frac{n+1}{2} \quad \text{و} \quad \left(\frac{n}{2}\right) + \varepsilon n > \frac{n}{2} + 1$$

بنابراین چه مقدار n زوج باشد و چه مقدار آن فرد باشد، آماره‌های ترتیبی‌ای که به ∞ میل می‌کنند مقدار حد M_{n} است. بنابراین نقطه فروریزش برای میانه $\frac{1}{2}$ است.

توجه داشته باشید که نقطه فروریزش نمی‌تواند از50% تجاوز کند، در صورت رخ‌دادن چنین حالتی می‌گوییم داده‌ها، آلوده شده‌اند. بنابراین ماکسیم نقطه فروریزش 0/5 است. برای مثال، میانه دارای یک نقطه فروریزش 0/5 است و میانگین پیراسته qام، دارای یک نقطهٔ فروریزش q است. آماره‌هایی با نقاط فروریزش زیاد آماره‌های استوار نامیده می‌شوند. (کسلا و برگر1900)