بسم الله الرحمن الرحيم

دانشکده علوم ریاضی

گروه آمار

پایان‌نامه

برای دریافت درجه کارشناسی ارشد در رشته
آمار، گرایش آمار ریاضی

عنوان

بررسی رفتار برآوردگر جک‌نایف ریج

استاد راهنما

دکتر محمد آرشی

دانشجو

مریم برزوئی بیدگلی

۱۳۹۳

پیر ما گفت خطا بر قلم صنع نرفت

آفرین بر نظر پاک خطا پوشش باد

«حافظ»

به

پدر و مادرم

و روح پاک الهه

سپاس گزاری...

سپاس بی‌کران پروردگار یکتا را که هستی‌مان بخشید، و به طریق علم و دانش رهنمون ساخت، و به همنشینی رهروان علم و معرفت، مفتخرمان نمود، و خوشه‌چینی از علم و آگاهی را روزی‌مان نهاد.

با سپاس از جناب آقای دکتر محمد آرشی، استاد راهنمای ارجمند که با ارائه رهنمودها، انتقادها و پیشنهادهایشان، در تمامی مراحل اجرای پایان‌نامه مرا حمایت و تشویق نمودند. از استادان شایسته جناب آقایان دکتر احمد نزاکتی و دکتر محمدرضا ربیعی که زحمت داوری این رساله را متقبل شدند، نهایت تشکر و قدردانی را دارم.

از تنها سرمایه زندگی‌ام، پدر بزرگوار و مادر عزیز، که هرچه امروز دارم از دعای خیر ایشان است و همواره قوت قلبی برای ادامه راهم بوده‌اند، تشکری ویژه و صمیمانه دارم. سپاس از خواهران و برادرانم که همواره در طول تحصیل حامی من در مواجهه با مشکلات بودند و هستی‌شان مایه‌ی دلگرمی من است.

بوسه می‌زنم بر دستان اولین معلمم سرکار خانم «سنجر» عزیز، که با همه وجودش آموخت تا بیاموزم و سپاسگزارم از جناب آقای «دکتر علی دولتی» استادی که سپیدی را بر تخته سیاه زندگیم نگاشت و برایم زندگی، بودن و انسان بودن را معنا کرد.

از استادان محترم جناب آقایان دکتر داود شاهسونی، دکتر حسین باغیشنی، دکتر علی مس فروش و سرکار خانم دکتر نگار اقبال، که در طول دوران تحصیلی‌ام در دوره کارشناسی ارشد، جهت آموزش و ارتقای علمی بنده، زحمت کشیده‌اند و صادقانه و عاشقانه تلاش می‌کنند تا نقّالی دانسته‌ها را به نقّادی اندیشه‌ها تبدیل سازند، قدردانی می‌نمایم. همچنین از جناب آقای دکتر مهدی کبرائی (دانشگاه صنعتی امیر کبیر)، به خاطر زحماتی که برای توصیف داده‌ها کشیدند، صمیمانه قدردانی می‌نمایم.

بسی شایسته است تشکر کنم از دوست عزیزم، خانم «مینا نوروزی راد» به پاس محبت بی‌دریغش که بدون هیچ چشم‌داشتی مشکلات مسیر را برایم تسهیل نمود.

به پاس قدردانی از دوستان خوبم خانم‌ها، مریم خان احمدی، سوسن عشوره، مهشید اخوان نژاد، عطیه سابقی، الهام آشوری و کوثر السادات تفاخ که لذت لحظه‌های ناب با هم بودن، آرامش روحی و آسایش فکری‌ام مدیون حضور سبز آن‌هاست.

مریم برزوئی بیدگلی

۱۳۹۳

تعهدنامه

اینجانب **مریم برزوئی بیدگلی** دانشجوی کارشناسی ارشد رشته **آمار** **دانشکده علوم ریاضی** دانشگاه شاهرود، نویسنده پایان‌نامه با عنوان **بررسی رفتار برآوردگر جک‌نایف ریج** ، تحت راهنمایی **دکتر محمد آرشی** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه شاهرود متعلق دارد، و مقالات مستخرج با نام "دانشگاه شاهرود" یا "Shahrood University" به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده) شده است، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

مریم برزوئی بیدگلی
۱۳۹۳

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

هنگامی‌که همخطی بین ستون‌های ماتریس طرح در رگرسیون خطی وجود دارد، استفاده از روش کمترین توان‌های دوم در برآورد ضرایب مدل، معمولاً باعث می‌شود که برآوردهای خیلی ضعیفی به‌دست آید. ثابت شده است که واریانس برآوردگرهای کمترین توان‌های دوم ضرایب رگرسیون ممکن است در حد قابل توجهی افزایش یابند و طول بردار برآورد کمترین توان‌های دوم پارامتر به طور متوسط خیلی زیاد می‌شود. در این راستا، یکی از راه‌های رفع مشکلات مطرح شده، استفاده از برآوردگرهای اریب است.

مسئله همخطی و کاربردی نبودن برآوردگر کمترین توان‌های دوم، منجر به گستردگی استفاده از برآوردگرهای اریب شده است. روش‌های متعددی مانند رگرسیون ریج، لیو و ... برای به‌دست آوردن برآوردهای اریب ضرایب رگرسیون مطرح شده است. یکی از این روش‌ها استفاده از برآوردگر اصلاح‌شده جک‌نایف لیو است، که اساس کار این برآوردگر این است که دو برآوردگر نوع لیو و جک‌نایف لیو را با هم ترکیب می‌کند و سپس برآوردگری ارائه می‌دهد که عملکرد بهتری نسبت به برآوردگر کمترین توان‌های دوم دارد. این برآوردگر اریب بوده اما، دارای واریانس کمتری نسبت به برآوردگر کمترین توان‌های دوم است.

در این پایان‌نامه مدل رگرسیون ریج را مورد بررسی قرار داده ایم. بدین منظور از نتایج برخی محققین بهره برده ایم که در نهایت منجر به برآوردگر جک‌نایف لیو اصلاح شده گردیده است. با توجه به پیچیدگی برآوردگر ارائه شده، از روش‌های شبیه‌سازی مونت‌کارلویی و جک‌نایف به منظور ارائه اریبی و مخاطره آن بهره برده ایم.

کلمات کلیدی: برآوردگر جک‌نایف، برآوردگر ریج، برآوردگر لیو، رگرسیون چندگانه، مخاطره، همخطی

پیش گفتار

آماردان‌ها از قدیم با مشکل انتخاب بین دو برآوردگر اریب مواجه بودند. در راستای انتخاب بین برآوردگرهای اریب، سعی نموده ایم با بهره‌گیری از تکنیک جک‌نایف برآوردگری را انتخاب کنیم که اریبی کمتری دارد. بدین منظور این تکنیک را برای برآوردگر اریب لیو به‌کار برده و برآوردگر جک‌نایف لیو اصلاح شده را معرفی کرده ایم.

در این راستا با مبنا قرار دادن مقاله‌ی آکدنیز دوران و آکدنیز[۱](۲۰۱۲) به دنبال کاهش اریبی برآوردهای اریب ضرایب رگرسیون می‌باشیم. آکدنیز دوران و آکدنیز (۲۰۱۲) در مقاله‌ی خود از روش جک‌نایف ارائه شده توسط هینکلی [۲](۱۹۷۷)، سینگ[۳]و همکاران (۱۹۸۶)، نیکوئیست [۴](۱۹۸۸) و باتا[۵]و همکاران (۲۰۰۸) برای محاسبه‌ی برآوردگر جک‌نایف لیو و برآوردگر جک‌نایف لیو تعمیم یافته، بهره برده اند. بدین منظور با یاری از نتایج مقاله آکدنیز دوران و آکدنیز (۲۰۱۲) و همچنین لی و یانگ[۶] (۲۰۱۲)، برآوردگر جک‌نایف لیو اصلاح شده را معرفی نموده ایم. با توجه به پیچیدگی برآوردگر ارائه شده، از روش‌های شبیه‌سازی مونت‌کارلویی و جک‌نایف به منظور محاسبه اریبی و مخاطره آن بهره برده ایم. این مجموعه شامل ۴ فصل و ۵ ضمیمه می‌باشد. مطالب هر فصل به طور مختصر عبارتست از:

- در فصل۱، مقدمات، نگاهی بر تاریخچه موضوع مورد بررسی آورده شده است. بنا به ضرورت، مقدماتی از رگرسیون خطی چندگانه، رگرسیون ریج، و برآوردگر نوع لیو ذکر گردیده است. در ادامه روش جک‌نایف تعریف و کاربرد آن در برآورد اریبی و واریانس ذکر شده است.
- در فصل۲، نوع دیگری از برآوردگر لیو را با عنوان برآوردگر جک‌نایف اصلاح شده لیو معرفی و کارایی آن را مورد بررسی قرار داده ایم. در نهایت با یک مثال عددی و همچنین شبیه‌سازی تلاش شده است که به برآوردگری که عملکرد بهتری نسبت به برآوردگر لیو تعمیم یافته دارد، دست یابیم.
- در فصل۳، برآوردگر معرفی شده توسط لی و یانگ (۲۰۱۲)، با عنوان برآوردگر لیو اصلاح شده مورد تحلیل قرار گرفته است. در این فصل نیز با یک مثال عددی و شبیه سازی عملکرد این برآوردگر نسبت به سایر برآوردگرهای مطرح شده، مورد بررسی و مقایسه قرار گرفته است.
- در فصل۴، برآوردگر لی و یانگ (۲۰۱۲) را با بهره‌گیری از تکنیک جک‌نایف به برآوردگری تبدیل کرده که دارای اریبی کمتری نسبت به برآوردگر قبلی است. برای مقایسه‌ی کارایی این برآوردگر با سایر برآوردگرها از روش‌های شبیه‌سازی مونت کارلویی استفاده کرده‌ایم.
- در ضمیمه۱، مقدمه‌ای بر جبر ماتریس‌ها به منظور سادگی درک اثبات قضایا و لم‌ها در متن پایان‌نامه مطرح گردیده است.

---

[۱] Akdeniz Duran and Akdeniz

[۲] Hinkley

[۳] Singh

[۴] Nyquist

[۵] Batah

[۶] Li and Yang

- در ضمیمه۲، تعاریف و لم‌های اساسی که در محاسبات به‌کار گرفته شده است، را بیان نموده ایم.
- در ضمیمه۳، توضیحاتی در خصوص آماره کاپا و عامل تورم واریانس به عنوان عاملی برای شناخت همخطی در مدل رگرسیون خطی داده شده است.
- در ضمیمه۴، برنامه‌های کامپیوتری با نرم افزارR ارائه شده است.
- در ضمیمه۵، خلاصه، بحث و نتیجه‌گیری و پیشنهادها برای ادامه‌ی کار در آینده مطرح گردیده است.

در این مجموعه، در مواردی که برهان از نویسنده بوده از علامت * و در صورتی که هم قضیه و هم برهان از نویسنده می‌باشد، از نماد ** استفاده شده است.

لیست مقالات مستخرج از پایان‌نامه

۱. برزوئی بیدگلی م. و آرشی م.، (۱۳۹۳). "کارایی برآوردگر جک‌نایف ریج"، دوازدهمین کنفرانس آمار ایران، کرمانشاه، ایران.

۲. برزوئی بیدگلی م. و آرشی م.، (۱۳۹۳). "کاربرد برآوردگر لیو اصلاح شده در صنعت پتروشیمی"، نخستین کنفرانس بین المللی تحقیق در عملیات برای تصمیم‌سازی بهینه در صنایع نفت، گاز، پتروشیمی و پالایش و پخش، تهران، ایران.

۳. برزوئی بیدگلی م. و آرشی م.، (۱۳۹۳). "برآوردگر جک‌نایف لیو اصلاح شده"، مجله علوم آماری، تحت داوری.

فهرست نشانه‌ها، نمادها و علائم اختصاری

$\mathbf{X}$	ماتریسX
X	بردار تصادفیX
$\mathbf{x}$	بردار تصادفی
x	یافته‌ی بردار تصادفی X
A'	ترانهاده‌ی ماتریس A
$X_{(i)}$	i- امین نمونه جک‌نایف
$\hat{\gamma}_{OLSE}$	برآوردگر کمترین توان‌های دوم γ
$\hat{\gamma}(k)$	برآوردگر ریج γ
$\hat{\gamma}(d)$	برآوردگر لیو γ
$\hat{\gamma}(\mathbf{D})$	برآوردگر لیو تعمیم یافته γ
$\hat{\gamma}_{JGLE}(d)$	برآوردگر جک‌نایف لیو γ
$\hat{\gamma}_{JGLE}(\mathbf{D})$	برآوردگر جک‌نایف لیو تعمیم یافته γ
$\hat{\gamma}_{MJGLE}(\mathbf{D})$	برآوردگر اصلاح شده جک‌نایف لیو تعمیم یافته γ
$\hat{\gamma}_{JMLE}(d,b)$	برآوردگر جک‌نایف لیو اصلاح شده γ
$\chi^{۲}(q)$	توزیع کی- دو مرکزی با درجه آزادی q
$N_p(\mu,\mathbf{\Sigma})$	نرمال p- متغیره با میانگین μ و واریانس Σ
$\kappa(\mathbf{X})$	عدد شرطی $\mathbf{X}$
VIF_j	عامل تورم واریانس برای مشاهده j- ام
$\hat{\sigma}^{۲}$	برآورد واریانس جامعه
ρ	ضریب همبستگی
$MSEM$	ماتریس میانگین توان دوم خطا
$SMSE$	اثر ماتریس میانگین توان دوم خطا
$ABIAS$	اثر ماتریس قدر مطلق اریبی
d_{mm}	برآورد نوع لیو تعمیم یافته مقدار d
$\hat{\beta}_{OLSE}$	برآوردگر کمترین توان‌های دوم β
$\hat{\beta}(k)$	برآوردگر ریج β
$\hat{\beta}(d)$	برآوردگر لیو β
$\hat{\beta}_{ME}$	برآوردگر آمیخته β
$\tilde{\beta}$	برآوردگر کمترین توان‌های دوم محدود شده β
$\hat{\beta}_{MRE}(k,b)$	برآوردگر ریج اصلاح شده β
$\hat{\beta}_{MLE}(k,b)$	برآوردگر لیو اصلاح شده β

فهرست نشانه‌ها، نمادها و علائم اختصاری

LSE	برآوردگر کمترین توان‌های دوم
LE	برآوردگر نوع لیو
GLE	برآوردگر لیو تعمیم یافته
JLE	برآوردگر جک‌نایف لیو
$JGLE$	برآوردگر جک‌نایف لیو تعمیم یافته
$MJLE$	برآوردگر جک‌نایف اصلاح شده لیو
$MJGLE$	برآوردگر جک‌نایف اصلاح شده لیو تعمیم یافته
RE	برآوردگر ریج
MRE	برآوردگر ریج اصلاح شده
MLE	برآوردگر لیو اصلاح شده
$JMLE$	برآوردگر جک‌نایف لیو اصلاح شده

فهرست مطالب

لیست تصاویر ص

لیست جداول ض

۱ مقدمه و دورنما ۱

۱.۱ مقدمه و تاریخچه ۱

۲.۱ رگرسیون خطی چندگانه ۴

۱.۲.۱ برآورد پارامترهای مدل ۴

۲.۲.۱ همخطی در رگرسیون چندگانه ۹

۳.۲.۱ روش‌هایی برای برخورد با همخطی چندگانه ۱۰

۳.۱ رگرسیون ریج ۱۱

۴.۱ برآوردگر نوع لیو ۱۵

۱.۴.۱ برآوردگر لیو تعمیم یافته ۱۶

۲.۴.۱ مثال عددی ۱۷

۵.۱ برآوردگر آمیخته ۲۱

۶.۱ روش جک‌نایف ۲۲

۲ برآوردگر جک‌نایف اصلاح شده لیو ۳۰

۱.۲ مقدمه ۳۰

۲.۲ معرفی برآوردگر جک‌نایف لیو تعمیم یافته ۳۰

۱.۲.۲ مثال عددی ۳۴

۳.۲ برآوردگر جک‌نایف اصلاح شده لیو تعمیم یافته ۳۵

۴.۲ کارایی برآوردگر جک‌نایف اصلاح شده لیو تعمیم یافته ۳۸

۱.۴.۲ مثال عددی ۳۹

۵.۲ مقایسه برآوردگر جک‌نایف اصلاح شده لیو تعمیم یافته و برآوردگر جک‌نایف لیو تعمیم یافته ۴۱

۶.۲ مثال عددی ۴۲

۷.۲ مطالعه شبیه‌سازی ۴۷

۳ برآوردگر لیو اصلاح شده لی و یانگ ۵۰
۱.۳ مقدمه ۵۰
۲.۳ برآوردگر لیو اصلاح شده ۵۴
۳.۳ مقایسه برآوردگرها ۵۵
۱.۳.۳ مقایسه بین برآوردگر لیو اصلاح شده و لیو ۵۶
۲.۳.۳ مقایسه بین برآوردگرهای لیو ، ریج و لیو اصلاح شده ۵۷
۳.۳.۳ مقایسه بین برآوردگرهای ریج اصلاح شده و لیو اصلاح شده ۶۰
۴.۳ مثال عددی ۶۲
۵.۳ مطالعه شبیه‌سازی ۶۵

۴ برآوردگر جک‌نایف اصلاح شده لی و یانگ ۶۸
۱.۴ مقدمه ۶۸
۲.۴ معرفی برآوردگر جک‌نایف لیو اصلاح شده ۶۹
۳.۴ اریبی و واریانس برآوردگر ۷۱
۴.۴ کارایی برآوردگر جک‌نایف لیو اصلاح شده ۷۳
۵.۴ مطالعه شبیه‌سازی ۷۴
۶.۴ نتیجه‌گیری ۷۵
۷.۴ پیشنهادات برای آینده تحقیق ۷۶

آ جبر ماتریس‌ها ۷۸

ب تعاریف، لم‌ها و قضایای اساسی ۸۲

پ شاخص‌های همخطی ۸۵
پ.۱ آماره کاپا ۸۵
پ.۲ عامل تورم واریانس ۸۵

ت برنامه‌های کامپیوتری با نرم افزار R ۸۷
ت.۱ برنامه مربوط به مثال ۲.۴.۱ . . . . . ۸۷
ت.۲ برنامه مربوط به مثال ۱.۲.۲ . . . . . ۹۱
ت.۳ برنامه مربوط به مثال ۱.۴.۲ . . . . . ۹۴
ت.۴ برنامه مربوط به مثال ۶.۲ . . . . . ۹۷
ت.۵ برنامه مربوط به بخش ۷.۲ . . . . . ۱۰۳
ت.۶ برنامه مربوط به مثال ۴.۳ . . . . . ۱۰۶

ت.۷ برنامه مربوط به بخش ۵.۳ . . . . . ۱۱۱
ت.۸ برنامه مربوط به بخش ۵.۴ . . . . . ۱۱۴

مراجع ۱۱۸

واژه‌نامه فارسی به انگلیسی ۱۲۳

واژه‌نامه انگلیسی به فارسی ۱۲۵

نمایه ۱۲۷

لیست تصاویر

۱.۱ عملکرد برآوردگر نوع لیو به ازای مقادیر مختلف d ، $0 < d < 1$ ۲۰

۱.۲ عملکرد برآوردگر لیو به ازای مقادیر مختلف d ، $0 < d < 1$ ۳۶

۲.۲ عملکرد برآوردگر لیو به ازای مقادیر مختلف d ، $0 < d < 1$ ۴۰

۳.۲ مقایسه برآوردگرها به ازای مقادیر مختلف d ، $0 < d < 1$ ۴۶

۱.۳ برآورد MSE برای برآوردگرهای RE و MRE ۶۴

۲.۳ برآورد MSE برای برآوردگرهای LE و MLE ۶۴

لیست جداول

۱.۱ داده‌های در صد تبدیل n- هپتان به استیلن ۱۸
۲.۱ برآورد $\hat{\beta}(d)$ برای مقادیر مختلف d ۱۹

۱.۲ برآورد $\hat{\beta}(d)$ برای مقادیر مختلف d ۳۵
۲.۲ برآورد $\hat{\beta}(d)$ برای مقادیر مختلف d ۴۱
۳.۲ همبستگی بین متغیرها ۴۳
۴.۲ برآورد β برای مقادیر مختلف d ۴۵
۵.۲ برآورد β برای مقادیر مختلف d ۴۵
۶.۲ برآورد β برای مقادیر مختلف d ۴۷
۷.۲ برآورد $MSEM$ برای $OLSE$ ، LE ، JLE و $MJLE$ به ازای (الف) $n = ۱۵$ و $p = ۳$، (ب) $n = ۵۰$ و $p = ۳$ ، (پ) $n = ۱۰۰$ و $p = ۳$ ۴۸
۸.۲ برآورد قدر مطلق اریبی برای $OLSE$ ، LE ، JLE و $MJLE$ به ازای (الف) $n = ۱۵$ و $p = ۳$ ، (ب) $n = ۵۰$ و $p = ۳$، (پ) $n = ۱۰۰$ و $p = ۳$ ۴۹

۱.۳ برآورد MSE برای LSE ، MLE ، LE ، RE و MRE ۶۳
۲.۳ برآورد MSE برای LSE ، MLE ، LE ، RE و MRE به ازای $\gamma = ۰/۷۰$. ۶۶
۳.۳ برآوردMSE برای LSE ، MLE ، LE، RE و MRE به ازای $\gamma = ۰/۹۰$ . . ۶۶
۴.۳ برآورد MSE برای LSE ، MLE ، LE ، RE و MRE به ازای $\gamma = ۰/۹۹$. ۶۶

۱.۴ برآورد MSE برای LSE ، LE ، MLE و $JMLE$ به ازای $\gamma = ۰/۷۰$ ۷۵
۲.۴ برآوردMSE برای LSE ، LE ، MLE و $JMLE$ به ازای $\gamma = ۰/۹۰$ ۷۵
۳.۴ برآورد MSE برای LSE ، LE ، MLE و $JMLE$ به ازای $\gamma = ۰/۹۹$ ۷۵

فصل ۱

مقدمه و دورنما

۱.۱ مقدمه و تاریخچه

در کتب آماری، رگرسیون به عنوان روش تعیین و تحلیل روابط نادقیق بین متغیرهای آماری تعریف می‌شود. بنابراین در رگرسیون دو مطلب تعیین روابط بین متغیرها و تحلیل روابط به‌دست آمده مورد توجه قرار می‌گیرد.

رگرسیون در لغت به معنای «برگشت» است. دلیل این نام‌گذاری به مطالعات گالتون[۱] (۱۸۸۶) در مورد رابطه بین قد پسر و قد پدر مربوط می‌شود. گالتون با رسم نمودار قد پسر در مقابل قد پدر (نمودار پراکنش) بر اساس تعداد زیادی از داده‌ها (مشاهدات) به این نتیجه رسید که اغلب افراد بلند قد، پسرانی کوتاه قدتر و اکثر افراد کوتاه قد، پسرانی بلند قدتر از خود دارند و این پدیده را «برگشت به میانگین[۲]» نامید و از آن پس نام رگرسیون بر روش‌های بررسی روابط بین متغیرهای آماری باقی ماند. تعیین روابط بین متغیرهای آماری را رگرسیون توصیفی و تحلیل این روابط را رگرسیون استنباطی می‌نامند.

به عنوان مثالی از یک مسئله که در آن تحلیل رگرسیونی می‌تواند مفید واقع شود، فرض کنید یک مهندس صنایع، توسط یک سازنده نوشابه استخدام شده که محصول تحویلی و عملیات سرویس و خدمت رسانی برای ماشین‌های فروش را تجزیه و تحلیل کند. او می‌خواهد بداند که زمان لازم برای این‌که یک فرد تحویل‌دهنده، این ماشین را سرویس دهد بستگی به تعداد موارد محصول تحویل شده دارد یا خیر. مهندس ۲۵ فروشنده جزئی را که دارای ماشین فروش می‌باشند، به تصادف بازدید کرده است و زمان تحویل (بر حسب دقیقه) و حجم محصول تحویل داده شده (بر حسب مورد) برای هر فقره، مشاهده می‌شود.

رابطه بین زمان تحویل با متغیرهای ذکر شده در بالا به شکل معادله یا الگویی است که متغیر وابسته (زمان تحویل) را به یک یا چند متغیر پیش‌بین[۳] مربوط می‌کند. در رگرسیون خطی متغیر پاسخ پیوسته

---

[۱] Francis Galton

[۲] Regression to the mean

[۳] Predictor

است ولی متغیرهای پیش‌بین می‌توانند پیوسته، گسسته یا رسته‌ای[۴] باشند. در این مثال زمان تحویل متغیر وابسته بر حسب دقیقه محاسبه می‌شود و متغیرهای پیش‌بین تعداد مختلف محصول تحویل شده می‌باشد. متغیر پاسخ را با y و متغیرهای پیش‌بین را با $X_1, X_2, \cdots, X_p$ نشان می‌دهیم.

متغیرهای پیش‌بین را «متغیرهای مستقل» نیز می‌نامند، اما بکار بردن این اصطلاح ممکن است تا حدی گمراه کننده باشد، زیرا همیشه $X_1, X_2, \cdots, X_p$ ، مستقل از یکدیگر نیستند و معمولاً از y مستقل نیستند. متغیر پاسخ نیز، گاهی اوقات «متغیر وابسته» نامیده می‌شود. برای توضیحات بیشتر، نیرومند (۱۳۸۷) را ببینید.

تحلیل رگرسیونی کاربردی‌ترین و مهم‌ترین روش آماری در تحلیل‌های آماری، اقتصادی، مالی، صنعتی و ... می‌باشد؛ به‌خصوص در ۴ دهه اخیر به طور ماهرانه‌ای در حل و بررسی مسائل کاربردی مورد استفاده قرار گرفته است. در بین تعداد زیاد شاخه‌های تحلیل رگرسیونی، روش برآورد کمترین توان‌های دوم تعمیم یافته[۵]، بر اساس نظریه معروف گوس- مارکف[۶] (کاریا و کوراتا[۷]، ۲۰۰۴ را ببینید) روشی پایه‌ای می‌باشد و هنوز نقشی اساسی در بسیاری از جنبه‌های نظری و کاربردی استنباط آماری به‌خصوص مدل رگرسیون خطی ایفا می‌کند.

اگر هیچ رابطه خطی بین متغیرهای پیش‌بین وجود نداشته باشد، این متغیرها متعامد[۸] نامیده می‌شوند. هنگامی‌که متغیرهای پیش‌بین متعامد باشند، برآوردگر کمترین توان‌های دوم پر کاربرد خواهند بود. اما متأسفانه، در بسیاری از موارد در عمل، متغیرهای پیش‌بین متعامد نیستند. گاهی اوقات، عدم وجود تعامد در متغیرهای پیش‌بین، مسأله بسیار حادی نیست. اما اگر در یک الگوی رگرسیون متغیرهای پیش‌بین به صورت کامل و یا تقریباً کامل، وابسته خطی باشند، در این صورت ممکن است استنباط بر اساس این الگوی رگرسیون تا حدی بسیار گمراه‌کننده و پر اشتباه باشد. در این هنگام می‌گوییم که میان متغیرها در الگوی رگرسیون همخطی وجود دارد.

مسئله همخطی و کاربردی نبودن برآوردگر کمترین توان‌های دوم، منجر به استفاده گسترده ای از برآوردگرهای اریب شده است. برخی روش‌های برآوردیابی اریب برای رفع مشکل همخطی، عبارتند از برآورد انقباضی استاین[۹]، برآورد ریج[۱۰] ، برآورد نوع لیو[۱۱] و ... (برای مطالعه بیشتر در زمینه برآوردگرهای بهبود یافته، به لیو[۱۲] ۲۰۰۳، راداکریشنا رائو و توتنبرگ[۱۳] ۱۹۹۵، سوئیندل[۱۴] ۱۹۷۶ و حسن زاده بشتیان ۱۳۸۸ مراجعه کنید).

---

[۴] Categorical
[۵] Generalized least squares
[۶] Gauss- Markov
[۷] Kariya and Kurata
[۸] Orthogonal
[۹] Stein shrinkage estimation
[۱۰] Ridge estimation
[۱۱] Liu-type estimation
[۱۲] Liu
[۱۳] Radhakrishna Rao and Tuotenburg
[۱۴] Swindel

یکی از مباحث مهم در استنباط آماری، بررسی ویژگی‌های اصلی برآوردها از جمله اریبی، واریانس و توزیع نمونه‌ای آن‌هاست. استفاده از برخی روش‌های معمول برای این منظور، عمدتاً پیچیده و حتی غیر ممکن هستند و در برخی از آنان نیاز به برقراری فرض‌های تئوری پیچیده وجود دارد. در چنین شرایطی، استفاده از روش‌های بازنمونه‌گیری می‌تواند راهگشا باشد، زیرا مهم‌ترین مزیت این روش‌ها عدم نیاز به محاسبات تئوری و برقرای فرض‌های پیچیده است. در سال‌های اخیر توجه زیادی به روش‌های بازنمونه‌گیری در رگرسیون شده است. روش‌های بوت استرپ[۱۵] و جک‌نایف[۱۶] عمده ترین روش بازنمونه‌گیری می‌باشند. روش جک‌نایف در سال ۱۹۴۹ توسط کوئینلی معرفی شد. روش بوت استرپ برای اولین بار توسط بردلی افرن[۱۷] در سال ۱۹۷۹ در مقاله‌ای با عنوان «بوت استرپ: نگاهی دیگر به روش جک‌نایف»، معرفی و در این مقاله از آن برای برآورد اریبی، واریانس و توزیع نمونه‌ای آماره‌ها استفاده شد. بوت استرپ معرفی شده توسط افرن، بوت استرپ ناپارامتری[۱۸] بود. ولی بعدها انواع مختلف بوت استرپ از جمله بوت استرپ پارامتری[۱۹]، بوت استرپ هموار[۲۰]، بوت استرپ بیزی[۲۱] و بوت استرپ دوگانه[۲۲] نیز معرفی شدند. برای آگاهی بیشتر به روبین[۲۳] (۱۹۸۱)، سیلورمن و یانگ[۲۴] (۱۹۸۷)، وایند[۲۵] (۱۹۹۵)، و دیویسن و هینکلی[۲۶] (۱۹۹۷) مراجعه کنید.

استفاده از روش‌های بوت استرپ در رگرسیون اولین بار توسط دیوید فریدمن[۲۷] در مقاله‌ای با عنوان «بوت استرپ کردن مدل‌های رگرسیون» در سال ۱۹۸۱ مطرح شد.

کاربرد روش‌های بوت استرپ در رگرسیون بر اساس ثابت بودن و یا تصادفی بودن متغیرهای پیش‌بین است. در صورت ثابت بودن متغیرهای پیش‌بین از روش بازنمونه‌گیری از مانده‌ها و در صورت تصادفی بودن، از روش بازنمونه‌گیری از مشاهدات استفاده می‌شود.

از زمان چاپ مقاله ای توسط افرن و گانگ[۲۸] در سال ۱۹۸۳ توجه بسیاری از آماردانان به روش جک‌نایف جلب شد و اهمیت و کاربرد آن روز به روز آشکارتر گردید، به طوری‌که از اواخر دهه ۷۰ تا اواسط دهه ۸۰، بیش از ده هزار مقاله در مورد بوت استرپ، ویژگی‌ها و کاربردهای آن در زمینه‌های مختلف نوشته شد. نیکوئیست در سال ۱۹۸۸ تکنیک جک‌نایف را برای رگرسیون ریج به کار برد و به این نتیجه رسید که برآوردگر ریج جک‌نایف دارای پارامتر اریبی کمتر و واریانس بیشتری نسبت به

---

[۱۵] Bootstrap
[۱۶] Jackknife
[۱۷] Bradley Efron
[۱۸] Non-parametric bootstrap
[۱۹] Parametric bootstrap
[۲۰] Smoothed bootstrap
[۲۱] Bayesian bootstrap
[۲۲] Double bootstrap
[۲۳] Rubin
[۲۴] Silverman and Yang
[۲۵] Vinod
[۲۶] Davison and Hinkley
[۲۷] Freedman
[۲۸] Efron and Gong