دانشگاه فردوسی مشهد
دانشکده علوم ریاضی

پایان نامه کارشناسی ارشد در رشته آمار ریاضی

عنوان :

آزمون نرمال بودن در حضور داده‌های پرت

استاد راهنما:

دکتر هادی جباری نوقابی

استاد مشاور:

دکتر ناصررضا ارقامی

نگارنده:

حسین صابر جبدرق

بهمن ۱۳۸۹

اظهارنامه

اینجانب **حسین صابر جبدرق** دانشجوی دوره کارشناسی ارشد رشته **آماریاضی** دانشکده ریاضی دانشگاه فردوسی مشهد نویسنده رساله/پایان نامه **آزمون نرمال بودن در حضور داده‌های پرت** تحت راهنمایی **دکتر هادی جباری نوقابی** متعهد می‌شوم:

- تحقیقات در این رساله/پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در رساله/پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه فردوسی مشهد می‌باشد و مقالات مستخرج با نام « دانشگاه فردوسی مشهد » و یا « Ferdowsi University of Mashhad » به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی رساله/پایان نامه تأثیرگذار بوده‌اند در مقالات مستخرج از رساله/ پایان نامه رعایت شده است.
- در کلیه مراحل انجام این رساله/پایان نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این رساله/پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است، اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه‌ای، نرم افزارها و تجهیزات ساخته شده) متعلق به دانشگاه فردوسی مشهد می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در رساله/پایان نامه بدون ذکر مرجع مجاز نمی‌باشد.

بسمه تعالي

Graduate Studies Thesis\Dissertation Information
Ferdowsi University of Mashhad

Author: Hossein Saber Jabdaragh
Supervisor(s): Dr. Hadi Jabbari Noughabi
Advisor(s): Prof. Naserreza Arghami

Faculty: Mathematical Sciences	**Department:** Statistics	**Specialization:** Mathematical Statistics

Approval Date:	**Defence Date:** 23/1/2011
M.Sc. ● **Ph.D.** ○	**Number of Pages:** 107

Abstract:

Statistical models are often based on normal distributions and procedures for testing this distributional assumption are needed. Many goodness-of-fit tests suffer from the presence of outliers, in the sense that they may reject the null hypothesis even in the case of a single extreme observation. In this thesis, we show a possible extension of the Shapiro-Wilk test that is not affected by such a problem. The presented method is inspired by the forward search (FS), a new, recently proposed by Daniele Coin (2008), diagnostic tool. An application to univariate observations shows how the procedure is able to capture the structure of the data, even in the presence of outliers.

Signature of Supervisor: **Date:**	**Key Words:** **1.** Computational algorithm **2.** Forward search **3.** Least absolute deviation **4.** Outlier **5.** Shapiro-Wilk test

تقدیم به

پدر و مادر مهربان و عزیزم

قدردانی

بدین وسیله مراتب قدردانی و تشکر خود را نسبت به استاد راهنمای بزرگوارم آقای دکتر هادی جباری نوقابی و استاد مشاورم آقای دکتر ناصررضا ارقامی ابراز می دارم و برای ایشان توفیق روزافزون آرزومندم که بدون راهنمایی های ایشان در مراحل تحقیق پایان نامه، به انجام رساندن این پژوهش میسّر نبود. همچنین سپاس و قدردانی خود را به اساتید محترم دکتر جعفر احمدی و دکتر غلام رضا محتشمی که قبول زحمت نموده و پایان نامه‌ام را مورد مطالعه و داوری قرار داده‌اند ، تقدیم می‌نمایم.

از کارمندان دانشکده، واحد انتشارات، اداره آموزش و بخش کتابخانه، دوستان خوبم آقایان بابک نصیری، سعید امیرنژاد، عبدالسعید توماج ، یاسر سنچولی، مجتبی اصفهانی و همه کسانی که به نوعی بر گردن بنده حقی دارند، سپاسگزارم.

بسمه تعالی .

مشخصات رساله/پایان نامه تحصیلی دانشجویان .

دانشگاه فردوسی مشهد

عنوان رساله/پایان نامه: آزمون نرمال بودن در حضور داده‌های پرت

نام نویسنده: حسین صابر جبدرق

نام استاد(ان) راهنما: دکتر هادی جباری نوقابی

نام استاد(ان) مشاور: دکتر ناصررضا ارقامی

دانشکده : علوم ریاضی	**گروه:** آمار	**رشته تحصیلی:** آمار ریاضی
تاریخ تصویب:		**تاریخ دفاع:** ۱۳۸۹/۱۱/۳
مقطع تحصیلی: **کارشناسی ارشد** ●	**دکتری** ○	**تعداد صفحات:** ۱۰۷

چکیده رساله/پایان نامه : اغلب مدل‌های آماری روی توزیع‌های نرمال پایه ریزی می‌شوند و روش‌هایی برای آزمون کردن این توزیع مورد نیاز است. بسیاری از آزمون‌های نیکویی برازش از حضور داده‌های پرت تأثیر منفی می‌پذیرند، به این معنی که ممکن است آن‌ها فرضیه صفر را حتی به خاطر وجود یک مشاهده بسیار بزرگ یا بسیار کوچک رد کنند. در این مطالعه بسطی از آزمون شاپیرو- ویلک را ارائه می‌دهیم که از چنین مشکلی تأثیر نمی‌پذیرد. روش حاضر الهام گرفته از روش جستجوی پیشرو[1] (FS) است که یک روش تشخیصی جدید است و اخیراً توسط دانیل کوین (۲۰۰۸) پیشنهاد شده است. با به کار گیری مشاهدات یک متغیر نشان داده می‌شود که این روش قادر است ساختار داده‌ها را حتی در حضور داده‌های پرت تشخیص دهد.

کلید واژه:

۱. اگوریتم محاسبه
۲. جستجوی پیشرو
۳. کمترین انحراف مطلق
۴. داده‌ی پرت
۵. آزمون شاپیرو-ویلک

امضای استاد راهنما:

تاریخ:

[1] Forward Search

فهرست مندرجات

فهرست جداول

عنوان صفحه

فهرست شکل‌ها

عنوان صفحه

نمادها

برآوردگر رگرسیونی کمترین مربعات میانه ..... LMS

برآوردگر رگرسیونی کمترین مربعات خطا ..... LS

برآوردگر رگرسیونی کمترین مقدار مطلق ..... L_1

الگوریتم روش کمترین انحراف مطلق ..... LAD

تابع توزیع تجربی ..... EDF

پیش‌گفتار

در این تحقیق آزمون نرمال بودن نمونه‌ای از مشاهدات که دارای داده‌های پرت است، بررسی می‌شود. نرمال بودن یکی از عمومی‌ترین فرض‌ها در توسعه و استفاده از روش‌های آماری است. بنابراین روش‌هایی برای آزمون کردن این توزیع (توزیع نرمال) مورد نیاز است. این مسأله توسط محققین زیادی مورد مطالعه و بررسی قرار گرفته است و در طول این بررسی آزمون‌هایی را یافتیم که تصور نمی‌کردیم وجود داشته باشند! در این بررسی حدود چهل روش آزمون را یافتیم که برای نرمال بودن پیشنهاد شده است و به همان تعداد روش‌هایی مثل رسم نمودار، آزمون داده‌های پرت، آزمون‌های نیکویی برازش تعمیم یافته و آزمون‌های دیگر که در کشف غیر نرمال بودن داده‌ها مفید هستند، وجود دارد. بنابراین، این بررسی ذره‌ای از این دریای بی‌کران است که مطمئناً یک بررسی جامع با تمام جزئیات نیست.

روش‌های زیادی برای آزمون نرمال بودن نمونه‌های یک متغیره در ادبیات آماری پیشنهاد شده است. فقط با در نظر گرفتن آزمون‌های با فرضیه صفر مرکب، آزمون‌های نیکویی برازش برای نرمال بودن را می‌توان در چهار گروه طبقه بندی کرد. اولی شامل اندازه فاصله بین تابع توزیع نظری و تابع توزیع تجربی است. دومین دسته شامل آماره‌هایی مرکب از

ضرایب چولگی و کشیدگی است. سوّمین خانواده با تعمیم آزمون نیکویی برازش پیرسن[۱] پایه ریزی می‌شود. چهارمین کلاس مربوط به روش‌های رگرسیونی است. بالاخره آخرین آن‌ها آزمون‌های نیکویی برازش مبتنی بر آنتروپی می‌باشد. برای اطلاعات بیشتر در مورد آزمون‌های نیکویی برازش مبتنی بر آنتروپی به پایان‌نامه خانم فاطمه یوسف‌زاده (۱۳۸۷) مراجعه کنید.

مشکل عمومی بسیاری از این آزمون‌ها حساسیت به حضور داده‌های پرت در نمونه است. در حقیقت، تنها یک مشاهده پرت می تواند منجر به رد فرضیه صفر (فرضیه نرمال بودن) شود حتی اگر اکثر داده‌ها از توزیع نرمال به دست آمده باشند.

در این تحقیق یک روش اصلی که با وفق دادن روش جستجوی پیشرو (FS) به آزمون‌های نیکویی برازش حاصل می شود، را معرفی می کنیم. بدین وسیله مشکل تأثیر مشاهدات پرت روی آزمون نیکویی برازش حل می‌شود. به طور اساسی FS از یک زیر مجموعه پرت از مشاهدات شروع می کند و مقیاس برای پیشرفت در جستجو (افزایش زیر مجموعه با استفاده از افزودن یک یا چند مشاهده در هر مرحله) مجموعه ای از روش‌های تشخیصی است که در طول جستجو به کار می‌روند (برای اطلاعات بیشترآتکینسون و ریانی ۲۰۰۰a، ۲۰۰۰b، ۲۰۰۱، ۲۰۰۲a، ۲۰۰۲b، ۲۰۰۴، آتکینسون و همکاران ۲۰۰۴ را ببینید).

به منظور وفق دادن FS برای آزمون نرمال بودن، به یک روش برای انتخاب یک زیر مجموعه پرت از مشاهدات و روش دیگر برای آزمون نرمال بودن نیاز داریم.

---

[۱] Pearson Goodness of Fit Test

این تحقیق شامل چهار فصل می‌باشد. فصل اول مقدمه و کلیات می‌باشد. در فصل اول تاریخچه‌ای از آزمون نرمال بودن، آزمون‌های نیکویی برازش مهم از جمله آزمون شاپیرو-ویلک، تعریف داده‌های پرت و روش‌های تشخیص این داده‌ها، برآوردگر رگرسیونی کمترین میانه مربعات و در کل مفاهیم و تعاریف مورد نیاز مطرح می‌شود. فصل دوم شامل دو روش است که محققین برای کشف داده‌های پرت در مشاهداتی که از جامعه نرمال به دست آمده‌اند، پیشنهاد کرده‌اند. در فصل سوم روش پیشنهادی برای کشف داده‌های پرت در توزیع نرمال و در صورت وجود، یافتن زیر مجموعه‌ای نرمال از نمونه‌ای تصادفی مورد بحث قرار می‌گیرد. این فصل را با * مشخص کردیم که یک روش جدید و پیشنهاد خودمان است. بالاخره، فصل چهارم به روش جستجوی پیشرو آزمون شاپیرو-ویلک برای نرمال بودن اختصاص یافته است که ساختار داده‌ها را با وجود داده‌های پرت به خوبی تشخیص می‌دهد.

در نهایت یاد آوری می‌شود که در تنظیم این پایان نامه برای برگردان انگلیسی به فارسی از واژه نامه آمار و ریاضی استفاده شده است. در صورتی که واژه‌ای در این واژه نامه وجود نداشته است، با نظر استاد راهنما برگردان فارسی آن به کار برده شده است.

فصل ۱

مقدمه و کلیات

۱.۱ تاریخچه

روش‌های آماری مثل آزمون استودنت، آزمون‌هایی برای ضرایب رگرسیونی، آنالیز واریانس و آزمون فیشر برای همگنی واریانس، همگی دارای یک فرض اساسی هستند که باید نمونه از توزیع نرمال به دست آمده باشد. البته در روش‌های آماری یا باید فرض نرمال بودن را آزمون کنیم که در صورت رد فرضیه صفر از روش‌های ناپارامتری برای بررسی داده‌ها استفاده کنیم، یا این که نشان دهیم در صورتی که داده‌ها دارای توزیع نرمال نباشند، تأثیری بر روی روش‌های بررسی مشاهدات ندارد. به عبارت دیگر باید نشان دهیم که نرمال بودن یا نبودن مشاهدات هیچ فرقی ندارند. بسیاری از محققان آماری با ارزیابی دامنه تأثیر تخطی از این فرض روی سطح معنی داری آزمون یا اثر بخشی برآورد کردن پارامتر، را بررسی کرده‌اند.

ارزیابی محققان در مورد تأثیرات تخطی از فرض نرمال بودن روی روش های استاندارد آماری به تاریخ قبل از مقاله بارتلت (۱۹۳۵) روی آزمون استودنت برمی‌گردد. فیشر (۱۹۶۴) در مورد این مسأله تحقیق کرد و نتایج مربوط به انباشتگی آماره های چولگی و کشیدگی در آزمون های نرمال بودن را توسعه داد.

مسأله بررسی نرمال بودن توزیع برای استواری [۱] آزمون فرضیه‌ها در ادبیات آماری مورد توجه قرار گرفته است. مرور این ادبیات و آزمون های نرمال بودن، تلاش و همت بسیاری از نظریه پردازان و شاغلان آماری را نشان می‌دهد.

پیتمن (a ۱۹۳۷ و b ۱۹۳۷) مسأله حساسیت آزمون استودنت و آنالیز واریانس یک طرفه را بررسی کرد و بدین ترتیب با استفاده از یک تبدیل نظری به نتایج مهمی برای عدم حساسیت آزمون استودنت و آنالیز واریانس یک طرفه به تخطی از نرمال بودن را ارائه نمود.

[۱] Robustness