بسم الله الرحمن الرحیم

دانشکده علوم

پایان نامه کارشناسی ارشد در رشته آمار ریاضی

بررسی مجموعه های اطمینان همزمان برای نسبت ترکیبات خطی از پارامترها درتوزیعهای چند متغیره و یک متغیره

توسط:

مریم طالبی مقدم

استاد راهنما :

دکتر عبدالرسول برهانی حقیقی

شهریور ۸۸ — ۸۷

به نام خدا

بررسی مجموعه های اطمینان همزمان برای نسبت ترکیبات خطی از پارامترها درتوزیعهای چند متغیره و یک متغیره

به وسیله‌ی:

مریم طالبی مقدم

پایان‌نامه

ارائه شده به تحصیلات تکمیلی دانشگاه به عنوان بخشی

از فعالیت‌های لازم برای اخذ درجه کارشناسی ارشد

در رشته‌ی:

آمار ریاضی

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی شده توسط کمیته پایان‌نامه با درجه: عالی

دکتر عبدالرسول برهانی حقیقی، استادیار بخش آمار(رئیس کمیته)..

دکتر مینا توحیدی، استادیار بخش آمار..

دکتر محمود خراتی کوپائی، استادیار بخش امار..

شهریورماه ۱۳۸۷

تقدیم به پدر ومادر مهربانم

آنان که همچون کوهی استوار در کنارم حضور دارند تا از تندباد حوادث نهراسم و در جاده پرپیچ وخم زندگی به پیش روم. آنان که نگاه پر محبتشان شوق زیستن را به من اعطا می کند.

سپاسگزاری

سپاسگزارم از تمام کسانی که کشتی سرگردانی مرا ناخدا شدند تا از طوفان های هراس انگیز جهالت به سوی ساحل زیبای علم رهنمون شوم و با صبر و حمایت خویش مرا به ادامه این راه امیدوار ساختند. من در این برهه از زمان، نهایت سپاس خود را از استاد راهنمایم ، آقای دکتر عبدالرسول برهانی حقیقی و اساتید مشاورم خانم دکتر مینا توحیدی و آقای دکتر محمود خراتی کوپایی و همچنین از آقای علی اکبرجعفری ابراز می‌کنم که در نهایت صداقت و بزرگواری گامهای سست مرا از آغاز تا کنون یاری نموده اند . به آن امید که شاید من نیز روزی یاریگر اندیشه های روشن آیندگان باشم.

چکیده

بررسی مجموعه های اطمینان همزمان برای نسبت ترکیبات خطی از پارامترها درتوزیعهای چند متغیره و یک متغیره

به وسیله ی:

مریم طالبی مقدم

در این پایان نامه مجموعه و فواصل اطمینان همزمان برای دو تابع برآورد شدنی و نسبت آن دو تحت یک مدل خطی و مجموعه و فواصل اطمینان همزمان برای نسبت ترکیبات خطی از پارامترها درتوزیعهای چند متغیره بررسی می شود. این پایان نامه روشهایی برای بدست آوردن فواصل اطمینان همزمان و روشهای جدیدی برای بدست آوردن مجموعه‌های اطمینان همزمان بر پایه t چند متغیره ارائه می گردد که سطح اطمینان مورد نظر را بطور دقیق کنترل می کند. بر این اساس این پایان نامه شامل پنج فصل می‌باشد. فصل اول مقدمه و پیشینه تحقیق می‌باشد. در فصل دوم فواصل اطمینان همزمان برای نسبت های تکی بیان می شود. مجموعه های اطمینان همزمان دقیق برای چندین نسبت در آنالیز واریانس یک‌طرفه در فصل سوم توضیح داده می‌شود. فواصل اطمینان همزمان تقریبی برای چندین نسبت در فصل چهارم بیان می شود. کاربرد روشهای ذکر شده و شبیه سازی و نتایج در فصل پنجم و ششم بیان می شود.

فهرست مطالب

عنوان صفحه

فصل اول: مقدمه و پیشینه تحقیق ۱

فصل دوم: فواصل اطمینان همزمان برای نسبتهای تکی ۴

۲-۱-مقدمه ۵

۲-۲-تعاریف و اصطلاحات مورد نیاز ۵

۲-۳-روش فیلر ۶

۲-۳-۱-روش فیلر در مورد مجموعه اطمینان برای نسبت میانگینهای نرمال ۷

۲-۳-۲-روش فیلر برای فاصله اطمینان همزمان برای نسبتهای دو تابع خطی ۸

۲-۴-روش S شفه ۹

۲-۵-روش بونفرونی ۱۱

۲-۶-فواصل اطمینان همزمان بر اساس توزیع تا چند متغیره ۱۲

۲-۶-۱-روش Plug-in (جایگزینی) ۱۲

۲-۶-۲-روش بدترین حالت ۱۳

۲-۶-۳-مقایسه روشها ۱۶

فصل سوم: مجموعه اطمینان همزمان برای نیستهای چند متغیره در آنالیز واریانس یک طرفه ۱۸

۳-۱-مقدمه ۱۹

۳-۲-تعاریف و اصطلاحات مورد نیاز ۱۹

۳-۳-مجموعه‌های اطمینان همزمان ۲۳

۳-۴-روش تکراری ۲۴

۳-۴-۱-الگوریتم روش تکراری ۲۵

۳-۵-روش تست نقطه به نقطه.. ۲۷

۳-۶-مقایسه روشها.. ۲۸

فصل چهارم : فواصل اطمینان همزمان برای نیست‌های چند متغیره................................ ۲۸

۴-۱-مقدمه.. ۲۸

۴-۲-روش بونفرونی.. ۲۸

۴-۳-روش شفه.. ۲۹

۴-۴-روش $Mt_r(v, I_r)$.. ۳۰

۴-۵-روش جایگزینی (Plug-in).. ۳۱

۴-۶-روش باز نمونه گیری (Resampling).. ۳۲

۴-۶-۱-روش Tmax.. ۳۲

۶-۴-۲-روش SCI متعادل.. ۳۳

۴-۷-نسبت ترکیب های خطی پارامترها در یک مدل خطی عام.. ۳۵

۴-۸-مقایسات چند به یک.. ۳۷

فصل پنجم: کاربرد روشها و نتیجه گیری.. ۳۹

۵-۱-کاربرد اول.. ۴۰

۵-۲-کاربرد دوم.. ۴۲

۵-۳-کاربرد سوم.. ۴۴

۵-۴-کاربرد چهارم.. ۴۴

۵-۵-کاربرد پنجم.. ۴۵

۵-۶-نتیجه‌گیری.. ۴۶

فصل ششم: شبیه‌سازی.. ۴۸

۶-۱-مقدمه.. ۴۹

۶-۲-شبیه سازی.. ۴۹

۶-۳-نتیجه‌گیری.. ۵۰

ضمائم و پیوستها.. ۵۰

پیوست اول.. ۵۲

پیوست ا لف: قضیه و نامساوی سیداک..۵۲
پیوست ب: قضیه و نامساوی اسلیپین..۵۳
پیوست دوم:..۵۴
پیوست الف) برنامه کامپیوتری مثالها..۵۴
پیوست ب) برنامه کامپیوتری شبیه سازی..۶۰
واژه‌نامه..۷۵

واژه‌نامه انگلیسی به فارسی...۷۶
واژه‌نامه فارسی به انگلیسی...۷۹
فهرست منابع...۸۳

فهرست جدول‌ها

عنوان و شماره .... صفحه

جدول شماره ۱: مقدار بحرانی برای فاصله اطمینان همزمان مختلف........ ۱۶

جدول شماره ۲: ارزش ژنوتیپ یک گیاه........ ۴۱

جدول شماره ۳: انواع مختلف غالب بودن........ ۴۲

جدول شماره ۴: وزن ژنوتیپ ها........ ۴۲

جدول شماره ۵: فاصله اطمینان همزمان برای پارامترهادر کاربرد دوم........ ۴۳

جدول شماره ۶: فواصل اطمینان همزمان دو طرفه با ضریب اطمینان ۹۵٪ برای γ_2 و γ_1 .... ۴۴

جدول شماره ۷: احتمال پوشش در حالتی که ۲ = τ_1 و ۳ = τ_2 .... ۴۶

جدول شماره۸: احتمال پوشش در حالتی که ۰/۸ = τ_1 و ۰/۴ = τ_2 .... ۴۷

فهرست شکل‌ها

عنوان	صفحه
شکل شماره ۱- ناحیه اطمینان به روش S شفه	۹
شکل شماره ۲- نا حیه انتگرال گیری برای بدست آوردن روش بدترین حالت	۱۴
شکل شماره ۳- مجموعه اطمینان وفواصل اطمینان همزمان مثال سوم	۴۴
شکل شماره ۴- مجموعه اطمینان و فواصل اطمینان همزمان مثال چهارم	۴۵
شکل شماره ۵- مجموعه اطمینان و فواصل اطمینان همزمان یک طرفه مثال‌پنجم	۴۶

فصل اول:

مقدمه و پشینه تحقیق

فصل اول: مقدمه و پیشینه تحقیق

در بسیاری از کاربردهای آماری علاقمند به مقایسه دو میانگین از دو جامعه از طریق بررسی نسبت میانگین‌های آن دو جامعه می‌باشیم. هاشک و دیگران (Hauschke et al, ۱۹۹۹) در مطالعه‌ای برابری تاثیر دو دارو را بر اساس نسبت میانگین تاثیر دو دارو در دو گروه آزمایشی ارزیابی و بررسی کردند. مساله مشابه دیگری که در ژنتیک کمی اتفاق می‌افتد عبارت است از برآورد کردن نسبت غالب بودن از طریق اثرات افزایشی ژن و غالب بودن آن می‌باشد که این موضوع توسط بویتکس و دیگران(Boitex et al, ۲۰۰۴) مطالعه و بررسی شده است. همچنین یکی از کاربردهای بسیار مهم آن در علم پزشکی می‌باشد. برای پزشکان نسبت تفاضل میانگین دو روش درمان به تفاضل میانگین اثر بخشی دو روش درمان مهم می‌باشد. این علم تحت عنوان ICER (نسبتهای هزینه اثر بخشی) نام دارد که این موضوع توسط پولسکی و دیگران (Polesky et al, ۱۹۹۷) و بریگز ودیگران (Briggs et al, ۱۹۹۹) روشهای مختلف ساخت فاصله اطمینان برای ICER را مقایسه کردند. همچنین خلاصه‌ای از کاربردهای فواصل اطمینان همزمان برای نسبت پارامترها را می‌توان در بسیاری از مقالات مالی (Malley, ۱۹۸۲) و سایکز (Sykes, ۲۰۰۰) یافت. اگر که میانگینها با اثرات بر طبق یک مدل خطی برآورد شوند از روش فیلر (Filler, ۱۹۵۴) می‌توان استفاده نمود و یک فاصله اطمینان برای نسبت میانگینها ارائه داد. کندال (Kendal , ۱۹۷۹) روش فیلر را در مورد مجموعه‌های اطمینان برای نسبت میانگینهای نرمال ارائه داد.

در اغلب مواقع در استنباط آماری محقیقن علاقمند به برآورد فاصله‌ای همزمان نسبت ترکیبات چندین پارامتر از چند جامعه می‌باشند. کاکس (Cox, ۱۹۷۶) و کوشات (Koschat , ۱۹۸۷) روش فیلر را به بیش از دو جامعه تعمیم دادند. شفه (Scheffe , ۱۹۷۰) نشان داد که برآورد نسبت ها در ارتباط با برآورد یک بردار بی جهت می‌باشد و به موجب آن روشی برای برآورد مشترک نسبت‌های چند متغیره بر پایه روش تصویر مطرح کرد.

بعد از شفه، یانگ و همکاران (young et al, ۱۹۹۷) روش شفه را در مورد نسبت پارامترهای خطی و مدل‌های مختلط غیر خطی توسعه دادند. بنت (Bennett, ۱۹۹۱) مسئله رسم کردن فواصل اطمینان برای نسبت میانگین‌ها به طور همزمان برای توزیع نرمال دو متغیره را مورد بررسی قرار داد. مالی (Malley, ۱۹۸۲) ساخت فاصله‌های اطمینان همزمان در حالت چند متغیره برای چندین نسبت با استفاده از ماتریس واریانس کوواریانس ارائه نمود. هونگ و لی‌یو

(Hwang and liu, ۱۹۹۰) دنباله‌ای از فواصل اطمینان برای چندین نسبت ارائه دادندکه شبیه فاصله اطمینان به روش شفه می‌باشد. روشهای مختلف فواصل اطمینان برای نسبت پارامترها به طور همزمان در یک مدل خطی توسط پایفو و امریچ (Piepho and Emrich, ۲۰۰۵) ارائه شد. روش‌های جدیدی برای رسم کردن مجموعه های اطمینان همزمان برای نسبت میانگین ها توسط دیلبا و همکارانش (Dilba et.al , ۲۰۰۶) ارائه دادند که دقیقاً سطح اطمینان را کنترل می‌کند. همچنین روشهای مختلف فواصل اطمینان همزمان برای نسبت پارامترها معرفی کردند.

هدف ما در این پایان نامه معرفی فواصل اطمینان همزمان و مجموعه‌های اطمینان همزمان برای نسبت پارامترها می‌باشد. در فصل دوم به بررسی فواصل اطمینان همزمان تحت یک مدل خطی برای نسبتهای تکی می‌پردازیم. فصل سوم روشهای بدست آوردن مجموعه های اطمینان همزمان دقیق برای چندین نسبت در آنالیز واریانس یک طرفه بیان می‌کنیم. در فصل چهارم فواصل اطمینان همزمان نسبتهای چند متغیره را معرفی می‌کنیم. کاربرد این روشهاو شبیه سازی و نتیجه گیری در فصل پنجم بیان می‌کنیم.

فصل دوم:

فواصل اطمینان همزمان برای نسبت های تکی

فصل‌دوم:فواصل اطمینان همزمان برای نسبتهای تکی

۲-۱-مقدمه

دراین فصل به بررسی فواصل اطمینان همزمان برای دو تابع برآورد شدنی ونسبت آن دو پرداخته می شود. روشهای فیلر، S -شفه ، بونفرونی و t چند متغیره مطرح می شود. دو روش جدید که بر پایه توزیع tچند متغیره می‌باشد، معرفی می‌شود.

۲-۲- تعاریف و اصطلاحات مورد نیاز

مدل خطی در نظر گرفته شده به صورت $y = X\beta + e$ می‌باشد که در آن X نشان دهنده ماتریس طرح، β نشان دهنده بردار پارامتر وخطا، $e \sim MVN(0,\sigma^2 I)$ می‌باشد. که‌ماعلاقمند به یافتن فواصل اطمینان همزمان برای دو تابع برآورد شدنی، $\gamma_1 = C_1'\beta$ و $\gamma_2 = C_2'\beta$ و نسبت آنها $\theta = \dfrac{\gamma_1}{\gamma_2}$ تحت مدل خطی هستیم به طوری که $C = (C_1, C_2)$ یک بردار می باشد. از کمترین توان دوم برای برآورد کردن$\gamma = (\gamma_1, \gamma_2)$ استفاده می کنیم که

$$\hat{\gamma} = C'(X'X)^{-} X'y \qquad (۲-۲-۱)$$

وبرآورد واریانس $\hat{\gamma}$ برابراست:

$$S = C'(X'X)^{-} C s^2 \qquad (۲-۲-۲)$$

به طوری که

$$s^2 = \frac{y'\left[I - X(X'X)^{-}X'\right]y}{n - rank X} \qquad (۲-۲-۳)$$

و n تعداد مشاهدات در y می باشد.

هدف بدست آوردن فواصل مشترک برای $\gamma_1 = K_1'\gamma$ ،$\gamma_2 = K_2'\gamma$ که $K_1 = (1,0)'$ و $K_2 = (0,1)'$ و $\theta = \frac{\gamma_1}{\gamma_2}$ می‌باشد که θ را می توان به صورت $\gamma_1 - \theta\gamma_2 = K_3'\gamma = 0$ نوشت که $K_3' = (1,-\theta)$ می باشد. برای بدست آوردن فاصله اطمینان برای θ به یک معادله درجه دوم به صورت $A\theta^2 + B\theta + c \le 0$ خواهیم رسید که با تعیین علامت کردن حدود θرا بدست خواهیم آورد. لازم به ذکر است که اگر $A < 0$ و $\Delta < 0$ معادله درجه دوم همواره منفی خواهد بود و تمام خط حقیقی فاصله تعریف شده برای θ می باشد. در این حالت، داده ها هیچ اطلاعاتی در مورد θ نخواهند داشت. اگر $\Delta < 0$ و $A > 0$ این معادله همواره مثبت می شود که این حالت نمی تواند اتفاق بیافتد زرب (Zerb, ۱۹۷۸). در کل پایان نامه منظور از SCI ،فاصله اطمینان همزمان و منظور از SCS ،مجموعه اطمینان همزمان می‌باشد.

۲-۳- روش فیلر

برای بدست آوردن فاصله اطمینان برای θ از فرم کلی زیر استفاده می کنیم.

$$\frac{(K_3'\hat{\gamma})^2}{K_3'SK_3'} \le m^2 \qquad (۲-۳-۱)$$

برا ی بدست آوردن m کافیست که توزیع $\frac{(K_3'\hat{\gamma})^2}{K_3'SK_3}$ را پیدا کنیم. $K_3'\hat{\gamma}$ دارای توزیع نرمال با میانگین صفر و $\mathrm{var}(K_3'\hat{\gamma})$ می‌باشدو $\frac{(n - rank(X))(K_3'SK_3')}{\mathrm{var}(K_3'\hat{\gamma})}$ دارای توزیع χ^2 با $v = n - rank(X)$ درجه آزادی است. در نتیجه

$$\frac{(K_3'\hat{\gamma})^2}{K_3'SK_3'} \sim F_{(1:v:1-\alpha)} \qquad (۲-۳-۲)$$

براساس روش فیلر فاصله اطمینان ٪ $(1 - \alpha)$ ۱۰۰ برای θ برابر است با:

$$\left[\theta : \frac{(K_3'\hat{\gamma})^2}{K_3'SK_3'} \leq F_{1;\upsilon;1-\alpha}\right] \qquad (۲-۳-۳)$$

ویک فاصله اطمینان برای θ براساس توزیع t برابر است با:

$$\left[\theta : \frac{\left|K_3'\hat{\gamma}\right|}{\sqrt{K_3'SK_3'}} \leq t_{\upsilon;1-\alpha/2}\right] \qquad (۲-۳-۴)$$

ایـن فاصـله اطمینــان بـرای γ_2 ،γ_1 بـا قـرار دادن K_1, K_2 بجـای K_3 و جـایگزین کـردن $\hat{\gamma}$ بوسیله $(\hat{\gamma}-\gamma)$ درنامعادلات بالا بدست می‌آید.

۲-۴-روش S شفه

برای بدست آوردن فاصله اطمینان برای γ_1و γ_2 و θ به روش شفه نیاز به قضیه زیر داریم.

قضیه ۱ـ اگر L ماتریس معین مثبت باشد آنگاه برای هر h داریم:

$$\max_{h:h\neq 0}\left\{\frac{(h'b)^2}{h'Lh}\right\} = b'L^{-1}b$$

اثبات:برای هر بردار uو v و عدد ثابت a داریم

$$0 < \left\|v - au\right\|^2 = a^2\left\|u\right\|^2 - 2au'v + \left\|v\right\|^2 = \left(a\left\|u\right\| - \frac{u'v}{\left\|u\right\|}\right)^2 + \left\|v\right\|^2 - \frac{(u'v)^2}{\left\|u\right\|^2}$$

$u \neq 0$ در نظر می‌گیریم . با قرار دادن $a = \dfrac{u'v}{\left\|u\right\|^2}$ از نامساوی کوشی شوارتز داریم :

$$\left\| v-\frac{u'v}{\|u\|^2}u \right\| = \|v\|^2 - \frac{(u'v)^2}{\|u\|^2} \geq 0$$

بنابراین $(u'v)^2 \leq \|v\|^2\|u\|^2$ و داریم

$$\max_{v:v\neq 0}\left\{\frac{(u'v)^2}{v'v}\right\} \leq u'u = \|u\|^2$$

از طرفی برای $v=u$ تساوی برقرار می‌شود پس $\max_{v:v\neq 0}\left\{\frac{(u'v)^2}{v'v}\right\} = u'u$ حال چون L ماترس معین مثبت است یک ماتریس R معکوس پذیر وجود دارد بطوریکه $L = RR'$. بنابراین با قرار دادن $v = R'h$ و $u = R^{-1}b$ داریم

$$\max_{h}\left\{\frac{(h'b)^2}{h'Lh}\right\} = b'L^{-1}b$$

■

برای بدست آوردن فاصله اطمینان از فرم کلی زیر استفاده می کنیم.

$$\frac{|K'(\hat{\gamma}-\gamma)|}{\sqrt{K'SK}} \leq m \qquad (۱-۴-۲)$$

برای هر $K \neq 0$. اگر $K = (0,1)'$ فاصله اطمینان برای γ_1 داریم. اگر $K = (1,0)'$ فاصله اطمینان برای γ_2 بدست می‌آید و برای $K = (-1,\theta)'$ فاصله اطمینان برای θ. بنابراین باید m را طوری تعیین کرد که بطور همزمان فواصل اطمینان برای هر سه پارامتر دارای ضریب اطمینان $1-\alpha$ باشد. برای این منظور از روش زیر استفاده می کنیم.

$$P_r\left[\frac{|K'(\hat{\gamma}-\gamma)|}{k'Sk} \leq m \quad \forall K\right] = P_r\left[\frac{(K'\hat{\gamma}-K'\gamma)^2}{K'SK'} \leq m^2 \quad \forall K\right] =$$
$$P_r\left[\max_K \frac{(K'\hat{\gamma}-K'\gamma)^2}{K'SK'} \leq m^2 \quad \forall K\right]$$

بنابراین طبق قضیه۱ ، احتمال بالابرابر است با

$$= P_r\left[(\hat{\gamma}-\gamma)' S^{-1}(\hat{\gamma}-\gamma) \le m^2 \qquad \forall K\right] \tag{۲-۴-۲}$$

بنابراین با استفاده از آماره $T^2 - Hoteling$ داریم:

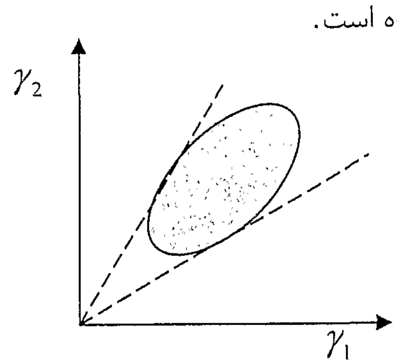
$$(\hat{\gamma}-\gamma) S^{-1}(\hat{\gamma}-\gamma) \sim 2F(2,\upsilon) \tag{۳-۴-۲}$$

پس فاصله به صورت زیر خواهد بود

$$\frac{\left(K'(\hat{\gamma}-\gamma)\right)}{K'SK} \le 2F_\alpha(2,\upsilon) \qquad \forall K \tag{۴-۴-۲}$$

این تذکر الزامی است چون این فواصل برای هر K ساخته شده اند بنابراین برای بعضی از K های خاص(مثل $K=(1,0)$ و $K=(0,1)$ و $K=(-1,\theta)$) فواصل فوق محافظه کارانه خواهدبود.

ناحیه اطمینان مشترک برای (γ_1,γ_2) یک بیضی می شود.اگر بیضیِ اطمینان مبدا را تحت پوشش قرار ندهد ناحیه اطمینان برای θ از طریق دو تانژانت (مماس) که از مبدا عبور می کندمشخص می شود. این نواحی در شکل(۱) نشان داده شده است.



شکل شماره ۱- ناحیه اطمینان به روش S شفه