بسم الله الرحمن الرحیم

دانشکده علوم پایه
گروه آمار

پایان‌نامه جهت اخذ درجه کارشناسی ارشد
رشته آمار ریاضی

عنوان :

فواصل اطمینان آزاد توزیع برای چندک های جامعه بر اساس آماره های ترتیبی از نمونه مجموعه رتبه دار

استاد راهنما:
دکتر بهاء الدین خالدی

نگارش:
حسنا حاتمی

بهمن ۱۳۹۰

دانشکده علوم پایه

گروه آمار

پایان‌نامه جهت اخذ درجه کارشناسی ارشد

رشته آمار ریاضی

نام دانشجو:

حسنا حاتمی

تحت عنوان :

فواصل اطمینان آزاد توزیع برای چندک های جامعه بر اساس آماره های ترتیبی از نمونه مجموعه رتبه دار

در تاریخ ۱۳۹۰/۱۱/۱۱ توسط هیأت داوران زیر بررسی و با درجه به تصویب نهایی رسید.

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

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

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

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

مرا کسی نساخت، خدا ساخت، نه آنچنان که کسی خواست، که من کس نداشتم، کس من خدا بود، خدا را سپاس.

الهی تو را سپاس که به برگیری قطره ای از اقیانوس بی کران دانشت یاریمان دادی و در بیت زیر سکنایمان که:

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

اکنون که با لطف و یاری پروردگار متعال 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 ارایه رساله مرا یاری نمودند صمیمانه قدر دانی 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.

نهایت سپاس خود را به خواهران عزیزم پیشکش می نمایم که وجودشان مرا سراسر لطف بوده است و مهربانی.

سپاسگزار زحمات بی دریغ اساتید بزرگوارم آقای دکتر سیاره، آقای دکتر هاشمی، آقای دکتر نیاپرست و آقای دکتر قزوینی نژاد و هم چنین خانم دکتر شرفی، هستم.

حسنا حاتمی

بهمن ۱۳۹۰

تقدیم به پدر و مادر عزیز

و خواهران مهربانم

آنان که آفتاب مهرشان در آستانه ی

قلبم همیشه پابرجاست و هرگز

غروب نخواهد کرد.

چکیده

در این رساله ابتدا روش نمونه گیری مجموعه رتبه دار معرفی می شود و بر اساس آماره های ترتیبی نمونه مجموعه رتبه دار یک فاصله اطمینان دقیق برای چندک p–ام جامعه ساخته می شود. سپس با استفاده از این فواصل اطمینان یک آزمون ناپارامتری دو نمونه ای برای تفاوت چندک بین دو جامعه به دست آورده می شود. یک برآورد نقطه ای، فاصله اطمینان و آزمون فرض آزاد توزیع برای چندک p–ام جامعه بر اساس طرح نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده بنا می شود. در نهایت یک استنباط دو نمونه ای آزاد توزیع بر اساس طرح نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده برای پارامتر تغییر مکان بین دو توزیع انجام می شود. نشان داده می شود که طرح نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده در برابر هر خطای رتبه بندی استوار است و کارایی بیشتری نسبت به طرح های رقیب یعنی نمونه مجموعه رتبه دار و نمونه تصادفی ساده در استنباط آماری به دست می دهد.

کلمات کلیدی:

آزمون ناپارامتری، خطای رتبه بندی، رتبه بندی قضاوتی، زیر مجموعه به طور جزئی مرتب شده

فهرست مطالب

عنوان | صفحه

۱ نمونه گیری مجموعه رتبه دار
تعاریف و مفاهیم اولیه ... ۱
۱-۱ آماره های ترتیبی ... ۲
۱-۲ نمونه گیری مجموعه رتبه دار ... ۵
۱-۲-۱ مکانیسم رتبه بندی ... ۷
۱-۲-۲ تاریخچه نمونه گیری مجموعه رتبه دار ... ۹
۱-۳ آماره های ترتیبی از نمونه مجموعه رتبه دار ... ۱۱
۱-۴ برآورد میانگین تابعی از مشاهدات با RSS ... ۱۵
۱-۵ برآورد چندک ها با نمونه گیری مجموعه رتبه دار ... ۱۹
۱-۵-۱ چندک های نمونه ای برای نمونه مجموعه رتبه دار ... ۱۹
۱-۵-۲ روش های استنباط برای چندک های جامعه با RSS ... ۲۱
۱-۵-۳ کارایی مجانبی RSS نسبت به SRS در برآورد چندک ها ... ۲۳
۱-۶ برآورد تابع چگالی بر اساس RSS ... ۲۵

۲ آزمون ناپارامتری دو نمونه ای دقیق برای اختلاف چندک بین دو جامعه بر اساس
نمونه مجموعه رتبه دار ... ۲۸
۲-۱ فاصله اطمینان آزاد توزیع برای چندک ها بر اساس ORSS ... ۲۹
۲-۲ آزمون دقیق دو نمونه ای ... ۳۳
۲-۳ رتبه بندی ناقص ... ۳۷
۲-۳-۱ برآوردگر ... ۳۸
۲-۳-۲ استفاده از برآوردگر به دست آمده در ساختن آزمون فرض ... ۴۰
۲-۴ یک مثال عددی ... ۴۴

۳ استنباط بر روی چندک های جامعه بر اساس نمونه گیری از زیر مجموعه های
به طور جزئی مرتب شده ... ۴۶

۱-۳ نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده ۴۷
۲-۳ استنباط بر روی چندک جامعه ۵۰
۳-۳ آزمون فرض ۵۹
۴-۳ شواهد تجربی ۶۳

۴ استنباط دو نمونه ای آزاد توزیع بر اساس نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده ۶۹
۱-۴ آزمون من ویتنی - ویلکاکسون ۷۰
۲-۴ استنباط دو نمونه ای بر اساس طرح PROSS ۷۵
۳-۴ آزمون فرض ۸۴
۴-۴ آزمون مجموع - رتبه طبقه بندی شده ۸۶
۵-۴ شواهد تجربی ۸۸

آ فهرست قضایای استفاده شده ۹۶
آ-۱ فهرست قضایا ۹۷

منابع و مآخذ ۹۹

واژه‌نامه فارسی به انگلیسی ۱۰۱

واژه‌نامه انگلیسی به فارسی ۱۰۳

پیشگفتار

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

مک این تایر (۱۹۵۲) [۱] برای اولین بار یک روش نمونه گیری را معرفی کرد که در سال های بعد به نمونه گیری مجموعه رتبه دار معروف 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. این روش نمونه گیری شامل انتخاب $n^۲$ واحد به طور تصادفی از جامعه است. این واحدها به n مجموعه، هر کدام با n واحد تقسیم می شوند. بدون هیچ 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 $(r = ۱, ..., n)$. این فرایند را یک سیکل می نامیم. اگر این سیکل را m بار تکرار کنیم آن گاه یک نمونه مجموعه رتبه دار به اندازه mn به صورت $\{X_{[r]i} : r = ۱, ..., n, i = ۱, ..., m\}$ در اختیار خواهیم داشت. فرایند رتبه بندی در نمونه گیری مجموعه رتبه دار در معرض خطاست. نمونه گیری مجموعه رتبه دار رتبه دهنده را مجبور می کند که حتی اگر نتواند با اطمینان بالا رتبه ی داده ها را مشخص کند در یک مجموعه به هر واحد یک رتبه را اختصاص دهد که این منجر به خطای رتبه بندی می شود. ازترک (۲۰۱۱) [۲] روش نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده را معرفی کرد. نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده شبیه به نمونه گیری مجموعه رتبه دار است با این تفاوت که در این روش ابتدا زیر مجموعه های به طور جزئی مرتب شده ایجاد می شود و داده های که نمی توان با اطمینان بالا به آنها یک رتبه معین را داد در یک زیر مجموعه قرار می گیرند سپس یک واحد برای اندازه گیری دقیق به طور تصادفی از یکی از این زیر مجموعه های به طور جزئی مرتب شده درون هر مجموعه انتخاب می شود.

در فصل اول از انجایی که نمونه مجموعه رتبه دار مشاهدات مستقل و هم توزیع نمونه تصادفی ساده را با آماره های ترتیبی مستقل و غیر هم توزیع جایگزین می کند ابتدا به بررسی مفهوم آماره های ترتیبی می پردازیم سپس نمونه مجموعه رتبه دار را معرفی می کنیم و آماره های ترتیبی از این روش نمونه گیری را معرفی می کنیم. در ادامه ی فصل نمونه گیری مجموعه رتبه دار را در حالت ناپارامتری مورد بررسی قرار می دهیم یعنی به مساله برآورد میانگین تابعی از مشاهدات، بررسی چندک های نمونه ای بر اساس نمونه

[۱] McIntyre

[۲] Ozturk

مجموعه رتبه دار و برآورد تابع چگالی جامعه بر اساس نمونه مجموعه رتبه دار می پردازیم.
در فصل دوم یک آزمون ناپارامتری دو نمونه ای دقیق برای اختلاف چندک بین دو جامعه انجام می دهیم و از آنجایی که این آزمون فرض بر اساس فواصل اطمینان درون یابی شده برای چندک های هر جامعه به طور مجزا، ساخته می شود ابتدا یک فاصله اطمینان آزاد توزیع برای چندک p–ام جامعه بر اساس آماره های ترتیبی نمونه مجموعه رتبه دار می سازیم. سپس از این فواصل اطمینان درون یابی شده برای ساختن آزمون فرض تفاوت چندک بین دو جامعه استفاده می کنیم.
در فصل سوم استنباط آماری را بر روی چندک های جامعه بر اساس نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده انجام می دهیم. بنابراین ابتدا روش نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده را معرفی می کنیم و بر اساس این روش نمونه گیری یک برآورد نقطه ای، فاصله اطمینان و آزمون فرض برای چندک p–ام جامعه می سازیم.
در نهایت در فصل چهارم یک استنباط دو نمونه ای آزاد توزیع بر اساس نمونه گیری از زیر مجموعه های به طور جزئی مرتب شده انجام می دهیم. یک برآوردگر نقطه ای، فاصله اطمینان و آزمون فرض برای پارامتر تغییر مکان بین توزیع دو جامعه می سازیم. نشان داده می شود که این طرح نمونه گیری در برابر هر خطای رتبه بندی استوار است و کارایی بیشتری در استنباط آماری نسبت به طرح های رقیب یعنی نمونه مجموعه رتبه دار و نمونه تصادفی ساده به دست می آورد.

فصل ۱

نمونه گیری مجموعه رتبه دار

تعاریف و مفاهیم اولیه

در این فصل به معرفی روش نمونه گیری مجموعه رتبه دار و مفاهیم مرتبط با آن می پردازیم و از آن جایی که این روش نمونه گیری مشاهدات مستقل و هم توزیع نمونه تصادفی ساده را با آماره های ترتیبی مستقل و غیر هم توزیع جایگزین می کند ابتدا در بخش اول این فصل به بررسی مفهوم آماره های ترتیبی می پردازیم و تابع توزیع و تابع چگالی r–امین آماره ترتیبی هم در حالتی که نمونه مستقل و هم توزیع باشد (iid) و هم در حالتی که نمونه مستقل اما غیر هم توزیع باشد $(inid)$ ارائه می دهیم. در بخش دوم به معرفی روش نمونه گیری مجموعه رتبه دار می پردازیم سپس یک تاریخچه مختصر درباره ی آن ارائه می دهیم و به اختصار نتایج به دست آمده در زمینه نمونه گیری مجموعه رتبه دار را بیان می کنیم. در بخش سوم آماره های ترتیبی از این روش نمونه گیری را معرفی می کنیم و به بررسی خواص توزیعی این آماره های ترتیبی می پردازیم. در بخش چهارم به مساله برآورد میانگین تابعی از مشاهدات با نمونه گیری مجموعه رتبه دار می پردازیم. در بخش پنجم به بررسی چندک ها از نمونه مجموعه رتبه دار و خواص آن ها می پردازیم. در بخش ششم یک برآوردگر سازگار برای تابع چگالی با استفاده از نمونه گیری مجموعه رتبه دار معرفی می کنیم.

۱ - ۱ آماره های ترتیبی

فرض کنید n متغیر تصادفی $X_1, X_2, ..., X_n$ را به ترتیب غیر نزولی مرتب کنیم و به صورت

$$X_{(1:n)} \leq X_{(2:n)} \leq \cdots \leq X_{(n:n)}$$

نمایش دهیم، سپس $X_{(r:n)}$ را r–امین آماره ترتیبی می نامیم.

در تعریف آماره های ترتیبی هیچ محدودیتی بر روی اینکه X_i ها مستقل یا هم توزیع باشند وجود ندارد. اما نتایج بسیار زیادی درباره آماره های ترتیبی تحت فرض کلاسیک که X_i ها iid باشند وجود دارد.

حال فرض کنید نمونه تصادفی ساده (SRS)[1] $X_{SRS} = \{X_1, X_2, ..., X_n\}$ از یک توزیع پیوسته با تابع توزیع تجمعی $F(x)$ و تابع چگالی $f(x)$ گرفته شود و $X_{OS} = \{X_{(1:n)} \leq X_{(2:n)} \leq \cdots \leq X_{(n:n)}\}$ نشان دهنده ی آماره های ترتیبی نمونه تصادفی ساده $(OSRS)$[2] باشد. فرض کنید $F_{(r:n)}(x)$ و $f_{(r:n)}(x)$ به ترتیب نشان دهنده ی تابع توزیع تجمعی و تابع چگالی احتمال r–امین آماره ترتیبی باشد و هم چنین $f_{(r,s:n)}(x, y)$ نشان دهنده تابع چگالی توام r–امین و s–امین آماره ترتیبی باشد. بنابراین $F_{(r:n)}(x)$ و

[1] Simple Random Sampling

[2] Ordered Simple Random Sampling

$f_{(r:n)}(x)$ و $f_{(r,s:n)}(x,y)$ به صورت زیر داده می شوند:

$$F_{(r:n)}(x) = \sum_{i=r}^{n} \binom{n}{i} [F(x)]^{i} [1-F(x)]^{n-i}, \quad (1-1)$$

$$f_{(r:n)}(x) = \frac{n!}{(r-1)!(n-r)!} [F(x)]^{r-1} [1-F(x)]^{n-r} f(x), \quad (2-1)$$

$$\begin{aligned} f_{(r,s:n)}(x,y) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} [F(x)]^{r-1} [F(y)-F(x)]^{s-r-1} \\ &\times [1-F(y)]^{n-s} f(x) f(y) \quad x<y. \end{aligned} \quad (3-1)$$

برای جزئیات بیشتر می توان به آرنولد و همکاران (۱۹۹۲) [1] و دیوید و ناگاراجا (۲۰۰۳) [2] مراجعه کرد.
آماره های ترتیبی به دست آمده از یک نمونه مجموعه رتبه دار مستقل و غیر هم توزیع هستند بنابراین ضروری است که بعضی جزئیات را درباره آماره های ترتیبی متناظر با متغیرهای $inid$ ارائه دهیم.
حال فرض کنید که $X_1, X_2, ..., X_n$ متغیرهای مستقل باشند که هر X_i دارای تابع توزیع تجمعی $F_i(x)$ و تابع چگالی $f_i(x)$ باشد و $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ نشان دهنده آماره های ترتیبی از این نمونه باشد سپس تابع توزیع تجمعی و تابع چگالی احتمال r–امین آماره ترتیبی $X_{r:n}$ بر اساس نتایج به دست آمده در دیوید و ناگاراجا (۲۰۰۳) به صورت زیر خواهد بود:

$$F_{r:n}(x) = \sum_{i=r}^{n} \sum_{S_i^{[n]}} \left\{ \prod_{l=1}^{i} F_{jl}(x) \prod_{l=i+1}^{n} [1-F_{jl}(x)] \right\}, \quad (4-1)$$

$$f_{r:n}(x) = \frac{1}{(r-1)!(n-r)!} \sum_{P^{[n]}} \left\{ \prod_{k=1}^{r-1} [F_{ik}(x)] f_{ir}(x) \prod_{k=r+1}^{n} [1-F_{ik}(x)] \right\}, \quad (5-1)$$

که در آن $\sum_{S_i^{[n]}}$ نشان دهنده ی مجموع بر روی همه ی جایگشت های $(j_1, j_2, ..., j_n)$ از $(1,2,...,n)$ است که در آن $j_1 \leq ... \leq j_i$ و $j_{i+1} \leq ... \leq j_n$ است و $\sum_{P^{[n]}}$ نشان دهنده مجموع بر روی همه ی $n!$ جایگشت از $(1,2,...,n)$ است. به علاوه تابع چگالی توام r–امین و s–امین آماره ترتیبی به صورت زیر داده شده است:

$$\begin{aligned} f_{r,s:n}(x,y) &= \frac{1}{(r-1)!(s-r-1)!(n-s)!} \sum_{P^{[n]}} \left\{ \prod_{k=1}^{r-1} [F_{ik}(x)] f_{ir}(x) \right. \\ &\times \left. \prod_{k=r+1}^{s-1} [F_{ik}(y)-F_{ik}(x)] f_{is}(y) \prod_{k=s+1}^{n} [1-F_{ik}(y)] \right\}, \ x<y. \end{aligned} \quad (6-1)$$

در دو لم بعدی دو ویژگی مهم را برای آماره های ترتیبی از یک نمونه iid بیان می کنیم که به وفور از آن ها استفاده خواهیم کرد:

[1] Arnold et al.

[2] Dived and Nagaraja

لم ۱-۱-۱. *برای هر $x \in \mathbf{R}$ داریم:*

$$\frac{۱}{n}\sum_{r=۱}^{n} f_{(r:n)}(x) = f(x) \qquad (۱-۷)$$

و آن را برابری اساسی می نامیم.

برهان. با استفاده از (۱ − ۲) برای هر $x \in \mathbf{R}$ داریم:

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

□

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

$$F_{(r:n)}(x) = B_{r,n-r+۱}(F(x)) \qquad (۱-۸)$$

که در آن $B_{a,b}(y)$ نمایانگر تابع توزیع تجمعی برای توزیع بتا با پارامترهای a و b است.

برهان. ابتدا از رابطه (۱ − ۲) داریم

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

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

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

بنابراین برهان کامل است. □

۲ - ۱ نمونه گیری مجموعه رتبه دار

مطالب این بخش از چن و همکاران (۲۰۰۴)[۱] اقتباس شده است.
یکی از قسمت های 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 سوال های مورد نظر به دست آورد. یکی از مکانیسم های متداول و معتبر برای جمع آوری داده ها، نمونه گیری تصادفی ساده (SRS) است. در این روش نمونه گیری، داده ها به صورت تصادفی از جامعه انتخاب می شوند.
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 گیری اندازه گیری کرد. برای وضعیت هایی که اندازه گیری واحدهای نمونه ممکن است سخت باشد، مثلا هزینه بردار، مخرب یا وقت گیر باشد اما رتبه بندی واحدها در یک مجموعه کوچک، آسان یا ارزان است می توان از نمونه گیری مجموعه رتبه دار(RSS)[۲] استفاده کرد.
فرض اساسی در نمونه گیری مجموعه رتبه دار این است که جامعه نامتناهی باشد. هم چنین فرض کنید که اگر یک مجموعه از داده ها از جامعه بیرون کشیده شود بتوان آن ها را با وسایل معینی بدون اندازه گیری دقیق رتبه بندی کرد.
فرم اصلی RSS که اولین بار توسط مک این تایر (۱۹۵۲) [۳]معرفی شد به صورت زیر است:
ابتدا یک نمونه تصادفی ساده به حجم n از جامعه بیرون کشیده می شود و این n واحد نمونه گیری با توجه به متغیر مورد مطالعه رتبه بندی می شوند این رتبه بندی بدون اندازه گیری دقیق و به صورت قضاوتی است. سپس یک واحد با رتبه ۱ برای اندازه گیری دقیق انتخاب می شود و بقیه واحدها کنار گذاشته می شوند. دوباره یک نمونه به اندازه n از جامعه انتخاب می شود و واحدهای نمونه به صورت قضاوتی رتبه بندی می شوند و واحد با رتبه ۲ برای اندازه گیری دقیق انتخاب می شود. این فرایند تا زمانی ادامه پیدا می کند که یک نمونه به اندازه n از جامعه بیرون کشیده شود و واحد با رتبه n برای اندازه گیری دقیق انتخاب شود. این فرایند یک سیکل نامیده می شود. اگر فرایند سیکل را m بار تکرار کنیم سپس یک

[۱] Chen et al

[۲] Ranked Set Sampling

[۳] McIntyre

نمونه مجموعه رتبه دار به اندازه $N = mn$ به صورت زیر در اختیار داریم:

$$\left\{X_{[r]i};\ r = ۱, ..., n;\ i = ۱, ..., m\right\}$$

اگر در مکانیسم رتبه بندی مشاهده ی دارای رتبه قضاوتی r–ام برابر با r–امین آماره ترتیبی باشد آنگاه رتبه بندی کامل است و نمونه مجموعه رتبه دار شامل آماره های ترتیبی مستقل است و ما از علامت $X_{(r)i}$ برای نشان دادن این نمونه استفاده می کنیم در غیر اینصورت رتبه بندی ناقص است و از علامت $X_{[r]i}$ برای نمایش نمونه استفاده می شود. روش ذکر شده نمونه گیری مجموعه رتبه دار متعادل نامیده می شود.

حالت کلی تر برای به دست آوردن یک نمونه مجموعه رتبه دار به حجم N به صورت زیر است:
ابتدا N مجموعه هر یک با اندازه k به طور تصادفی از جامعه تحت مطالعه بیرون کشیده می شود. سپس واحدهای هر مجموعه با یک مکانیسم معین به غیر از اندازه گیری دقیق رتبه بندی می شوند. سپس یک واحد و تنها یک واحد برای اندازه گیری دقیق از هر مجموعه انتخاب می شود و درنهایت N واحد اندازه گیری شده بر روی متغیر مورد مطالعه داریم. اگر فرض کنیم n_i تعداد واحدهای اندازه گیری شده با رتبه i باشد در اینصورت $N = \sum_{i=۱}^{k} n_i$. j–امین واحد اندازه گیری شده با رتبه i را با $X_{[i]j}$ نشان می دهیم و نمایش RSS به صورت زیر است:

$$\begin{array}{cccc} X_{[۱]۱} & X_{[۱]۲} & ... & X_{[۱]n_۱} \\ X_{[۲]۱} & X_{[۲]۲} & ... & X_{[۲]n_۲} \\ ... & ... & ... & ... \\ X_{[k]۱} & X_{[k]۲} & ... & X_{[k]n_k} \end{array}$$

که آن را نمونه مجموعه رتبه دار نامتعادل می نامیم. اگر $n_۱ = n_۲ = ... = n_k$ در این صورت RSS نامتعادل به RSS متعادل تبدیل می شود.
معمولا در شرایط زیر نمونه گیری مجموعه رتبه دار به کار می رود:

۱. رتبه بندی یک مجموعه از واحدها با توجه به متغیر مورد مطالعه بر اساس تجهیزات بصری یا وسایل معین نسبت به اندازه گیری دقیق به آسانی انجام شود.

۲. رتبه بندی با توجه به متغیر همراه انجام شود که منظور از متغیر همراه متغیری است که هدف اصلی نمونه گیری نیست اما وابستگی زیادی به متغیر مورد مطالعه دارد.

۱-۲-۱ مکانیسم رتبه بندی

در بخش ۱ − ۲ روش کلی RSS را معرفی کردیم. این روش یک طرح دو مرحله ای است. در اولین مرحله، یک نمونه تصادفی از جامعه گرفته می شود و یک مکانیسم معین برای رتبه بندی واحدها به کار گرفته می شود. در دومین مرحله، اندازه گیری دقیق متغیر مورد مطالعه بر روی واحدهای انتخاب شده بر اساس اطلاعات رتبه بندی در مرحله اول انجام می شود. این رتبه بندی قضاوتی بر اساس متغیر مورد مطالعه انجام می گیرد. مکانیسم های رتبه بندی دیگری غیر از این مکانیسم رتبه بندی وجود دارد که دراینجا به ذکر دو مورد از این مکانیسم ها می پردازیم.

فرض کنید با مکانیسم رتبه بندی اصلی مک این تایر (۱۹۵۲) شروع کنیم. اگر رتبه بندی کامل باشد یعنی رتبه واحدها با ترتیب عددی واحدها مطابقت داشته باشد در این صورت واحدهای اندازه گیری شده آماره های ترتیبی با تابع چگالی $f_{(r:n)}$ یعنی تابع چگالی r–امین آماره ترتیبی از توزیع F هستند.

یک مکانیسم رتبه بندی سازگار گفته می شود اگر رابطه اساسی زیر برقرار باشد:

$$F(x) = \frac{۱}{n}\sum_{r=۱}^{n} F_{[r:n]}(x) \qquad \forall x \in \mathbf{R} \tag{۱-۹}$$

بنابراین از لم ۱ − ۱ − ۱ واضح است که مکانیسم رتبه بندی کامل با توجه به متغیر مورد مطالعه سازگار است. در ادامه دو مکانیسم رتبه بندی سازگار را بیان می کنیم:

۱. مکانیسم رتبه بندی ناقص با توجه به متغیر مورد مطالعه: وقتی خطای رتبه بندی وجود دارد تابع چگالی آماره رتبه ای با رتبه r به صورت $f_{(r:n)}(x)$ نیست. در این حالت تابع توزیع تجمعی $F_{[r:n]}(x)$ را می توانیم بر اساس $F_{(s:n)}(x)$ به صورت زیر بیان می کنیم:

$$F_{[r:n]}(x) = \sum_{s=۱}^{n} p_{sr} F_{(s:n)}(x)$$

که در آن p_{sr} احتمال این است که واحدی که از نظر عددی دارای رتبه s–ام است به عنوان آماره ترتیبی قضاوتی r–ام انتخاب شود. زمانی که هر واحد در مجموعه باید به یک کلاس قضاوتی اختصاص داده شود برای هر $s = ۱, ..., k$ مجموع بر روی r باید مساوی با یک شود یعنی $\sum_{r=۱}^{n} p_{sr} = ۱$. هم چنین اگر هر مجموعه در فرایند انتخاب RSS به طور کامل به صورت قضاوتی رتبه بندی شود سپس هر آماره ترتیبی حقیقی با یک آماره ترتیبی قضاوتی جفت می شود بنابراین برای $r = ۱, ..., k$ مجموع بر روی s مساوی با یک می شود یعنی $\sum_{s=۱}^{n} p_{sr} = ۱$.

بنابراین :

$$\begin{aligned}\frac{۱}{n}\sum_{r=۱}^{n}F_{[r:n]}(x) &= \frac{۱}{n}\sum_{r=۱}^{n}\sum_{s=۱}^{n}p_{sr}F_{(s:n)}(x)\\ &= \frac{۱}{n}\sum_{s=۱}^{n}(\sum_{r=۱}^{n}p_{sr})F_{(s:n)}(x)\\ &= F(x)\end{aligned}$$

در حالت کلی ساختار یک RSS وابسته به مقادیر در نظر گرفته شده برای p_{sr} ها است. اگر برای همه ی r ها و s ها p_{sr} را برابر با $\frac{۱}{n}$ در نظر بگیریم سپس

$$F_{[r:n]}(x) = \frac{۱}{n}\sum_{s=۱}^{n}F_{(s:n)}(x) = F(x) \qquad r = ۱, ..., n$$

این حالت را رتبه بندی تصادفی گویند.

۲. رتبه بندی با توجه به متغیر همراه: موقعیت های عملی بسیاری وجود دارد که رتبه بندی متغیر مورد مطالعه X مشکل است اما یک متغیر همراه Y به آسانی می تواند رتبه بندی شود. بنابراین از یک متغیر همراه برای رتبه بندی واحدهای نمونه استفاده می کنیم. در این روش ابتدا n نمونه تصادفی ساده به اندازه n از جامعه انتخاب می کنیم سپس هر نمونه را نسبت به متغیر Y رتبه بندی می کنیم و در نمونه i–ام X متناظر با زوجی که Y دارای رتبه عددی i–ام است برای اندازه گیری مشخصه مورد نظر به عنوان عضو i–ام نمونه انتخاب می کنیم. برای دسترسی به اندازه نمونه بیشتر می توان این چرخه را m بار تکرار کرد. توجه کنید که Y ها مرتب شده اند ولی X ها مرتب شده نیستند زیرا رتبه بندی برحسب مقادیر Y بوده است. فرض کنید $Y_{(r)}$ نشان دهنده r–امین آماره ترتیبی از Y و $X_{[r]}$ نشان دهنده r–امین آماره ترتیبی قضاوتی متناظر با Y باشد. هم چنین فرض کنید $f_{X|Y_{(r)}}(x|y)$ تابع چگالی شرطی X به شرط $Y_{(r)} = y$ باشد و $g_{(r:n)}(y)$ تابع چگالی حاشیه ای $Y_{(r)}$ باشد. از آنجایی که $(X_i, Y_i), i = ۱, ..., n,$ متغیرهای تصادفی مستقل و هم توزیع هستند، تابع چگای شرطی $X_{[r]}$ به شرط $Y_{(r)}$ به صورت زیر است:

$$f_{X_{[r]}|Y_{(r)}}(x|y) = f_{X|Y}(x|y)$$

اکنون با استفاده از رابطه بالا داریم

$$\begin{aligned}f_{X_{[r]},Y_{(r)}}(x, y) &= f_{X_{[r]}|Y_{(r)}}(x|y)g_{(r:n)}(y)\\ &= f_{X|Y}(x|y)g_{(r:n)}(y)\end{aligned}$$

بنابراین

$$f_{[r:n]}(x) = \int f_{X|Y}(x|y) g_{(r:n)}(y) dy.$$

بنابراین به سادگی نتیجه می گیریم که:

$$\begin{aligned} f(x) &= \int \sum_{r=1}^{n} \frac{1}{n} f_{X|Y}(x|y) g_{(r:n)}(y) dy \\ &= \frac{1}{n} \sum_{r=1}^{n} f_{[r:n]}(x) \end{aligned}$$

اگر از یک متغیر همراه برای رتبه بندی استفاده کنیم در این صورت مقدار افزایش در دقت برآوردگر، بستگی به میزان وابستگی بین دو متغیر دارد. در این روش به طور طبیعی انتظار می رود که اگر ضریب وابستگی بین دو متغیر صفر باشد آن گاه دقت برآوردگر کاهش می یابد ولی هرچه ضریب وابستگی مثبت بیشتر شود، دقت برآوردگر نیز افزایش می یابد.

برای جزئیات بیشتر درباره ی متغیر همراه و مفاهیم مرتبط با آن می توان به یانگ (۱۹۷۷) [۱]مراجعه کرد.

۱-۲-۲ تاریخچه نمونه گیری مجموعه رتبه دار

ایده نمونه گیری مجموعه رتبه دار اولین بار توسط مک این تایر به منظور پیدا کردن یک روش با کارایی بیشتر برای برآورد بازده چراگاه به کار گرفته شد و به صورت تجربی به این نتیجه دست یافت که این روش نمونه گیری بهتر از نمونه تصادفی ساده است. او از

$$\hat{\mu}_{RSS} = \frac{1}{mn} \sum_{r=1}^{n} \sum_{i=1}^{m} X_{[r]i}$$

به عنوان یک برآوردگر برای μ استفاده کرد و نشان داد که $\hat{\mu}_{RSS}$ کارایی بیشتری نسبت به میانگین نمونه تصادفی ساده یعنی $\bar{X}$ دارد. مک این تایر (۱۹۵۲) مشاهده کرد که کارایی نسبی که به صورت نسبت واریانس میانگین نمونه مجموعه رتبه دار نسبت به واریانس میانگین نمونه تصادفی ساده در اندازه یکسان است، برای ۵ توزیع بزرگتر یا مساوی $\frac{n+1}{2}$ است. هرچند که او نتوانست یک نظریه ریاضی جالب برای آن ارئه دهد اما با این وجود، او را به عنوان اولین کسی که نمونه گیری مجموعه رتبه دار را معرفی کرد، می شناسند. هرچند که از آن زمان تا سال ها این روش نمونه گیری بدون پیشرفت باقی ماند اولین نتایج

[۱] Yang