بسم‌الله الرّحمن الرّحیم

برآورد پارامترمقیاس در خانواده‌های توزیع مکان_مقیاس متقارن توسط برآوردگرهای BLUE و OUAE برای نمونه‌های کامل و سانسور شده

بوسیله

مهدی خدادادی‌پور

پایان نامه

ارائه شده به دانشکده تحصیلات تکمیلی بعنوان بخشی

از فعالیت‌های تحصیلی لازم برای اخذ

درجه کارشناسی ارشد

دررشته

آمار ریاضی

از

دانشگاه شیراز

شیراز، ایران

ارزیابی وتصویب شده توسط کمیته‌پایان نامه بادرجه: بسیار خوب

امضاء اعضاء کمیته پایان نامه

دکتر ناهید سنجری فارسی‌پور، دانشیار بخش آمار دانشگاه شیراز (رئیس کمیته)

دکتر احمد رضا سلطانی استاد بخش آمار دانشگاه شیراز ....

دکتر فریبرز حیدری استادیار بخش آمار دانشگاه شیراز ....

آبان ماه ۱۳۷۸

تقدیم به:

پدر و مادر عزیزم و همسر مهربانم که همواره در
طول زندگی مدیون بزرگواریهایشان خواهم بود

سپاسگزاری

اینک که به یاری پروردگار متعال پایان‌نامه خود را به اتمام رسانده‌ام بر عهده من است که از زحمات و رهنمودهای همیشه راهنمایم، استاد ارجمند جناب خانم دکتر ناهید سنجری فارسی‌پور تشکر کنم که اگر نبود زحمات بیدریغ و بزرگوارانه ایشان بی‌شک اتمام این رساله میسر نمی‌گشت و شایسته است که از زحمات بی‌شائبه اعضای محترم کمیته پایان‌نامه جناب آقای دکتر احمد رضا سلطانی و همچنین از آقای دکتر فریبرز حیدری سپاسگزاری نمایم.

همچنین در پایان از پدر و مادر عزیزم که اولین معلم زندگی‌ام هستند، همهٔ آنانی که در تحصیل علم و دانش مرا مشوق بوده‌اند و راه را بر من هموار ساختند صمیمانه تشکر می‌کنم.

چکیده

برآورد پارامتر مقیاس در خانواده‌های توزیع مکان- مقیاس توسط برآوردگرهای (BLUE) و (OUAE) برای نمونه‌های کامل و سانسور شده

توسط

مهدی خدادادی‌پور

یک برآوردکننده جدیدی برای برآورد پارامتر مقیاس، توزیع‌های مکان - مقیاس متقارن (با فرض مشخص بودن پارامتر مکان) معرفی می‌گردد که یک برآوردکننده بهینه و بصورت یک ترکیب خطی از قدر مطلق آماره‌های ترتیبی می‌باشد. این برآوردکننده را برآوردکننده بهینهٔ نااریب خطی (OUAE) می‌نامیم. با محاسبه گشتاورهای قدر مطلق از آماره‌های ترتیبی، کارآئی نسبی این برآوردگر جدید را نسبت به بهترین برآوردکننده نااریب خطی (BLUE) برای توزیعهای مستطیلی - نرمال - وایبل دوگانه - نمائی دوگانه محاسبه می‌گردد. به طوری که نشان داده می‌شود که در بیشترین حالات (برای نمونه‌های مختلف) برآوردکننده جدید کاراتر از برآوردکننده BLUE می‌باشد.

فهرست مطالب

عنوان | صفحه

عنوان صفحه

عنوان	صفحه

فهرست جداول

عنوان | صفحه

عنوان صفحه

فصل اوّل

مقدمه و تاریخچه

هدف اصلی از ارائه مطالب مندرج در این پایان‌نامه، معرفی برآوردکننده جدیدی بنام برآوردکننده نااریب بهینه خطی؛ (OUAE) پارامتر مقیاس (Scale) توسط قدر مطلق آماره‌های ترتیبی در خانوادهٔ توزیع‌های مکان ـ مقیاس (Location - Scale) (با مشخص بودن پارامتر مکان) می‌باشد.

همچنین محاسبه آن برای یک نمونه تصادفی به اندازه $n = ۳(۱)۱۰$ برای توزیع‌های مستطیلی متقارن ـ نرمال ـ نمائی دوگانه (لاپلاس) ـ وایبل دوگانه.

هدف دیگر مورد نظر بررسی کارائی نسبی برآوردکننده OUAE نسبت به برآورد BLUE (بهترین برآوردکننده نااریب خطی توسط آماره‌های ترتیبی) پارامتر مقیاس می‌باشد که نشان داده می‌شود عموماً برای نمونه‌های تصادفی به اندازه‌ها مختلف در توزیع‌های ذکر شده برآوردکننده OUAE کاراتر از برآوردکننده BLUE است.

برای محاسبه هر دو برآورد، نیاز به محاسبهٔ گشتاور اوّل و دوّم و گشتاور حاصلضربی از قدر مطلق آماره‌های ترتیبی استاندارد شده و همچنین آماره‌های ترتیبی استاندار شده توزیع‌های ذکر شده می‌باشد که چگونگی روش محاسبه گشتاورها هدف دیگر ارائه مطالب می‌باشد.

بنابراین برای بررسی اهداف فوق؛ ارائه مطالب لازم در ۳ فصل، دنبال خواهد شد.

فصل دوّم: به معرفی برآوردکننده OUAE پارامتر مقیاس و برآوردکننده BLUE پارامترهای مکان ـ مقیاس ـ توزیع‌های مکان ـ مقیاس و روش محاسبه آن می‌پردازد.

فصل سوّم: برآوردکننده‌های OUAE و BLUE پارامتر مقیاس در توزیع‌ها مثلثی متقارن و نرمال (با مشخص بودن پارامتر مکان) برای نمونه‌ها تصادفی به اندازه $n = ۳(۱)۱۰$ محاسبه می‌گردد

و طبق جداولی، ضرایب برآوردها و واریانس آنها و کارائی نسبی OUAE نسبت به BLUE برای نمونه‌های فوق ارائه می‌گردد و نتایجی از این جداول بدست می‌آید.

فصل چهارم: همانند فصل سوّم برآوردکننده OUAE و BLUE پارامتر مقیاس برای توزیع‌های نمائی دوگانه و وایبل دوگانه برای نمونه‌های تصادفی به اندازه $n = ۳(۱)۱۰$ محاسبه می‌گردد و جداولی از ضرایب برآوردها و واریانس آنها و کارائی نسبی OUAE نسبت به BLUE بدست می‌آید که نتایجی قابل توجه از این جدول‌ها ارائه می‌گردد.

قبل از پرداختن به مطالب فصل‌های ۲ تا ۴ نیاز به یادآوری مطالبی می‌باشیم که در فصلهای مذکور مورد استفاده قرار می‌گیرند که به اختصار مطالب یادآوری می‌شوند.

۱_۱_ آماره‌های ترتیبی

فرض کنید $X_۱, \ldots, X_n$ متغیرهای تصادفی و پیوسته باشند که هر یک دارای تابع توزیع $F(x)$ و تابع چگالی $f(x)$ باشند، متغیرهای تصادفی $X_{۱:n}, \ldots, X_{n:n}$ را در نظر بگیریم با این فرض که این نمونه از جایگشت مجموعهٔ $X_۱, \ldots, X_n$ به طوری که به صورت صعودی مرتب شده‌اند، حاصل شده است این نمونه را، آماره‌های ترتیبی گویند که متغیرهایی وابسته می‌باشند.

مبحث آماره‌های ترتیبی، بیشتر در مورد خواص و کاربرد متغیرهای تصادفی مرتب شده و توابعی از آنها بحث و گفتگو می‌کند.

مثال ۱_۱: دستگاهی دارای n قسمت می‌باشد زمانهای خراب شدن این قسمتها عبارتند از $X_۱, \ldots, X_n$ که دارای فرضهای ارائه شده در قسمت فوق هستند، آنگاه X_k عبارتست از زمانی که طی آن قسمت k از کار می‌افتد، اگر دستگاهی به مجرد اینکه یکی از قسمتها از کار بیفتد متوقف گردد، آنگاه $X_{۱:n} = min(X_۱, \ldots, X_n)$ عبارت است زمان خراب شدن دستگاه. اگر دستگاه تا زمانی که nامین قسمت خراب شود متوقف نشود آنگاه $X_{n:n} = max(X_۱, \ldots, X_n)$ عبارت

است از زمان خراب شدن دستگاه.

قضیه: (۱ـ۱) فرض کنید $X_{۱}, \ldots, X_n$ یک نمونه تصادفی به اندازه n از یک تابع پیوسته با تابع چگالی $f(x)$ و توزیع $F(x)$ باشد، بنابراین تابع چگالی rامین آماره ترتیبی $X_{r:n}$ بصورت

$$f_r(x) = D_{r,n}\,[f(x)]^{r-۱}\,[۱ - f(x)]^{n-r}\,f(x)$$

است و تابع چگالی توأم rامین، sامین آماره ترتیبی بصورت

$$\begin{aligned} f_{r,s}(x,y) &= D_{r,s,n} f(x) f(y) F(x)^{r-۱}\,[F(y) - F(x)]^{s-r-۱} \\ &\times [۱ - f(y)]^{n-s} \qquad ; x \leq y \end{aligned}$$

می‌باشد، همچنین تابع چگالی توأم n آماره ترتیبی بصورت

$$\begin{aligned} f_{۱,۲,\ldots,n}(x_{۱}, \ldots, x_n) &= n!\pi_{i=۱}^{n} f(x_i) \\ & -\infty < x_{۱} < x_{۲} < \ldots < x_{۲} < +\infty \end{aligned}$$

می‌باشد.

$$\begin{aligned} D_{r,n} &= \frac{n!}{(r-۱)!(n-r)!} \\ D_{r,s,n} &= \frac{n!}{(r-۱)!(s-n-r)!(n-s)!} \end{aligned}$$

اثبات: [۲]

n قطعه را در نظر می‌گیریم، برای آزمون طول عمر، قطعات را مورد آزمایش قرار می‌دهیم تا n قطعه از کار بیافتد یا در اصطلاح شکست مشاهده شود که زمان شکست هر قطعه، زمان طول عمر قطعه می‌باشد، بعنوان مثال در یک کاربرد نمونه‌ای پزشکی یا طول عمر ممکن است شروع یک بیماری مزمن تا لحظه مرگ باشد یا در قابلیت اعتماد صنعتی زمان شکست به عنوان طول عمر قطعات دستگاهها در نظر گرفته می‌شود.

فرض کنید طول عمر هر قطعه متغیر تصادفی با تابع توزیع پیوسته $F(x;\theta)$ و تابع چگالی $f(x;\theta)$ باشند، که θ پارامتر مجهول می‌باشد، (طول عمر قطعات از هم مستقل می‌باشند) در نتیجه با توجه به مطالب فوق یک نمونه تصادفی مرتب شده $(X_{1:n},\dots,X_{n:n})$ خواهیم داشت.

حالتی را در نظر بگیرید که ما مجبور شویم بنا به دلایلی آزمایش را قطع کنیم قبل از اینکه همهٔ قطعات از کار بیافتد، در نتیجه زمان شکست یا طول عمر بعضی از قطعات در دسترس ما نیست، پس در این حالت نمونه تصادفی مرتب شده را سانسور شده گویند در اینجا به تعریف دو نوع از نمونه‌های سانسور شده می‌پردازیم.

۱_۱_۱ سانسور نوع I (سانسور زمانی)

فرض کنید که تصمیم بگیریم که آزمایش فوق را در یک زمان مشخص t متوقف کنیم، که در این صورت زمانهای شکست قطعات، شکست خورده قبل از زمان t ثبت می‌شود، نمونه بدست آمده را در این حالت سانسور نوع I یا سانسور زمانی گویند که سانسور راست نیز می‌نامند زیرا داده‌های سمت راست نمونه مرتب شده (داده‌ها بزرگ) سانسور شده است. واضح است که در این حالت تعداد مشاهدات آماره‌های ترتیبی یک متغیر تصادفی است. هر گاه متغیر تصادفی را R تعریف کنیم، R مقادیر زیر را می‌تواند بگیرد

$$R = ۰, ۱, \dots, n$$

که تابع درستنمائی آن بصورت

$$L(\theta|r,\underline{x}) = \begin{cases} \frac{n!}{(n-r)!}\pi_{i=۱}^{r} f(x_i)\,[۱-F(t,\theta)]^{n-r} & ۰ < r_۱ < n \\ & x_۱ < x_۲ < \dots < x_r < t \\ \{۱-F(t,\theta)\}^n & r = ۰ \quad t < x_۱ \end{cases} \tag{۱.۱}$$

می‌باشد. ($\underline{x}$ یک بردار از مشاهدات آماره‌ها ترتیبی سانسور شده است)

اثبات [۲]:

۲ـ۱ـ۱ـ سانسور نوع II (سانسور شکستی)

فرض کنید که بنا به دلایلی تصمیم بگیریم که آزمایش را وقتیکه rامین شکست رخ داد متوقف کنیم، در این حالت نمونه مرتب شده $(X_{1:n}, \ldots, X_{r:n})$ را نمونه سانسور شده نوع II یا سانسور شکستی نامند. در این حالت $X_{r:n}$ زمان شکست rامین قطعه، متغیر تصادفی می‌باشد با تابع درستنمایی به صورت

$$L(\theta|r, \underline{x}) = \frac{n!}{(n-r)!} \pi_{i=1}^{r} f(x_i) \left[1 - F(x_r, \theta)\right]^{n-r} \qquad x_1 < \ldots < x_r \quad (1.2)$$

اثبات[۲]:

همانطور که در سانسور نوع I اشاره گردید چون داده‌های بزرگ سانسور شده است در این حالت نمونه را نمونه سانسور راست نیز نامیده می‌گردد. همچنین در حالتی که داده‌ها کوچک سانسور شوند نمونه $(X_{r+1:n}, \ldots, X_{n:n})$ را سانسور چپ گویند. حتی داده‌ها می‌تواند از طرف چپ و راست سانسور شوند که نمونه $(X_{r_1+1:n} \quad \ldots, X_{n-r_2:n})$ را سانسور دو طرفه گویند که r_1 تعداد مشاهدات سانسور شده طرف چپ و r_2 تعداد مشاهدات سانسور شده طرف راست را نشان می‌دهد. در صورتیکه $r_1 = r_2$ باشد نمونه را نمونهٔ سانسور شده متقارن می‌نامند و هر گاه $r_1 = r_2 = 0$ باشد نمونه تصادفی را نمونه کامل گویند.

لازم به ذکر است به طور کلی هر زیر مجموعه از آماره‌های ترتیبی $(X_{1:n}, \ldots, X_{n:n})$ برای هر توزیعی، را سانسور نوع II نامند.

۲ـ۱ـ مدل‌های خطی و برآوردهای حداقل مربعات

در این بخش مختصراً در مورد مدل‌های خطی و برآوردهای حداقل مربعات (Least -

(Squares و نتایج حاصل از آن به بحث می‌پردازد که در فصل بعد مورد استفاده قرار می‌گیرد.

پرداختن به مسائل در مورد مدل‌های خطی، بدون استفاده از ماتریس‌ها امکان‌پذیر نمی‌باشد.

اینکه به معرفی چند نماد و تعریف می‌پردازیم. ماتریس $A = (a_{ij})$ را در نظر بگیریم که درایه ijام آن a_{ij} می‌باشد، ترانهادهٔ ماتریس A را $A' = (a_{ji})$ تعریف می‌کنیم که ماتریسی است که جای سطر و ستونش عوض شده باشد، همچنین هر گاه A یک ماتریس مربع و معکوس‌پذیر باشد، معکوس ماتریس A را با نماد A^{-1} نشان می‌دهیم.

ماتریس A را ماتریس متقارن گوئیم هر گاه $A = A'$ باشد. هر گاه ماتریس $n \times n$، A متقارن و برای

$$\forall y \neq \underline{o} \in R^n \quad , \quad y'Ay > 0$$

ماتریس A را ماتریس معین مثبت (positive definit) گویند که چنین ماتریس معکوس‌پذیر می‌باشد.

۱_۲_۱_ تعریف مدل‌های خطی:

متغیرهای تصادفی وجود دارند که مشاهدات آنها در یک آزمایش اغلب وابسته به یک یا چند متغیر دیگر می‌باشد. بعنوان مثال در یک آزمایش شیمیایی، متغیر سرعت واکنش وابسته به متغیرهای درجه حرارت و زمان واکنش می‌باشد.

متغیر تصادفی Y_i را در نظر می‌گیریم بطوریکه مشاهده، y_i وابسته به مقادیر مشاهده شده

$$a_{i0}, a_{i1}, \ldots, a_{ip} \quad i = 1, \ldots, n \qquad n \geq p + 1$$

باشد با فرض اینکه

$$E(Y_i) \;=\; \sum_{k=0}^{p} a_{ik} B_k \qquad (1.3)$$