بسم الله الرحمن الرحيم

کلیه حقوق مادی مترتب بر نتایج مطالعات، ابتکارات و
نوآوری‌های ناشی از تحقیق موضوع این پایان‌نامه
متعلق به دانشگاه رازی است.

دانشکده علوم پایه
گروه آمار

پایان‌نامه جهت اخذ درجه کارشناسی ارشد
رشته آمار ریاضی

عنوان :

آزمون هابی درباره ی چندک های جامعه با استفاده از نمونه گیری مجموعه رتبه دار

استاد راهنما:
دکتر بهاء الدین خالدی

نگارش:
سمیه امیریان

بهمن ۱۳۹۰

دانشکده علوم پایه
گروه آمار

پایان‌نامه جهت اخذ درجه کارشناسی ارشد
رشته آمار ریاضی

نام دانشجو:
سمیه امیریان

تحت عنوان :

آزمون هایی درباره ی چندک های جامعه با استفاده از نمونه گیری مجموعه رتبه دار

در تاریخ ۱۳۹۰/۱۲/۳ توسط هیأت داوران زیر بررسی و با درجه به تصویب نهایی رسید.

استاد راهنمای پایان‌نامه دکتر بهاء الدین خالدی با مرتبه‌ی علمی امضاء:

استاد داور داخل گروه دکتر داود قزوینی نژاد با مرتبه‌ی علمی امضاء:

استاد داور خارج گروه دکتر خلیل شفیعی با مرتبه‌ی علمی امضاء:

... و در آغاز هیچ نبود کلمه بود و آن کلمه خدا بود ...

مرا کسی نساخت، خدا ساخت، نه آنچنان که کسی می خواست، که من کسی نداشتم کسم خدا بود خدا را سپاس. الهی تو را سپاس که به برگیری قطره ای از اقیانوس بیکران علم و دانشت یاریمان دادی و در بیت زیر سکنایمان کرد:

آب دریا را اگر نتوان کشید هم به قدر تشنگی باید چشید

اکنون که با لطف و یاری پروردگار انجام این رساله به اتمام رسیده است، بر خود لازم می دانم از کلیه عزیزانی که در این راه مرا یاری نموده اند صمیمانه قدردانی نمایم، از جناب آقای دکتر بهاء الدین خالدی استاد عزیزم که با نظرات و پیشنهادات ارزنده و زحمات بی دریغشان در طی مراحل مختلف اجرا، تدوین و ارائه رساله مرا یاری نموده اند صمیمانه قدردانی می نمایم. از پدر و مادر فداکارم که در کلیه مراحل تحصیل مشوق و راهنمای من بودند و با نور شمع وجودشان، روشنی بخش راهم گردیدند، خاضعانه سپاسگزارم و بوسه بر دستان پر مهرشان می نهم که هر چه دارم از وجود پاک و مهربان آنهاست و به یکایک تارهای سپید مویشان هزاران دین دارم. نهایت سپاس خود را به برادران عزیزم پیشکش می نمایم که وجودشان مرا سراسر لطف بوده است و مهربانی.

سپاسگزار زحمات بی دریغ اساتید بزرگوار آقای دکتر سیاره، آقای دکتر هاشمی، آقای دکتر نیاپرست، آقای دکتر قزوینی نژاد، آقای دکتر شفیعی و خانم دکتر شرفی هستم.

سمیه امیریان

کرمانشاه- بهمن ماه ۱۳۹۰

تقدیم به پدر و مادر دلسوز

همسر عزیز
و برادران مهربانم

آنان که آفتاب مهرشان در آستانه ی
قلبم همیشه پابرجاست و هرگز
غروب نخواهد کرد.

سمیه امیریان

بهمن ۱۳۹۰

چکیده

نمونه گیری مجموعه رتبه دار یک روش نمونه گیری است که نسبت به نمونه گیری تصادفی ساده از نظر کاهش هزینه و زمان دارای اهمیت است. این روش زمانی مفید است که اندازه گیری واحدهای نمونه گیری، هزینه بر یا سخت باشد ولی رتبه بندی واحدها در یک مجموعه کوچک، آسان و ارزان است. در این پایان نامه طرح های مختلفی از نمونه گیری مجموعه رتبه دار شامل نمونه گیری مجموعه رتبه دار استاندارد، طرح مشاهدات دنباله ای، طرح میان دامنه، طرح نمونه ای ثابت، طرح نمونه مجموعه مرتب شده رتبه دار به منظور به دست آوردن کارایی بیشتر، معرفی می شود و همچنین از آزمون علامت به عنوان یکی از روشهای اساسی در ناپارامتری، برای آنالیز داده های به دست آمده بر اساس این طرحهای نمونه گیری برای آزمون میانه استفاده می کنیم و سپس با معرفی یک آماره آزمون علامت وزنی، یک آزمون علامت بهینه برای میانه جامعه به دست می آوریم و با استفاده از نتایج به دست آمده در مورد میانه، چندک های جامعه را مورد آزمون قرار می دهیم و آزمون علامت بهینه را برای چندک p –ام جامعه بدست می آوریم. در نهایت نمونه گیری مجموعه رتبه دار با متغیر همراه و نمونه گیری مجموعه رتبه دار دو متغیره را معرفی می کنیم و خواصی از آزمون علامت بر اساس این روشهای نمونه گیری مطالعه می شود.

واژه های کلیدی: مشاهدات دنباله ای، میان دامنه، طرح نمونه ای ثابت، نمونه مجموعه مرتب شده رتبه دار، کارایی، آزمون علامت، متغیر همراه.

فهرست مطالب

عنوان صفحه

۱ تعاریف ومفاهیم اولیه ۱

۱-۱ برخی ویژگیهای آماره های ترتیبی ۲

۱-۲ نمونه گیری مجموعه رتبه دار ۴

۱-۲-۱ برآورد میانگین با نمونه گیری مجموعه رتبه دار ۹

۱-۳ آزمون علامت ۱۵

۱-۴ کارایی مجانبی پیتمن ۱۷

۲ آزمون علامت برای میانه جامعه بر اساس نمونه مجموعه رتبه دار ۲۴

۲-۱ آزمون علامت بر اساس SRS و RSS ۲۵

۲-۱-۱ آزمون علامت بر اساس نمونه تصادفی ساده(SRS) ۲۵

۲-۱-۲ آزمون علامت بر اساس RSS استاندارد ۲۶

۲-۱-۳ مقایسه کارایی آزمونها بر اساس SRS و RSS ۳۰

۲-۲ آزمون علامت بر اساس طرحهای مختلف RSS ۳۱

۲-۲-۱ طرح مشاهدات دنباله ای ۳۱

۲-۲-۲ طرح میان دامنه ۳۴

۲-۲-۳ طرح نمونه ای ثابت ۳۸

۲-۳ آزمون علامت بر اساس ROSS ۴۱

۳ آزمون علامت بهینه برای چندک های جامعه بر اساس نمونه مجموعه رتبه دار ۴۸

۳-۱ آزمون علامت بهینه برای میانه جامعه بر اساس RSS ۴۹

۳-۲ آزمون علامت بهینه برای چندک p –ام جامعه ۵۷

۴ آزمون علامت جفتی با استفاده از نمونه گیری مجموعه رتبه دار دو متغیره . . ۶۸

۴-۱ نمونه گیری مجموعه رتبه دار دو متغیره ۶۹

۴-۱-۱ برآورد میانگین با استفاده از BVRSS ۷۵

۴-۲ نمونه گیری مجموعه رتبه دار با متغیر همراه ۷۷
۴-۳ آزمون علامت جفتی بر اساس نمونه مجموعه رتبه دار دو متغیره ۸۰

آ فهرست قضایای استفاده شده ۸۷
آ-۱ فهرست قضایا ۸۸

منابع و مآخذ ۸۹

واژه‌نامه فارسی به انگلیسی ۹۱

واژه‌نامه انگلیسی به فارسی ۹۳

پیشگفتار

یکی از مهمترین قسمت های علم آمار روش نمونه گیری است و به کارگیری یک روش نمونه گیری خوب همواره مورد علاقه آمارشناسان بوده است، به خصوص زمانی که اندازه مشخصه موردنظر هزینه بر یا وقت گیر است.

نمونه گیری مجموعه رتبه دار (RSS) یک روش نمونه گیری است که ممکن است برای افزایش دقت یک برآوردگر به کار گرفته شود. این روش زمانی مفید است که اندازه گیری واحدهای نمونه گیری، هزینه بر یا سخت باشد ولی رتبه بندی واحدها در یک مجموعه کوچک، آسان و ارزان است.

این روش نمونه گیری شامل انتخاب k^2 واحد به طور تصادفی از جامعه است. این واحدها به k مجموعه، هر کدام با k واحد تقسیم می شوند. بدون هیچ گونه اندازه گیری، واحدهای هر مجموعه را رتبه بندی می کنیم. این رتبه بندی می تواند بر اساس تجهیزات بصری یا قضاوت شخصی و یا هر وسیله دیگری صورت پذیرد. اکنون در مجموعه $r-$ ام، واحدی که دارای رتبه $r-$ ام است را برای اندازه گیری مشخصه مورد نظر انتخاب می کنیم $r = ۱, ..., k$. این فرآیند را یک چرخه می نامیم. اگر این چرخه را m بار تکرار کنیم آنگاه یک نمونه مجموعه رتبه دار به اندازه mk به صورت $\{X_{(r)i}; r = ۱, ..., k; i = ۱, ..., m\}$ در اختیار خواهیم داشت.

روش نمونه گیری مجموعه رتبه دار اولین بار توسط مک این تایر (۱۹۵۲) به منظور پیدا کردن یک روش با کارایی بیشتر برای برآورد میانگین محصول چراگاه معرفی شد. هر چند که او نتوانست یک نظریه ریاضی برای آن ارائه دهد اما با این وجود، او را به عنوان اولین کسی که نمونه گیری مجموعه رتبه دار را معرفی کرد، می شناسند. تا اینکه در سال ۱۹۶۸ تاکاهاشی و واکی موتو (۱۹۶۸) توانستند یک نظریه ریاضی برای آن ارائه دهند. بعد از آن روش های آماری زیادی شامل روش های پارامتری و ناپارامتری بر اساس RSS بررسی شد. اخیرا آزمون علامت به عنوان یکی از روش های اساسی در ناپارامتریک، برای آنالیز داده های RSS توسط برخی محققین انجام شده است. هتمانسپرگر ۱۹۹۵ برای اولین بار به مقایسه کارایی آزمون علامت بر اساس نمونه مجموعه رتبه دار و نمونه تصادفی ساده پرداخت. بعدها صورت های مختلفی از نمونه گیری مجموعه رتبه دار معرفی شد و کارایی این روش ها با نمونه گیری تصادفی ساده مقایسه شد. در فصل اول ابتدا برخی مفاهیم اولیه نظیر آماره های ترتیبی، نمونه گیری مجموعه رتبه دار، کارایی و آزمون علامت را بیان خواهیم کرد.

در فصل دوم طرح های مختلفی از نمونه گیری مجموعه رتبه دار شامل نمونه گیری مجموعه رتبه دار استاندارد، طرح مشاهدات دنباله ای، طرح میان دامنه، طرح نمونه ای ثابت، طرح نمونه مجموعه مرتب شده رتبه دار معرفی خواهد شد و به مطالعه آزمون علامت برای میانه جامعه بر اساس این طرح های نمونه گیری خواهیم پرداخت.

در فصل سوم یک آزمون علامت بهینه با بیشترین کارایی برای میانه جامعه به دست خواهیم آورد و با استفاده از نتایج به دست آمده در مورد میانه، چندک های جامعه را مورد آزمون قرار خواهیم داد.

در نهایت در فصل چهارم صورت های مختلفی از روش نمونه گیری مجموعه رتبه دار با نام نمونه گیری مجموعه رتبه دار با متغیر همراه و نمونه گیری مجموعه رتبه دار دو متغیره معرفی خواهد شد و کارایی آزمون علامت بر اساس این روش ها را با روش نمونه تصادفی ساده مقایسه خواهیم کرد.

فصل ۱

تعاریف ومفاهیم اولیه

این فصل در بردارنده تعاریف و مفاهیمی است که در فصل های بعدی مورد استفاده قرار می گیرند. بخش ۱ – ۱ به برخی ویژگی های آماره های ترتیبی اختصاص دارد. در بخش ۱ – ۲ روش نمونه گیری مجموعه رتبه دار را معرفی می کنیم و به مساله برآورد میانگین جامعه با این نوع نمونه گیری خواهیم پرداخت. در بخش ۱ – ۳ یکی از آزمون های ناپارامتری به نام آزمون علامت را معرفی خواهیم کرد و در نهایت در بخش ۱ – ۴ کارایی را که به عنوان یکی از معیارهای مقایسه آزمونها شناخته شده است را معرفی می کنیم.

۱ - ۱ برخی ویژگیهای آماره های ترتیبی

مطالب این بخش از دیوید [۱](۱۹۸۱) اقتباس شده است.
فرض کنید $X_۱$ ، $X_۲$ ، ... ، X_k متغیرهای تصادفی پیوسته مستقل با توزیع یکسان و دارای چگالی مشترک f و تابع توزیع تجمعی F هستند. اگر این متغیرها را به صورت افزایشی مرتب کرده و به صورت:

$$X_{(۱)} \leq X_{(۲)} \leq ... \leq X_{(k)}$$

نمایش دهیم، آنگاه این مقادیر مرتب شده به عنوان آماره های ترتیبی مربوط به متغیرهای تصادفی $X_۱$ ، $X_۲$ ، ... ، X_k شناخته می شوند. تابع چگالی توام آماره های ترتیبی به صورت زیر است:

$$f_{X_{(۱)},...,X_{(k)}}(x_۱, ..., x_k) = k! f(x_۱), ..., f(x_k) \quad x_۱ \leq x_۲ \leq ... \leq x_k$$

تابع چگالی $r-$ امین آماره ترتیبی را می توان به وسیله انتگرال گیری از تابع چگالی توام به دست آورد. تابع چگالی $X_{(r)}$ به صورت زیر است:

$$f_{(r)}(x) = k \binom{k-۱}{r-۱} F^{r-۱}(x)(۱ - F(x))^{k-r} f(x) \qquad (۱\text{-}۱)$$

لم زیر یک ویژگی مهم را برای آماره های ترتیبی بیان می کند که به وفور از آن استفاده خواهیم کرد.

لم ۱ - ۱ - ۱. *برای هر* $x \in R$ *داریم*

$$\frac{۱}{k} \sum_{r=۱}^{k} f_{(r)}(x) = f(x) \qquad (۱\text{-}۲)$$

و آن را برابری اساسی می نامیم.

[۱] David

برهان. با استفاده از (۱.۱) برای هر $x \in R$ داریم

$$\begin{aligned}
\frac{۱}{k}\sum_{r=۱}^{k} f_{(r)}(x) &= \frac{۱}{k}\sum_{r=۱}^{k} k\binom{k-۱}{r-۱}F^{r-۱}(x)(۱-F(x))^{k-r}f(x) \\
&= f(x)\sum_{r=۱}^{k}\binom{k-۱}{r-۱}F^{r-۱}(x)(۱-F(x))^{k-r} \\
&= f(x)\sum_{r=۰}^{k-۱}\binom{k-۱}{r}F^{r}(x)(۱-F(x))^{k-۱-r} \\
&= f(x)(F(x)+۱-F(x))^{k-۱} \\
&= f(x)
\end{aligned}$$

□

اکنون نمادهای زیر را در نظر بگیرید:

$$\mu = E(X)$$

$$\mu_{(r)} = E(X_{(r)})$$

$$\sigma^{۲} = E(X-\mu)^{۲}$$

$$\sigma^{۲}_{(r)} = E(X_{(r)}-\mu_{(r)})^{۲}$$

در این صورت با استفاده از برابری اساسی (۲.۱) داریم:

$$\frac{۱}{k}\sum_{r=۱}^{k}\mu_{(r)} = \mu$$

بنابراین

$$\sum_{r=۱}^{k}(\mu_{(r)}-\mu) = ۰ \qquad (۳\text{-}۱)$$

یک ویژگی مهم دیگر برای آماره های ترتیبی در لم زیر داده شده است:

لم ۱-۱-۲. *اگر $F_{(r)}(x)$ نمایانگر تابع توزیع r-امین آماره ترتیبی از یک نمونه تصادفی k تایی از توزیع F باشد آن گاه داریم*

$$F_{(r)}(x) = B(r, k-r+۱, F(x))$$

که در آن $B(r,s,x)$ نمایانگر تابع توزیع تجمعی برای توزیع بتا با پارامترهای r و s است.

برهان. ابتدا از رابطه(۱.۱) داریم

$$F_{(r)}(x) = \int_{-\infty}^{x} \frac{k!}{(r-۱)!(k-r)!} F^{r-۱}(t)(۱-F(t))^{k-r} dF(t)$$

سپس با تغییر متغیر $F(t) = u$ به این نتیجه می رسیم که

$$\begin{aligned} F_{(r)}(x) &= \int_{۰}^{F(x)} \frac{k!}{(r-۱)!(k-r)!} u^{r-۱}(۱-u)^{k-r} du \\ &= B(r, k-r+۱, F(x)) \end{aligned}$$

بنابراین برهان کامل است. □

۱ - ۲ نمونه گیری مجموعه رتبه دار

مطالب این بخش از مک این تایر [۱](۱۹۵۲) و پاتیل [۲] و همکاران (۱۹۹۹) اقتباس شده است.
یکی از قسمتهای کلیدی هر استنباط آماری این است که داده ها از یک مکانیسم مشخص طوری انتخاب شوند که آزمایشگر بتواند قضاوت های درست و معتبری در مورد سوالهای موردنظر به دست آورد. یکی از مکانیسم های متداول و معتبر برای جمع آوری داده ها، نمونه گیری تصادفی ساده است. در این روش نمونه گیری، داده ها به صورت تصادفی از جامعه انتخاب می شوند. توجه کنید که در این روش هیچ کنترلی روی اعضای جامعه برای ورود به نمونه وجود ندارد. در واقع همه چیز به شانس واگذار می شود. یک مکانیسم دیگر، نمونه گیری با طبقه بندی است که در آن، طبقات به گونه ای در نظر گرفته می شوند که همگنی درون طبقات زیاد است و از یک طبقه به طبقه دیگر تفاوت ها بیشتر می شود. برخی دیگر از این مکانیسم ها، نمونه گیری خوشه ای، نمونه گیری سیستماتیک و غیره است. با استفاده از هر یک از مکانیسم های ذکر شده در بالا می توان واحدهای نمونه را انتخاب کرده و مشخصه های مورد نظر را برای هر واحد نمونه اندازه گیری کرد. برای وضعیتهایی که اندازه گیری واحدهای نمونه هزینه بر، مخرب و یا وقت گیر اما رتبه بندی واحدها در یک مجموعه کوچک، آسان و ارزان باشد می توان از نمونه گیری مجموعه رتبه دار استفاده کرد. به عنوان یک مقدمه ساده برای معرفی این نوع نمونه گیری، مثال های زیر را در نظر بگیرید:

[۱] McIntyre

[۲] Patil

مثال ۱.۱: فرض کنید می خواهیم میانگین قد دانشجویان یک دانشگاه را با یک نمونه تصادفی شامل سه دانشجو برآورد می کنیم. ساده ترین راه، استفاده از نمونه تصادفی ساده است به این صورت که به طور تصادفی سه دانشجو را از میان دانشجویان دانشگاه انتخاب کرده و قد آنها را اندازه گیری می کنیم. در اینصورت میانگین حسابی قد این سه دانشجو، یک برآوردگر نااریب برای میانگین جامعه است. توجه کنید که در این روش، هیچ کنترلی روی اعضای جامعه برای ورود به نمونه صورت نمی گیرد. برای مثال ممکن است یک دانشجوی بلندقد و دو دانشجوی کوتاه قد انتخاب شوند و یا حتی سه دانشجوی بلندقد انتخاب شوند. بنابراین ممکن است برآوردهای خوبی نداشته باشیم. یک راه حل برای رفع این مشکل، افزایش اندازه نمونه است.

از طرف دیگر ممکن است از روش نمونه گیری مجموعه رتبه دار استفاده کنیم. به اینصورت که ابتدا به طور تصادفی سه دانشجو را برای صرف صبحانه دعوت می کنیم و آن ها را با مقایسه قدشان از طریق چشم رتبه بندی کنیم سپس دانشجوی دارای کوتاه ترین قد را به عنوان اولین عضو نمونه انتخاب کرده و قد او را اندازه گیری کنیم. اکنون برای نهار نیز سه دانشجو را به تصادف دعوت می کنیم و پس از رتبه بندی قد آنها، دانشجوی دارای رتبه دوم قد را به عنوان دومین عضو نمونه انتخاب کرده و قد او را نیز اندازه می گیریم. در نهایت برای شام نیز به طور مشابه عمل می کنیم ولی این بار بلندقدترین دانشجو را انتخاب کرده و قد او را به عنوان سومین عضو نمونه اندازه گیری می کنیم.

دو استفاده از روش نمونه گیری مجموعه رتبه دار در زیر آورده شده است.

مثال ۲.۱: یک استفاده از روش نمونه گیری مجموعه رتبه دار ارزیابی محدوده حرکت و چرای دام است. واحدهای نمونه گیری در چنین برآوردی به صورت $quadrats$ (قطعات زمین مستطیل شکل) تعریف شده اند. نحوه اندازه گیری هر $quadrats$ شامل علف چیده شده در داخل زمین و سپس خشک کردن و وزن کردن علوفه می باشد که روندی کاملا زمان بر و مخرب است. با این حال، می توان قطعات نزدیک بهم را نسبتا دقیق باقضاوت شخصی و یا بازدید دیداری رتبه بندی کرد. حال اگر تغییرات بین قطعاتی که در کنار هم قرار گرفته اند با تغییرات بین قطعات دور از هم ،یکسان باشند، آن گاه قطعات نزدیک به هم انتخاب شده، می توانند به عنوان نمونه تصادفی در نظر گرفته شوند. در چنین شرایطی RSS می تواند بهره وری (کارایی) را به میزان قابل توجهی افزایش دهد.

انتشارات $Halls$ و $Dell$ گزارشی از یک آزمایش منتشر کرده که در آن از روش RSS برای برآورد محصول علوفه جنگل های $shortleaf - loblolly$ (کاج کوتاه برگ) $pine - hardwood$ (کاج پهن برگ) در ناحیه ای به وسعت ۳۰۰ جریب در جنگلی در نزدیکی تگزاس، ایالات متحده استفاده شده است. در این تحقیق قطعات زمین به وسیله قاب های فلزی به مساحت ۳.۱ فوت مربع معین شده بودند. اندازه گیری یک$quadrats$ شامل چیدن $herbage$ (علوفه) و $browse$ (گیاهان قابل چرا) در سطح زمین و سپس

خشک کردن علوفه برای رسیدن به وزن ثابت در دمای $۷۰^{\circ}C$ می شود. فرآیند نمونه گیری به صورت زیر است: ابتدا مجموعه مکان ها توسط یک توری زنجیری معین شده اند، سپس در نقاط تصادفی در داخل دایره ای به شعاع ۱۳ فوت در هر مکان ۳ قاب فلزی روی زمین قرار داده شده است. سپس ۲ نفر هر کدام مستقلا ۳ قطعه زمین را رتبه بندی می کنند و همچنین علوفه ها و گیاهان قابل چرا هر کدام جداگانه رتبه بندی می شوند.برای $browse$ ، ۱۲۶ مجموعه و برای $herbage$ ، ۱۲۴ مجموعه رتبه بندی شده و چیده شده اند. مشخص شد که برای دستیابی به یک دقت یکسان در برآورد ۱۰۰ قطعه زمین انتخاب شده ،برای $browse$ ها تنها ۴۸ قطعه نیاز به اندازه گیری دارند و برای $herbage$ تنها ۷۳ قطعه زمین به اندازه گیری نیاز دارند.

مثال ۳.۱: یک استفاده دیگر از RSS برای برآورد میزان تراکم بوته ها در جنگل بلوط بود. در این مطالعه نمونه گیری شامل محاسبه شمار هر نوع پوشش گیاهی در بلوک هایی که به طور تصادفی انتخاب شده اند، می باشد، که نسبتا زمان بر است. اما رتبه بندی تعداد کمی از بلوک ها با بازرسی دیداری می تواند به آسانی انجام گیرد.

در دو مثال بالا رتبه بندی تعداد کمی از واحد های نمونه گیری با بازرسی دیداری انجام گرفته است. برای اینکه بازرسی چشمی امکان پذیر باشد، واحدهای نمونه گیری باید نزدیک به هم باشند. همانطور که در مثال ۱ ذکر شد اگر تغییرات بین واحدهای نزدیک بهم با تغییرات در بین واحدهای دور از هم یکسان باشد، یک مجموعه از واحدهای نمونه گیری نزدیک بهم می تواند به عنوان نمونه تصادفی در نظر گرفته شود. با این حال اگر تغییرات در بین واحدهای نزدیک بهم خیلی کوچکتر از تغییرات در بین واحدهای دور از هم باشد، دیگر یک مجموعه از واحدهای نزدیک بهم یک نمونه تصادفی از جمعیت مورد مطالعه نخواهد بود. برای غلبه بر این مشکل می توان تمام ناحیه مورد مطالعه را به قسمتهایی تقسیم کرد که در آن پراکندگی بومی هر قسمت تا اندازه بسیار دقیقی با سایر پراکندگی های دیگر یکسان می باشد و سپس از روش RSS برای هر قسمت استفاده کرد.

اکنون به طور رسمی نمونه گیری مجموعه رتبه دار را معرفی می نماییم. مکانیسم نمونه گیری مجموعه رتبه دار به صورت زیر است:

گام ۱: ابتدا $k^{۲}$ نمونه تصادفی را از جامعه مورد نظر انتخاب می کنیم. توجه کنید که هیچگونه اندازه گیری برای مشخصه مورد نظر صورت نمی گیرد.

گام ۲: k^2 واحد انتخاب شده را به طور تصادفی به k مجموعه با اندازه برابر به صورت زیر تقسیم می نماییم:

$$\begin{array}{llllll} X_{11} & , & X_{21} & , \ldots & X_{k1} & \text{مجموعه۱} \\ X_{12} & , & X_{22} & , \ldots & X_{k2} & \text{مجموعه۲} \\ \ldots & & \ldots & \ldots & \ldots & \\ X_{1k} & , & X_{2k} & , \ldots & X_{kk} & \text{مجموعهk} \end{array}$$

که در آن متغیر تصادفی X_{ij} معرف شاخص مورد نظر برای واحد i–ام در مجموعه j–ام است.

گام ۳: بدون هیچگونه آگاهی از مقادیر متغیر مورد نظر، واحدها را درون هر مجموعه رتبه بندی می کنیم. این رتبه بندی می تواند بر اساس تجهیزات بصری یا قضاوت شخصی و یا هر وسیله دیگری صورت پذیرد. توجه کنید که اگر اندازه مجموعه (k) بزرگ باشد آن گاه این امکان وجود دارد که رتبه بندی با خطا همراه باشد. بنابراین k را کوچک در نظر می گیریم. k معمولا ۳، ۴ یا ۵ است. اکنون با رتبه بندی واحدهای درون هر مجموعه داریم

$$\begin{array}{llllll} X_{(1)1} & , & X_{(2)1} & , \ldots & X_{(k)1} & \text{مجموعه۱} \\ X_{(1)2} & , & X_{(2)2} & , \ldots & X_{(k)2} & \text{مجموعه۲} \\ \ldots & & \ldots & \ldots & \ldots & \\ X_{(1)k} & , & X_{(2)k} & , \ldots & X_{(k)k} & \text{مجموعهk} \end{array}$$

گام ۴: ابتدا از مجموعه ۱، واحدی که دارای کوچکترین رتبه و از مجموعه ۲، واحدی که دارای رتبه دوم و به همین ترتیب از مجموعه k، واحدی که دارای بزرگترین رتبه است را انتخاب می نماییم. سپس مشخصه مورد نظر را برای واحدهای نمونه انتخاب شده، اندازه گیری کرده و آن را یک نمونه مجموعه رتبه دار به اندازه k می نامیم و به صورت زیر نمایش می دهیم:

$$X_{(1)1}, X_{(2)2}, ..., X_{(k)k}$$

توجه کنید که گرچه در ابتدا k^2 واحد را انتخاب کردیم ولی در نهایت تنها k واحد را اندازه گیری و به عنوان اعضای نمونه RSS در نظر می گیریم.

گام ۵: فرایند مربوط به گام های ۱ تا ۴ را یک چرخه می نامیم. با توجه به گام ۳ برای دستیابی به یک نمونه با اندازه بیشتر، می توان چرخه را به تعداد لازم تکرار کرد. بنابراین با m بار تکرار چرخه، یک نمونه مجموعه رتبه دار به اندازه mk به صورت زیر در اختیار داریم:

$$\{X_{(r)j}\ ; r = ۱, ..., k\ ; j = ۱, ..., m\} \qquad (۴-۱)$$

ملاحظات: ۱- با توجه به گام های ۱ تا ۵ ملاحظه می شود که هر مشاهده در نمونه مجموعه رتبه دار با اینکه آماره ترتیبی است ولی این مشاهدات مستقل اند و این استقلال به این دلیل است که آماره های ترتیبی از مجموعه های مستقل و متفاوتی انتخاب شده اند. ۲- از آنجایی که امکان دارد رتبه بندی با خطا همراه باشد لذا تعریف زیر را ارائه می دهیم:

تعریف ۱-۲-۱. اگر در مکانیسم رتبه بندی مشاهده دارای رتبه r-ام برابر r-امین آماره ترتیبی باشد آن گاه می گوییم رتبه بندی کامل است در غیر اینصورت می گوییم رتبه بندی ناقص است.

برای یک بحث بیشتر در مورد رتبه بندی ناقص و اثر آن بر روش نمونه گیری مجموعه رتبه دار به دل و کلاتر (۱۹۷۲) و دیوید و لواین (۱۹۷۲) مراجعه کنید.
۳- در برخی موارد رتبه بندی متغیر مورد نظر نیز مشکل است. در اینصورت از یک متغیر همراه برای رتبه بندی استفاده می کنیم که جزئیات مربوط به آن را در فصل چهارم بیان خواهیم کرد.
در ادامه تاریخچه نمونه گیری مجموعه رتبه دار را به اختصار بیان می کنیم.
ایده نمونه گیری مجموعه رتبه دار اولین بار توسط مک این تایر(۱۹۵۲) بیان شد. او این شیوه نمونه گیری را برای برآورد میانگین محصول یونجه مراتع به کار گرفت و به صورت تجربی به این نتیجه دست یافت که این روش نمونه گیری بهتر از روش نمونه گیری تصادفی ساده است اما نتوانست یک نظریه ریاضی برای آن ارائه دهد تا اینکه در سال ۱۹۶۸ تاکاهاشی و واکی موتو (۱۹۶۸) توانستند یک نظریه جالب برای آن ارائه دهند. آن ها نشان دادند که تحت رتبه بندی کامل، میانگین نمونه مجموعه رتبه دار یک برآوردگر نااریب برای میانگین جامعه است و همچنین واریانس آن کمتر از واریانس میانگین نمونه تصادفی ساده است. استوکس در سال ۱۹۷۷ استفاده از متغیر همراه را برای نمونه گیری مجموعه رتبه دار پیشنهاد کرد. در سال ۱۹۹۵ هتمانسپرگر (۱۹۹۵) آزمون علامت بر اساس RSS را معرفی کرد. در سال ۲۰۰۳ بای و چن (۲۰۰۳) نظریه پارامتری و ناپارامتری را برای نمونه گیری مجموعه رتبه دار توسعه دادند. در سالهای اخیر نیز پیشرفت های قابل ملاحظه ای در زمینه نمونه گیری مجموعه رتبه دار انجام شده است که برای آگاهی از آن ها می توانید به پاتیل و همکاران (۱۹۹۹) و چن (۲۰۰۷) مراجعه نمایید.