

تحلیل استوار واریانس بر اساس توزیع جایگشتی میانگین پیراسته

کوروش دادخواه و ادیس صمدی

گروه آمار، دانشگاه کردستان

تاریخ دریافت: ۱۳۹۶/۳/۲ تاریخ آخرین بازنگری: ۱۳۹۷/۲/۲

چکیده: ساختار آزمون تحلیل واریانس به گونه‌ای است که در صورت حضور داده‌های دورافتاده در بین مشاهدات، نتایج آزمون می‌تواند به اشتباه منجر به رد یا پذیرش فرض صفر شود. در این مقاله، روش استوار توزیع جایگشتی آماره F بر اساس میانگین پیراسته پیشنهاد شده است. این روش، به کمک توزیع جایگشتی تابعی از میانگین پیراسته، حساسیت نسبت به فرض‌های کلاسیک همچون نرمال بودن داده‌ها و حضور داده‌های دورافتاده را کاهش داده و اعتبار نتایج حاصل را تضمین می‌کند. روش پیشنهادی با روش تحلیل استوار واریانس بر اساس رهیافت جستجوی پیشرو مقایسه می‌شود. روش ارائه شده برخلاف روش مبتنی بر جستجوی پیشرو، ما را از فرضیات محدودکننده پارامتری بی‌نیاز و از نظر محاسباتی زمان کمتری را صرف می‌کند. نتایج بررسی‌های عددی روی خطای نوع اول و توان آزمون، حکایت از عملکرد خوب این روش استوار در مقایسه با روش رقیب دارد.

واژه‌های کلیدی: تحلیل واریانس، استواری، داده‌های دورافتاده، توزیع جایگشتی، جستجوی پیشرو.

۱ مقدمه

توجه به حضور مشاهدات دورافتاده در بین داده‌ها از مسائل مهم در تحلیل داده‌ها هستند. تنها یک داده تشخیص داده نشده می‌تواند کل نتایج تحلیل را فاقد اعتبار کند. داده‌های دورافتاده مشاهداتی ناسازگار هستند که با بخش عمده‌ای از داده‌ها اختلاف زیادی دارند. هاوکینز (۱۹۸۰) بیان می‌کند که «داده دورافتاده مشاهده‌ای است که آن‌قدر از بقیه داده‌ها انحراف داشته باشد که این ایده را در ذهن بیاورد که توسط روش

و مکانیزم دیگری تولید شده است».

در این مقاله دو هدف دنبال می‌شود: ابتدا آماره‌ای تعریف می‌شود که در حضور مشاهدات پرت استوار باشد. سپس به کمک توزیع جایگشتی آماره فوق، انجام آزمون تحلیل واریانس از فرضیات محدودکننده‌ای همچون نرمال بودن داده‌ها آزاد خواهد شد. این کار باعث می‌شود بتوان روش پیشنهادی را برای داده‌هایی که دارای توزیع‌های غیر نرمال و با دم‌های کلفت نیز هستند به کاربرد.

آزمون جایگشتی نوعی آزمون معنی‌داری آماری است که در آن توزیع نمونه آماره آزمون با محاسبه تمام مقادیر ممکن برای این آماره تحت تمام جایگشت‌های ممکن داده‌ها به دست می‌آید. یک آزمون جایگشتی در برخی موارد آزمون دقیق^۱ یا آزمون تصادفی^۲ نامیده می‌شود. آزمون جایگشتی به آزمون تصادفی کردن فیشر نیز معروف است. این آزمون را اولین بار فیشر (۱۹۳۴) ارائه داد به همین خاطر آن را آزمون تصادفی کردن فیشر می‌نامند.

آزمون جایگشتی برای حجم نمونه زیاد بسیار زمان‌بر است و به علت کمبود امکانات در دهه‌های پیشین، اجرای آزمون برای حجم نمونه زیاد تقریباً غیرممکن بوده است. به همین دلیل این آزمون در اوایل مطرح شدن، مخالفان و منتقدان بسیاری داشت. آن‌ها اعتقاد داشتند برای حجم نمونه زیاد، اجرای آزمون غیرممکن است. اما با ظهور محاسبات آماری و نرم‌افزارهای مدرن کامپیوتری این اعتقاد کم‌کم رنگ باخت. این سخن فیشر (۱۹۳۶) در مورد آزمون‌های جایگشتی نشان‌دهنده نقش مهم این آزمون در میان آزمون‌های آماری است: حقیقتاً یک آماردان این روش خسته‌کننده را بکار نمی‌برد، اما نتیجه‌گیری او در برابر این واقعیت که این اتخاذ تصمیم می‌توانسته با روش بسیار ابتدایی آزمون جایگشتی صورت بگیرد، هیچ توجیهی ندارد.

فیشر (۱۹۳۵) آزمون جایگشتی برای زوج‌های جور شده را مورد بحث قرار داد. بعد از فیشر، آماردانان زیادی به این آزمون پرداختند. به عنوان مثال، اولین شالوده نظری آن توسط پیتمن (۱۹۳۷) ارائه و بسط داده شد. ذکر این نکته جالب توجه است که درحالی‌که فیشر فرض تصادفی بودن نمونه را برای انجام آزمون لازم دانسته بود، پیتمن بعد از ارائه اصول آزمون جایگشتی برای مقایسه دو جامعه ادعا کرد که این آزمون می‌تواند حتی در مواردی که نمونه‌ها تصادفی نیستند نیز به کار رود. مطالعات بعدی توسط لهمن و استین (۱۹۴۹) انجام داده شد که شامل بسط و گسترش نظری آزمون جایگشتی بهینه بود. سپس هافدینگ (۱۹۵۲) بر روی رفتار مجانبی آزمون جایگشتی مطالعاتی را انجام داد و همچنین مطالعاتی نیز توسط باکس و اندرسن (۱۹۵۵) در زمینه‌های ذکر شده صورت گرفته است.

^۱Exact Test

^۲Randomization Test

در این مقاله هدف، ارائه روشی استوار^۳ برای تحلیل واریانس یک طرفه^۴ با اثرات ثابت^۵ است. به همین منظور، روش توزیع استوار جایگشتی^۶ بر اساس تابعی از میانگین پیراسته پیشنهاد و نتایج آن با روش مبتنی بر جستجوی پیشرو^۷ مقایسه می‌شود. از این رو در بخش ۲ ابتدا اثر مشاهدات پرت روی روش کلاسیک تحلیل واریانس نشان داده می‌شود. سپس در بخش ۳ روش استوار تحلیل واریانس بر اساس رهیافت جستجوی پیشرو به‌عنوان روش رقیب مرور می‌شود. روش پیشنهادی در بخش ۴ معرفی می‌شود. در بخش ۵ کارایی روش پیشنهادی و روش رقیب به‌وسیله شبیه‌سازی و داده‌های واقعی مورد ارزیابی قرار می‌گیرد و در پایان، جمع‌بندی کارهای انجام‌شده در بخش بحث و نتیجه‌گیری ارائه می‌شود.

۲ تحلیل واریانس یک طرفه در حضور داده‌های دورافتاده

تحلیل واریانس یکی از گسترده‌ترین فنون آماری مورداستفاده در آزمون اختلاف میانگین‌های چند جامعه است. تحلیل واریانس اولین بار توسط فیشر (۱۹۱۸) مطرح شد. اولین مرجع برای جدول تحلیل واریانس به اثر منتشرشده فیشر (۱۹۲۳) برمی‌گردد و در فیشر (۱۹۲۵) به‌طور کامل ارائه و موردبررسی قرارگرفته است.

فرض کنید از یک عامل (متغیر)، a تیمار یا سطح متفاوت وجود داشته باشد و هدف مقایسه آن‌ها باهم باشد. مدل برای هر مشاهده به‌صورت:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, n_i$$

است، که در آن مشاهده y_{ij} از تیمار i ام، μ میانگین مشترک تمام تیمارها، τ_i اثر i امین تیمار، ε_{ij} مؤلفه خطای تصادفی و μ_i میانگین i امین تیمار است. با فرض آنکه $E(\varepsilon_{ij}) = 0$ ، $Var(\varepsilon_{ij}) = \sigma^2 < \infty$ ، ε_{ij} ها مستقل و دارای توزیع نرمال با واریانس ثابت σ^2 هستند، تحلیل واریانس تک‌عاملی ارائه می‌شود. در مدل با اثرات ثابت، اثرهای تیماری معمولاً به‌صورت انحراف از میانگین کل تعریف می‌شوند، به‌طوری‌که

³Robust

⁴One-Way Analysis of Variance

⁵Fixed Effect

⁶Permutation Test

⁷Forward Search

$\sum \tau_i = 0$ است. برای آزمون فرض برابری اثر تیمارها، یعنی

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_a \\ H_1 : \mu_i \neq \mu_j \quad (i, j) \text{ جفت برای حداقل یک} \end{cases}$$

کل تغییرپذیری در داده‌ها را می‌توان به صورت

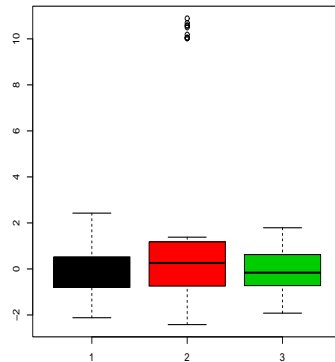
$$SS_{Total} = SS_T + SS_E$$

تجزیه کرد، که در آن مجموع توان‌های دوم حاصل از تیمارها (تغییرات بین تیمارها) و SS_E مجموع توان دوم حاصل از خطا (تغییرات درون تیمارها) هستند. چون مجموع درجات آزادی SS_T و SS_E ، یعنی $N - 1$ برابر کل درجات آزادی است، لذا می‌توان نشان داد که $\frac{SS_T}{\sigma^2}$ و $\frac{SS_E}{\sigma^2}$ متغیرهای تصادفی مستقل با توزیع کای دو هستند. پس اگر فرض صفر، یعنی فرض عدم وجود اختلاف در میانگین تیمارها درست باشد، آنگاه آماره

$$F = \frac{\frac{SS_T}{a-1}}{\frac{SS_E}{N-a}} \sim F_{(a-1, N-a)} \quad (1)$$

دارای توزیع فیشر است. فرض H_0 رد می‌شود اگر، $F > F_{(\alpha, a-1, N-a)}$. به دلیل وجود آماره \bar{y}_i در جفت SS_T و SS_E ، مقدار آماره F شدیداً تحت تأثیر حضور داده‌های دورافتاده قرار دارد. در حقیقت میانگین نمونه، تحت فرض نرمال، بهترین برآوردگر ناریب برای میانگین جامعه است، اما در صورت وجود آلودگی در داده‌ها، کارایی آن به شدت کاهش یافته و باعث شبه در اعتبار استنباط‌های آماری مبتنی بر آن می‌شود.

برای مثال یک مجموعه داده نمونه‌گیری شده با 50° مشاهده در سه گروه تحلیل واریانس یک طرفه متعادل را در نظر بگیرید. واحدها در هر گروه توسط توزیع نرمال استاندارد تولید می‌شود، اما گروه دوم توسط $U(10, 11)$ آلوده شده است. شکل ۱ نمودار جعبه‌ای نتایج این نمونه را نشان می‌دهد. آماره F دارای درجه‌های آزادی ۲ و ۱۴۷ و مقدار $9/0.24$ با p -مقدار $10^{-4} \times 2/0.1$ است، که به دلیل آلودگی در ناحیه رد می‌افتد.



شکل ۱. نمودار جعبه‌ای که نشان می‌دهد گروه ۲ شدیداً آلوده است.

۳ روش جستجوی پیشرو

روش جستجوی پیشرو که برای کشف اثر مشاهدات دورافتاده توسط برتاجینی و ورنال (۲۰۰۷) پیشنهاد داده شده، متشکل از سه مرحله است. اول، روی انتخاب زیرمجموعه اولیه تمرکز می‌کند، دوم، به اجرای جستجوی پیشرو اشاره می‌کند و سوم به نظارت و بررسی آماره‌ها در طول جستجو می‌پردازد. روش جستجوی پیشرو با ویژگی‌های مدل ANOVA سازگار است، خصوصاً حضور گروه‌ها در ساختار داده‌ها را در نظر می‌گیرد. در مدل‌سازی بر اساس مشاهدات $S^{(n)} = \{x_1, \dots, x_n\}$ نقطه شروع روش پیشرو، برازش مدل به زیرمجموعه اولیه‌ای از مشاهدات است، که باید فاقد داده پرت باشد و انتخاب آن‌ها می‌تواند به هر روش استواری صورت گیرد. برای مثال می‌توان از قدر مطلق فاصله مشاهدات از میانه هر گروه به‌عنوان معیاری برای اندازه‌گیری دوری از مرکز کمک گرفت. روش پیشنهادی در ساختار ANOVA، استفاده از مشاهدات y_{ij} است که در رابطه $\min |y_{ij} - \text{med}_i|$ صدق می‌کنند. که در آن $(i = 1, \dots, a)$ میانه گروه i ام است. تعداد اعضاء زیرمجموعه اولیه $S^{(*)}$ برابر یا بزرگ‌تر از a است. سپس همه مشاهدات با توجه به نزدیکی آن‌ها به مدل برازش داده شده مرتب شده و اندازه زیرمجموعه افزایش می‌یابد. این فرایند تا ورود همه مشاهدات به زیرمجموعه اولیه ادامه می‌یابد. از طریق جستجو سعی می‌شود علاوه بر شناسایی داده‌های دورافتاده، به تحلیل اثرشان روی برآورد پارامترها و بر استنباط روی مدل نیز پرداخته شود. در ادامه روش جستجوی پیشرو برای مدل تحلیل واریانس با اثرات ثابت بررسی می‌شود.

از دو روش متناسب^۸ و غیرمتناسب^۹ در اجرای الگوریتم استفاده می‌شود. تفاوت اساسی آن‌ها در تعداد واحدهایی است که در طول جستجو به زیرمجموعه اضافه می‌شوند. روش غیرمتناسب در هر مرحله فقط یک واحد جدید را اضافه می‌کند درحالی‌که در روش متناسب، حداقل مشاهدات لازم نسبت به ترکیب کلی (نسبت گروه) نمونه را به زیرمجموعه اضافه می‌کند. در حرکت از $S^{(m)}$ به $S^{(m+1)}$ بیشتر اوقات فقط یک واحد جدید به زیرمجموعه قبلی اضافه می‌شود. همچنین ممکن است وقتی یک واحد یا بیشتر خارج می‌شود دو یا چند واحد بیشتر وارد $S^{(m+1)}$ شوند. با این حال چنین رویدادی کاملاً غیرمعمول است و این تنها وقتی اتفاق می‌افتد که شامل یک واحد متعلق به یک خوشه از داده‌های دورافتاده باشد. زیرمجموعه‌ی $S^{(m)}$ دارای اندازه $m = \sum_{i=1}^a m_i$ است، که در آن m_i تعداد مشاهدات در گروه i در مرحله m است. اما در روش متناسب جستجوی پیشرو با اضافه کردن ℓ واحد به جای یک واحد به مرحله بعد حرکت می‌کند، که تحت شرط

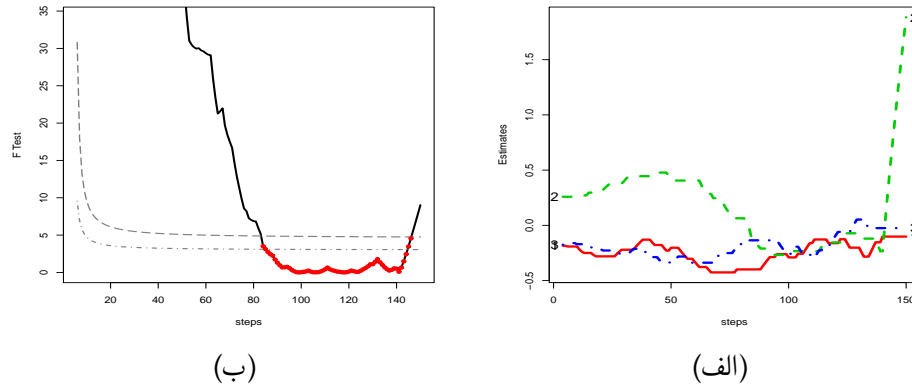
$$\frac{\ell_i}{\ell} \approx \frac{(n_i - \ell_i)}{n}$$

برابر $\ell = \sum_{i=1}^a \ell_i$ است. در این حالت ممکن است وقتی یک یا چند مشاهده خارج می‌شوند بیشتر از ℓ واحد جدید به زیرمجموعه اضافه شود. در هر مرحله از جستجو، جمع‌آوری اطلاعات روی برآورد پارامترها، مانده‌ها و دیگر آماره‌های مرتبط، راهنمایی محقق در تشخیص داده‌های دورافتاده و تحلیل اثرشان روی استنباط مدل را ممکن می‌سازد. در اینجا به بررسی ضرایب برآورد شده و آماره F اکتفا می‌شود. برای نشان دادن کاربرد و مزایای روش جستجوی پیشرو، از مجموعه داده استفاده‌شده مربوط به بخش ۲ استفاده می‌شود.

نمودار (الف) در شکل ۲ نشان می‌دهد که چگونه μ_i ها در طول جستجو تغییر می‌کنند. همان‌طور که انتظار می‌رود، قسمت اول جستجو به دلیل آلودگی شدید باهم تفاوت زیادی دارند، میانه‌های سه گروه به وضوح متفاوت هستند، بنابراین در این مرحله از جستجو اثر ابعاد کوچک زیرمجموعه‌ها، تغییرپذیری بالای برآوردها را به دنبال دارد. همان‌طور که جستجو ادامه می‌یابد، ضرایب برآورد شده ثابت و زمانی که تعداد مراحل نزدیک ۱۴۰ است تغییرپذیری برآوردها به سمت صفر میل می‌کنند. بعداً آن، برآورد μ_2 به شدت افزایش می‌یابد، که این نشان‌دهنده حضور داده‌های دورافتاده در گروه دوم است. نمودار (ب) در شکل ۲ چگونگی حرکت آماره F در طول روش جستجوی پیشرو را نشان می‌دهد. خطوط خط‌چین بیانگر چندک‌های ۹۵ درصد و ۹۹ درصد توزیع F فیشر با ۲ و ۳ - m درجه آزادی در هر مرحله است. خط

⁸Proportional

⁹Non-Proportional



شکل ۲. (الف) نمودارهای پیشرو ضرایب $\hat{\mu}_1$ تا $\hat{\mu}_3$ ، به ترتیب با خطوط شماره ۱ تا ۳ و (ب) آزمون F (خطوط خط چین، ناحیه پذیرش در سطوح معنی‌داری ۱ درصد و ۵ درصد هستند)

ممتد به مقادیر $F_{S(m)}$ به دست آمده در طی روند اشاره دارد. در ابتدا این مقادیر خیلی بزرگ هستند، همه آن‌ها در ناحیه رد قرار دارند و خارج از مقیاس نمودار واقع‌اند. پس از مدتی آماره کاهش می‌یابد و برای $83 \leq m \leq 144$ ، در ناحیه پذیرش قرار می‌گیرند.

آزمون F پیشرو می‌تواند به عنوان مجموعه $F_{FS} = \{F_{(k)}, \dots, F_{(n)}\}$ از آزمون‌های F کلاسیک در هر مرحله از جستجو تعریف شود. برای به دست آوردن آماره آزمون F پیشرو استوار، امکان جدا کردن نقطه برش از روند حرکت وجود دارد که با این کار گروه مشاهداتی که از قسمت عمده داده‌های دیگر تفاوت دارند را جدا می‌کند. همان‌طور که نشان داده شد، جستجوی نقطه برش نمی‌تواند به صورت ”خودکار” باشد و این کار بر اساس تحلیل گرافیکی انجام می‌پذیرد. کمترین نسبتی از داده‌های دورافتاده که باعث یک تغییر بزرگ بدون کنترل در مقدار برآورد می‌شود، نقطه فروزش نامیده می‌شود. استواری روش پیشرو از انتخاب یک برآوردگر بخصوص، با نقطه فروریزش بالا نتیجه نمی‌شود بلکه با گنجاندن تدریجی واحدها در زیرمجموعه‌ای که در مراحل اولیه خالی از داده‌های دورافتاده است صورت می‌گیرد.

۴ آزمون جایگشتی

برای انجام مسائل مربوط به آزمون فرض‌ها معمولاً علاوه بر فرض‌های زیربنایی چون تصادفی بودن نمونه‌گیری، مستقل بودن خطاها و ثابت بودن مقدار واریانس خطاها، به فرض نرمال بودن توزیع داده‌ها نیز

نیاز است. اما عموماً این فرض‌ها به‌ندرت در عمل رخ می‌دهند. بنابراین می‌توان از آزمون‌های جایگشتی که یکی از شاخه‌های آزمون‌های ناپارامتری است استفاده نمود. مزیت آزمون جایگشتی این است که نیازی به برقراری فرض نرمال ندارد. در سال‌های اخیر نیز مطالعات بسیاری بر روی آزمون‌های جایگشتی و کاربرد آن‌ها در روش‌های مختلف آماری و طرح آزمایش توسط آماردانان انجام شده است و با پیشرفت ابزارها و فنون محاسباتی، این روش توسط کامپیوترها و نرم‌افزارهای آماری انجام می‌شود. در ادامه روش آزمون جایگشتی برای k نمونه (هیگینز، ۲۰۰۴ و بوتار، ۲۰۰۷) معرفی می‌شود.

۱.۴ آزمون‌های جایگشتی k نمونه‌ای

فرض بر این است که واحدهای آزمایشی به k تیمار در یک طرح کاملاً تصادفی اختصاص داده شود یا اینکه مشاهدات به‌طور تصادفی از k جامعه انتخاب شده باشند. اگر تیمارها متفاوت باشند، فرض می‌شود که مشاهدات در حداقل یکی از تیمارها تمایل به بزرگ‌تر شدن از مشاهدات از حداقل یکی از تیمارهای دیگر دارد.

اگر مشاهدات از جوامع با توابع توزیع تجمعی $F_1(x), \dots, F_k(x)$ به‌طور تصادفی انتخاب شوند، آنگاه فرض صفر، برابری توزیع‌ها را آزمون می‌کند:

$$H_0 : F_1(x) = \dots = F_k(x)$$

$$H_1 : F_i(x) \leq F_j(x) \quad \text{یا} \quad F_i(x) \geq F_j(x)$$

برای حداقل یک جفت (i, j) با نابرابری اکید برای حداقل یک x

حالت خاصی از H_1 ، فرض مقابل انتقال به صورت $H_1 : F_i(x) = F(x - \mu_i)$ است. مدل تحلیل واریانس نیز شبیه این مدل است که در آن μ_i ها، میانگین‌ها هستند و F توزیع نرمال با میانگین صفر و واریانس σ^2 است. اندازه نمونه کل با $N = n_1 + \dots + n_k$ نشان داده می‌شود. اگر مشاهدات به‌صورت تصادفی از جوامع با توزیع نرمال با واریانس یکسان انتخاب شده باشند آنگاه آماره نسبتی F در رابطه (۱) تحت فرض H_0 دارای توزیع فیشر با درجه‌های آزادی $k - 1$ و $N - k$ است. ممکن است از این توزیع برای تعیین p -مقدار آزمون استفاده شود. با این حال اگر فرض نرمال قابل تأمین نباشد از آزمون F جایگشتی استفاده می‌شود.

۲.۴ مراحل انجام آزمون جایگشتی F

- ۱- آماره F برای داده‌های اصلی محاسبه و با F_{obs} نشان داده شود.
- ۲- همه جایگشت‌های ممکن N مشاهده از k تیمار محاسبه‌شود، که در آن n_i مشاهده در تیمار i ام و $\frac{N!}{n_1!n_2!\dots n_k!}$ تعداد جایگشت‌ها است. اگر امکان تولید همه جایگشت‌ها وجود نداشت، یک نمونه تصادفی از R جایگشت انتخاب شود. نمونه تصادفی $R = 1000$ جایگشت معمولاً برای بیشتر برنامه‌های کاربردی کافی است.
- ۳- برای هر جایگشت، آماره F محاسبه و با F^* نشان داده شود.
- ۴- مقدار p -مقدار از رابطه $p_value = \frac{\text{تعداد } F^* \geq F_{obs}}{\text{تعداد جایگشت‌های ارزیابی‌شده}}$ محاسبه شود.

۳.۴ فرم‌های جایگزین آماره F جایگشتی

مجموع توان‌های دوم کل به صورت

$$SS_{Total} = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$$

تعریف می‌شود. از ویژگی‌های مجموع توان‌های دوم در تحلیل واریانس داریم:

$$SS_{Total} = SS_T + SS_E$$

حال SS_{Total} دارای مقدار یکسان برای همه جایگشت‌های داده‌ها است، که با C نشان داده می‌شود. چون آماره

$$F = \frac{\frac{SS_T}{k-1}}{\frac{C-SS_T}{N-k}}$$

تابعی صعودی از SS_T است، آزمون جایگشتی بر اساس SS_T معادل استفاده از آن بر اساس F است. همچنین چون

$$SS_T = \sum_{i=1}^k n_i \bar{X}_i^2 - N \bar{X}^2$$

و \bar{X} دارای مقدار یکسان برای همه جایگشت‌های داده‌ها است، به شیوه هم‌ارزی می‌توان آزمون جایگشتی را بر مبنای

$$SS_X = \sum_{i=1}^k n_i \bar{X}_i^2$$

انجام داد.

۴.۴ آزمون جایگشتی بر اساس میانگین پیراسته

میانگین پیراسته معیاری دیگر از مرکز مجموعه‌ای از داده‌ها است که از آن در آزمون جایگشتی استفاده می‌شود. به گفته استیگلر (۱۹۷۶) میانگین پیراسته برای اولین بار توسط ژوزف گرگون (۱۸۲۱) ارائه شد. میانگین پیراسته به‌طور رسمی توسط مندلیف (۱۸۹۵) تعریف شد و اولین بار با این اسم توسط توکی (۱۹۶۲) نام‌برده شد. برای سال‌های بسیاری، میانگین پیراسته یک برآوردگر بسیار محبوب از پارامترهای مکانی^{۱۰} بوده است. در ادامه، برای نشان دادن کاربرد استفاده از این روش، از مجموعه داده استفاده‌شده در شکل ۱ مربوط به بخش ۲ استفاده می‌شود. مجموعه داده‌ها از سه گروه متعادل با توزیع نرمال استاندارد تشکیل شده است، اما گروه دوم به‌شدت توسط ۲۰٪ از مشاهدات با توزیع یکنواخت $U(10, 11)$ آلوده شده‌اند. همان‌طور که در بخش دوم مقاله بیان شد، p -مقدار در آزمون کلاسیک $10^{-4} \times 201$ است که به دلیل آلودگی در ناحیه رد می‌افتد. حال مقادیر مختلفی از پیراستگی مورد بررسی قرار می‌گیرد. p -مقدار آزمون جایگشتی بر اساس ۴ مقدار نمونه از این پیراستگی‌ها به‌صورت جدول ۱ به‌دست‌آمده است. در سطح معنی‌داری $\alpha = 0.05$ با پیراستگی ۱۸٪، p -مقدار در ناحیه رد قرار می‌گیرد، اما با قرار دادن میزان ۲۰٪ به بالا، p -مقدار در ناحیه پذیرش واقع می‌شود.

جدول ۱. جدول p -مقدارهای آزمون جایگشتی بر اساس ۴ مقدار مختلف پیراستگی

میزان پیراستگی	۰/۱۶	۰/۱۸	۰/۲۰	۰/۲۲
p -مقدار	۰/۰۰۰	۰/۰۳۱	۰/۴۸۰	۰/۴۹۹

¹⁰Location

۵ ارزیابی و مقایسه روش‌ها

در این بخش، بر اساس دو مدل شبیه‌سازی شده در ۵۰۰ مرتبه تکرار، عملکرد روش‌های کلاسیک، روش استوار مبتنی بر جستجوی پیشرو و روش پیشنهادی مبتنی بر جایگشت روی نرخ خطای نوع اول، توان آزمون و مدت‌زمان اجرای دو روش بررسی می‌شود. برای برآورد توزیع جایگشتی در روش پیشنهادی به یک نمونه $R_i = 1000$ تایی بسنده می‌گردد.

۱.۵ مقایسه روش‌ها بر اساس نرخ خطای نوع اول

مجموعه داده‌ای با سه گروه متعادل از اندازه نمونه افزایشی n_i مشاهده ($n_i = 20, 40, \dots, 200$); تنها گروه دوم توسط نرخ افزایشی $(\varepsilon = \varepsilon)$ ($i = 1, 2, 3$) از توزیع نرمال استاندارد در نظر بگیرد. بنابراین توزیع گروه دوم به صورت $(1-\varepsilon)N(0, 1) + \varepsilon N(4, 1)$ آلوده شده است. جدول ۲ فراوانی نسبی $rf_{(n_i, \varepsilon)}$ را برای ۵۰۰ مرتبه شبیه‌سازی که در آن آزمون $F \varepsilon N(4, 1)$ کلاسیک در سطح معنی‌داری $\alpha = 0.05$ در ناحیه رد می‌افتد را نشان می‌دهد، به عبارتی، تقریبی از احتمال خطای نوع اول را ارائه می‌دهد:

$$Pr(F_{(n_i, \varepsilon)} > F_{(0.05; df_1=2, df_2=2(n_i-1))}) \approx rf_{(n_i, \varepsilon)}$$

برای مثال، زوج $(n_i = 100, \varepsilon = 0.07)$ به این معنی است که ۷ مشاهده آلوده در گروه دوم قرار دارند و آزمون F کلاسیک، در ۵۰۰ بار تکرار آزمایش به غلط فرض صفر را ۱۷۸ مرتبه رد می‌کند بنابراین تقریب α مقدار ۰.۳۵۶ به دست می‌آید. همان‌طور که در بخش ۲ اشاره شد، در روش جستجوی پیشرو، نقطه برش بر اساس ارزیابی نمودارهای مختلف تعیین می‌شود. کاملاً واضح است که انجام یک شبیه‌سازی به‌منظور ارزیابی آزمون بدون در نظر گرفتن یک نقطه برش خودکار ناممکن است. بنابراین برای بررسی این موضوع با توجه به چگونگی تولید داده‌ها، در روش غیرمتناسب، فرایند جستجو در مرحله $\varepsilon \cdot n_i$ متوقف می‌شود زیرا این روش تضمین می‌کند که داده‌های دورافتاده در پایان جستجو وارد زیرمجموعه شوند. جدول ۳ فراوانی نسبی را برای ۵۰۰ بار شبیه‌سازی که در آن آزمون F پیشرو استوار در سطح معنی‌داری $\alpha = 0.05$ در ناحیه رد می‌افتد را نشان می‌دهد. با روش جستجوی پیشرو، احتمال پذیرش H_1 وقتی H_0 درست است همیشه کمتر از احتمال متناظر نمایش داده‌شده در جدول ۲ برای روش کلاسیک است. جدول ۴ مقدار پیراستگی را به میزان $\varepsilon\%$ انجام می‌دهد، بدین معنی که میزان پیراستگی را به‌صورت متغیر و متناسب با

جدول ۲. تقریب احتمال خطای نوع I آزمون F کلاسیک در حضور آلودگی

ε						
۱۰٪	۹٪	۸٪	۷٪	۶٪	۵٪	n_i
۰٫۰۹۰	۰٫۰۵۲	۰٫۰۶۸	۰٫۰۵۴	۰٫۰۵۲	۰٫۰۶۸	۲۰
۰٫۲۲۲	۰٫۱۲۴	۰٫۱۴۰	۰٫۰۵۰	۰٫۰۵۲	۰٫۰۷۸	۴۰
۰٫۳۷۶	۰٫۲۴۶	۰٫۲۰۲	۰٫۱۷۴	۰٫۱۱۴	۰٫۱۲۲	۶۰
۰٫۴۹۴	۰٫۴۴۰	۰٫۳۱۶	۰٫۲۲۰	۰٫۱۲۲	۰٫۱۳۰	۸۰
۰٫۶۵۸	۰٫۵۶۴	۰٫۴۴۰	۰٫۳۵۶	۰٫۲۶۲	۰٫۱۹۲	۱۰۰
۰٫۷۹۲	۰٫۵۹۴	۰٫۴۶۴	۰٫۴۰۴	۰٫۳۰۸	۰٫۲۳۶	۱۲۰
۰٫۸۶۰	۰٫۶۹۰	۰٫۶۱۶	۰٫۴۷۰	۰٫۳۶۶	۰٫۳۰۰	۱۴۰
۰٫۸۸۰	۰٫۸۰۴	۰٫۶۴۴	۰٫۵۲۲	۰٫۴۳۴	۰٫۲۹۲	۱۶۰
۰٫۹۴۸	۰٫۸۶۴	۰٫۷۳۶	۰٫۵۹۲	۰٫۴۴۶	۰٫۳۷۶	۱۸۰
۰٫۹۴۶	۰٫۹۰۰	۰٫۸۴۴	۰٫۷۱۲	۰٫۵۴۶	۰٫۴۰۴	۲۰۰

جدول ۳. تقریب احتمال خطای نوع I آزمون F پیشرو استوار (روش غیرمتناسب)

ε						
۱۰٪	۹٪	۸٪	۷٪	۶٪	۵٪	n_i
۰٫۰۶۰	۰٫۰۲۲	۰٫۰۶۴	۰٫۰۶۲	۰٫۰۷۸	۰٫۰۵۲	۲۰
۰٫۰۵۶	۰٫۰۷۰	۰٫۰۵۰	۰٫۰۵۰	۰٫۰۶۶	۰٫۰۴۶	۴۰
۰٫۰۸۰	۰٫۰۴۲	۰٫۰۹۰	۰٫۰۷۲	۰٫۰۵۴	۰٫۰۷۲	۶۰
۰٫۰۷۸	۰٫۰۷۰	۰٫۰۶۴	۰٫۰۸۲	۰٫۰۴۲	۰٫۰۵۰	۸۰
۰٫۰۸۲	۰٫۰۷۶	۰٫۰۷۰	۰٫۰۷۰	۰٫۰۶۰	۰٫۰۶۰	۱۰۰
۰٫۰۸۰	۰٫۰۸۴	۰٫۰۷۲	۰٫۰۶۲	۰٫۰۶۰	۰٫۰۵۲	۱۲۰
۰٫۰۷۶	۰٫۰۶۰	۰٫۰۷۲	۰٫۰۷۸	۰٫۰۵۲	۰٫۰۴۶	۱۴۰
۰٫۰۷۴	۰٫۰۹۰	۰٫۱۰۲	۰٫۰۷۲	۰٫۰۷۲	۰٫۰۶۴	۱۶۰
۰٫۰۸۸	۰٫۰۶۶	۰٫۰۷۸	۰٫۰۸۴	۰٫۰۵۴	۰٫۰۶۰	۱۸۰
۰٫۰۷۶	۰٫۰۶۸	۰٫۰۸۴	۰٫۰۶۶	۰٫۰۳۲	۰٫۰۷۴	۲۰۰

میزان آلودگی قرار داده‌ایم. همچنین نتایج جدول‌های ۵ و ۶ برای مقادیر ثابت پیراستگی به ترتیب ۱۵٪ و ۲۰٪ به دست آمده است. جداول ۴، ۵ و ۶ فراوانی نسبی را برای ۵۰۰ مرتبه شبیه‌سازی نشان می‌دهند که در آن آزمون F جایگشتی بر اساس میانگین پیراسته با مقدارهای به ترتیب ε ، ۱۵٪ و ۲۰٪ در سطح معنی‌داری $\alpha = ۰٫۰۵$ در ناحیه رد می‌افتد. با روش جایگشتی نیز همانند روش پیشرو، احتمال پذیرش H_1 وقتی H_0 درست است کمتر از احتمال متناظر نمایش داده شده در جدول ۲ برای روش کلاسیک است. نمودارهای شکل ۳ بیانگر عملکرد روش‌های جستجوی پیشرو و آزمون جایگشتی بر اساس پیراستگی‌های ε (میزان پیراستگی متناسب با میزان آلودگی)، ۱۵٪ و ۲۰٪ نسبت به روش کلاسیک در حضور درصد آلودگی‌های متفاوت ε هستند. می‌بینیم که در حضور آلودگی‌های ۵٪ و ۶٪، روش پیشرو و روش جایگشتی

جدول ۴. تقریب احتمال خطای نوع I آزمون F جایگشتی بر اساس میانگین پیراسته ε

ε						
۱۰%	۹%	۸%	۷%	۶%	۵%	n_i
۰٫۰۵۲	۰٫۰۵۸	۰٫۰۵۰	۰٫۰۵۴	۰٫۰۶۴	۰٫۰۶۴	۲۰
۰٫۱۰۰	۰٫۰۷۴	۰٫۰۷۶	۰٫۰۷۰	۰٫۰۶۲	۰٫۰۵۰	۴۰
۰٫۱۵۶	۰٫۱۵۸	۰٫۱۲۴	۰٫۱۱۶	۰٫۰۸۸	۰٫۰۷۲	۶۰
۰٫۲۰۸	۰٫۱۴۶	۰٫۱۲۸	۰٫۱۰۸	۰٫۱۱۴	۰٫۱۱۲	۸۰
۰٫۲۵۸	۰٫۲۴۴	۰٫۲۱۴	۰٫۱۳۶	۰٫۱۲۴	۰٫۱۲۴	۱۰۰
۰٫۳۳۴	۰٫۲۲۶	۰٫۲۱۰	۰٫۱۸۲	۰٫۱۲۰	۰٫۱۲۰	۱۲۰
۰٫۳۷۸	۰٫۳۴۰	۰٫۲۳۴	۰٫۲۱۰	۰٫۱۸۶	۰٫۱۵۰	۱۴۰
۰٫۴۲۶	۰٫۳۵۸	۰٫۲۷۸	۰٫۲۰۶	۰٫۱۶۲	۰٫۱۳۶	۱۶۰
۰٫۵۲۶	۰٫۴۰۸	۰٫۲۹۰	۰٫۲۴۰	۰٫۱۸۸	۰٫۱۵۲	۱۸۰
۰٫۵۵۰	۰٫۴۹۴	۰٫۳۹۰	۰٫۳۱۴	۰٫۲۱۴	۰٫۱۸۸	۲۰۰

جدول ۵. تقریب احتمال خطای نوع I آزمون F جایگشتی بر اساس میانگین پیراسته ۱۵%

ε						
۱۰%	۹%	۸%	۷%	۶%	۵%	n_i
۰٫۰۴۴	۰٫۰۵۰	۰٫۰۴۶	۰٫۰۵۴	۰٫۰۶۲	۰٫۰۶۰	۲۰
۰٫۰۸۴	۰٫۰۶۴	۰٫۰۶۰	۰٫۰۶۶	۰٫۰۵۶	۰٫۰۵۲	۴۰
۰٫۱۱۴	۰٫۱۱۲	۰٫۰۹۰	۰٫۰۹۲	۰٫۰۷۸	۰٫۰۶۴	۶۰
۰٫۱۶۶	۰٫۱۰۲	۰٫۰۹۲	۰٫۰۸۰	۰٫۰۸۶	۰٫۰۶۴	۸۰
۰٫۱۹۰	۰٫۱۷۰	۰٫۱۳۶	۰٫۰۸۴	۰٫۰۸۴	۰٫۰۹۸	۱۰۰
۰٫۲۲۶	۰٫۱۴۸	۰٫۱۵۶	۰٫۱۰۶	۰٫۰۹۲	۰٫۰۷۴	۱۲۰
۰٫۲۷۸	۰٫۲۱۸	۰٫۱۷۰	۰٫۱۶۴	۰٫۱۲۰	۰٫۰۸۸	۱۴۰
۰٫۳۰۴	۰٫۲۴۴	۰٫۱۷۲	۰٫۱۲۴	۰٫۰۸۸	۰٫۰۸۰	۱۶۰
۰٫۳۶۶	۰٫۲۷۲	۰٫۱۸۰	۰٫۱۴۲	۰٫۱۱۴	۰٫۱۰۰	۱۸۰
۰٫۴۰۰	۰٫۳۳۴	۰٫۵۶	۰٫۲۰۶	۰٫۱۳۶	۰٫۰۹۸	۲۰۰

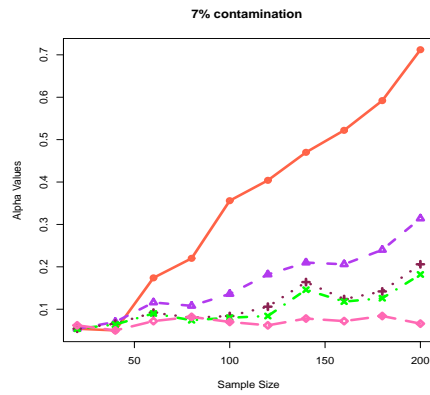
با هر سه میزان پیراستگی تا حجم ۶۰ تقریباً عملکرد یکسانی دارند. در حجم ۸۰ نیز غیر از روش جایگشتی با پیراستگی ε همگی عملکرد یکسانی دارند. اما از این مرحله به بعد هرچه اندازه نمونه افزایش می‌یابد روش پیشرو بهتر عمل می‌کند. روش‌های جایگشتی با پیراستگی ۱۵% و ۲۰% نیز پس از آن عملکرد یکسانی دارند. در حضور آلودگی بیشتر، روش جستجوی پیشرو نتیجه بهتری را ارائه می‌دهد. پس از آن در این دو نمودار می‌بینیم که هر چه اندازه نمونه بزرگ‌تر می‌شود روش جایگشتی با میانگین پیراسته ۲۰% عملکرد مناسب‌تری دارد.

جدول ۶. تقریب احتمال خطای نوع I آزمون F جایگشتی بر اساس میانگین پیراسته ۲۰٪

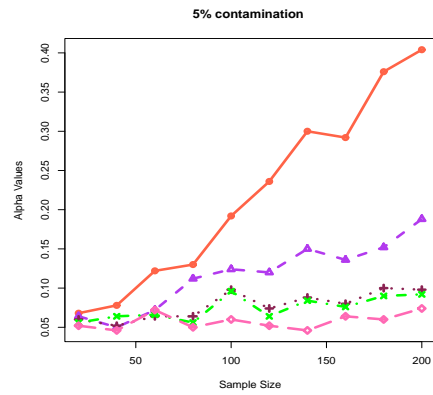
ε						
۱۰٪	۹٪	۸٪	۷٪	۶٪	۵٪	n_i
۰٫۰۳۸	۰٫۰۵۰	۰٫۰۴۴	۰٫۰۵۴	۰٫۰۵۲	۰٫۰۵۶	۲۰
۰٫۰۷۴	۰٫۰۶۲	۰٫۰۶۲	۰٫۰۶۴	۰٫۰۵۰	۰٫۰۶۴	۴۰
۰٫۰۹۸	۰٫۱۰۲	۰٫۰۸۶	۰٫۰۹۰	۰٫۰۷۸	۰٫۰۶۶	۶۰
۰٫۱۴۴	۰٫۰۹۴	۰٫۰۸۴	۰٫۰۷۴	۰٫۰۸۰	۰٫۰۵۶	۸۰
۰٫۱۶۸	۰٫۱۵۲	۰٫۱۲۰	۰٫۰۸۰	۰٫۰۸۴	۰٫۰۹۶	۱۰۰
۰٫۲۰۲	۰٫۱۲۲	۰٫۱۳۶	۰٫۰۸۴	۰٫۰۸۶	۰٫۰۶۴	۱۲۰
۰٫۲۳۶	۰٫۱۸۸	۰٫۱۴۶	۰٫۱۴۶	۰٫۱۰۰	۰٫۰۸۴	۱۴۰
۰٫۲۵۰	۰٫۲۰۶	۰٫۱۴۶	۰٫۱۱۸	۰٫۰۹۰	۰٫۰۷۶	۱۶۰
۰٫۲۹۶	۰٫۲۳۲	۰٫۱۶۰	۰٫۱۲۶	۰٫۱۰۶	۰٫۰۹۰	۱۸۰
۰٫۳۳۸	۰٫۲۸۸	۰٫۲۰۸	۰٫۱۸۲	۰٫۱۱۸	۰٫۰۹۲	۲۰۰

۲.۵ مقایسه روش‌ها بر اساس توان آزمون

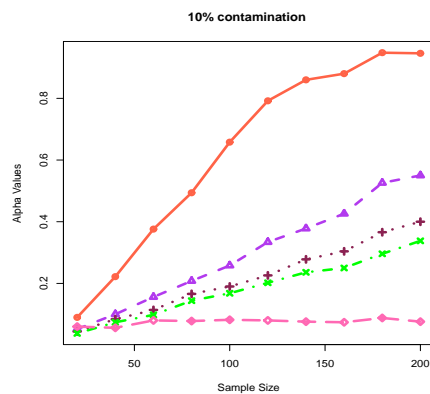
فرض کنید مجموعه داده‌ای با سه گروه متعادل ($n_i = 20, 40, \dots, 200; i = 1, 2, 3$) از اندازه نمونه افزایشی n_i مشاهده داریم. مشاهدات گروه اول و سوم از توزیع نرمال استاندارد تولید می‌شوند ولی قسمت اعظم مشاهدات گروه دوم از توزیع $N(4, 1)$ و بقیه مشاهدات توسط نرخ افزایشی $\varepsilon\%$ ($\varepsilon = 0.05, \dots, 0.10$) از توزیع $N(-46, 1)$ آلوده شده است. بنابراین توزیع گروه دوم به صورت: $(1 - \varepsilon)N(4, 1) + \varepsilon N(-46, 1)$ است. جدول ۷ فراوانی‌های نسبی $r.f.(n_i, \varepsilon)$ را برای ۵۰۰ بار شبیه‌سازی که در آن آزمون F کلاسیک در سطح معنی‌داری $\alpha = 0.05$ به درستی در ناحیه رد می‌افتد را نشان می‌دهد. به عبارتی، توان آزمون را برای روش کلاسیک نمایش می‌دهد. مشاهده می‌شود که توان آزمون در حضور آلودگی ۵٪ و در نمونه‌های بزرگ‌تر نتایج بهتری را نشان می‌دهد و در بقیه حالات صفر است، که این نشان از ناکارآمدی و مشکل روش کلاسیک در حضور آلودگی است. روش جستجوی پیشرو و جایگشتی نیز مشابه روش کلاسیک در ۵۰۰ بار شبیه‌سازی در سطح معنی‌داری $\alpha = 0.05$ مورد ارزیابی قرار گرفت. ملاحظه می‌شود که تمام مقادیر توان آزمون در سطوح مختلف آلودگی و اندازه‌های نمونه متفاوت برای این دو روش، نتیجه ایده‌آل ۱۰۰ درصد را برای توان آزمون نشان می‌دهند. نمودارهای شکل ۴ به وضوح عملکرد یکسان روش جستجوی پیشرو و روش جایگشتی با پیراستگی‌های $\varepsilon\%$ ، ۱۵٪ و ۲۰٪ را نشان می‌دهد که در حضور آلودگی‌های ε درصد ($\varepsilon = 0.05, 0.06, \dots, 0.10$) و همچنین اندازه نمونه‌های ۲۰، ۴۰، ...، ۲۰۰ عملکرد یکسان و خوبی را ارائه داده‌اند. همچنین مشکل روش کلاسیک در حضور آلودگی را به وضوح نشان می‌دهد.



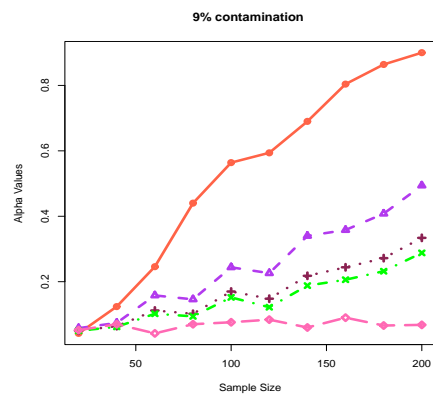
آلودگی ۷٪



آلودگی ۵٪

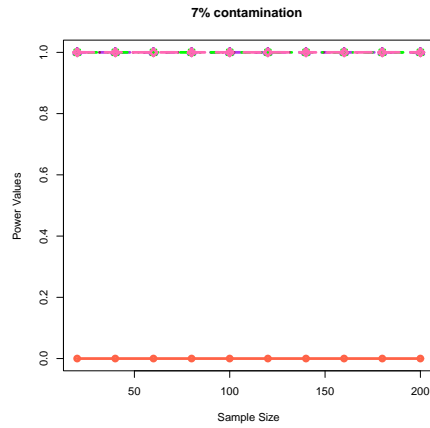


آلودگی ۱۰٪

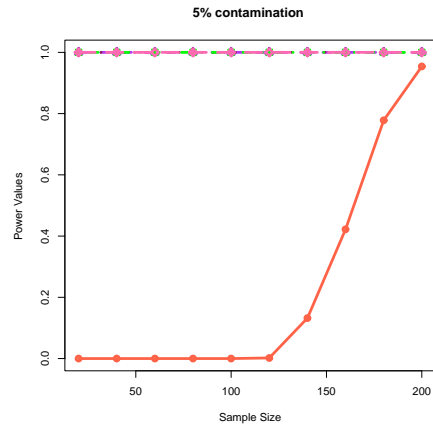


آلودگی ۹٪

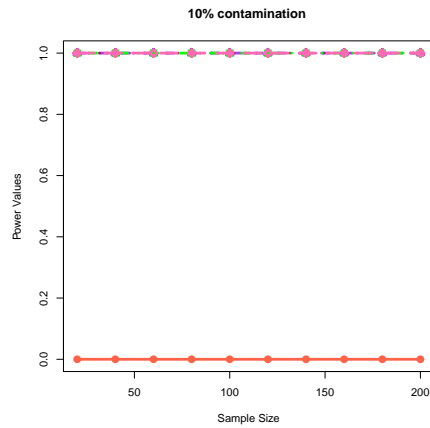
شکل ۳. نمودارهای میانگین نرخ خطای نوع اول در روش‌های کلاسیک، جستجوی پیشرو و جایگشتی (خطوط با علامت دایره برای روش کلاسیک، علامت مثلث جایگشتی با پیراستگی ۵٪، علامت بعلاوه پیراستگی ۱۵٪، علامت ضربدر پیراستگی ۲۰٪ و لوزی جستجوی پیشرو است) در حضور آلودگی‌های ۰/۰۵، ۰/۰۷، ۰/۰۹، ۰/۱۰.



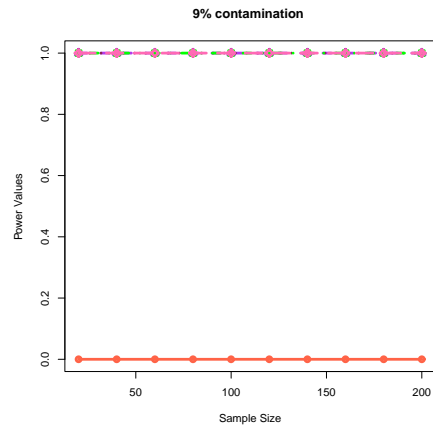
آلودگی ۷٪



آلودگی ۵٪



آلودگی ۱۰٪



آلودگی ۹٪

شکل ۰۴. نمودارهای میانگین توان آزمون (خطوط با علامت دایره برای روش کلاسیک، علامت مثلث جایگشتی با پیراستگی ϵ ، علامت بعلاوه پیراستگی ۱۵٪، علامت ضریب پیراستگی ۲۰٪ و لوزی جستجوی پیشرو است که به غیر از کلاسیک همگی عملکرد یکسانی دارند) در حضور آلودگی‌های ۰/۱۰، ۰/۰۹، ۰/۰۷، ۰/۰۵.

جدول ۷. تقریب توان آزمون F کلاسیک در حضور آلودگی

ε						
۱۰%	۹%	۸%	۷%	۶%	۵%	n_i
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۲۰
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۴۰
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۶۰
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۸۰
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۱۰۰
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۱۲۰
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۱۳	۱۴۰
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۴۲	۱۶۰
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۷۸	۱۸۰
۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۰۰	۰٫۹۵	۲۰۰

۳.۵ مقایسه روش‌های جستجوی پیشرو و آزمون جایگشتی

در این بخش مدت‌زمان اجرای دو روش جستجوی پیشرو و جایگشتی، تحت شرایط یکسان در تولید اعداد شبیه‌سازی در نرم‌افزار R مقایسه می‌شوند. برای بررسی این موضوع از مدل شبیه‌سازی مربوط به خطای نوع اول استفاده شده است. جدول ۸ برآورد تقریبی مدت‌زمان اجرای روش جستجوی پیشرو را بر اساس

جدول ۸. تقریب مدت‌زمان اجرای آزمون F پیشرو (روش غیرمتناسب)

ε						
۱۰%	۹%	۸%	۷%	۶%	۵%	n_i
۰٫۳۳	۰٫۳۳	۰٫۳۲	۰٫۳۲	۰٫۳۲	۰٫۳۲	۲۰
۰٫۷۱	۰٫۷۱	۰٫۷۰	۰٫۷۰	۰٫۷۱	۰٫۷۲	۴۰
۱٫۱۲	۱٫۱۲	۱٫۱۲	۱٫۱۲	۱٫۱۳	۱٫۱۲	۶۰
۱٫۵۸	۱٫۵۸	۱٫۵۷	۱٫۵۹	۱٫۵۹	۱٫۶۰	۸۰
۲٫۰۸	۲٫۰۸	۲٫۰۶	۲٫۰۸	۲٫۰۱	۲٫۱۳	۱۰۰
۲٫۶۱	۲٫۶۲	۲٫۶۲	۲٫۶۲	۲٫۶۲	۲٫۶۸	۱۲۰
۳٫۱۷	۳٫۱۴	۳٫۱۸	۳٫۱۹	۳٫۱۹	۳٫۲۱	۱۴۰
۳٫۷۶	۳٫۷۸	۳٫۸۰	۳٫۸۰	۳٫۸۴	۳٫۸۱	۱۶۰
۴٫۴۴	۴٫۳۹	۴٫۴۴	۴٫۴۹	۴٫۴۹	۴٫۴۷	۱۸۰
۵٫۱۷	۵٫۱۱	۵٫۲۱	۵٫۱۷	۵٫۱۸	۵٫۲۲	۲۰۰

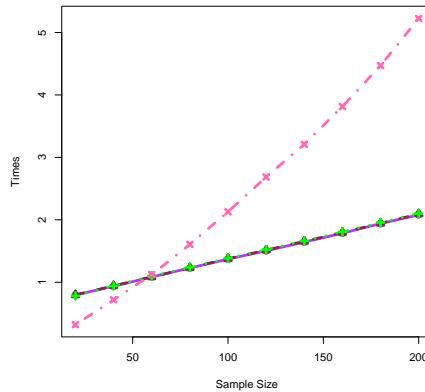
آلودگی‌ها و اندازه نمونه‌های مختلف نمایش می‌دهد. همان‌طور که انتظار می‌رود میزان آلودگی هیچ تأثیری روی مدت‌زمان اجرا نمی‌تواند داشته باشد اما اندازه نمونه‌های مختلف، زمان‌های متفاوتی را نشان می‌دهد. برای مثال، در ستون مربوط به آلودگی ۵٪، برای حجم نمونه ۲۰، دارای میانگین مدت‌زمان اجرای $۰٫۳۲ \times ۸۰$

ثانیه است. هر چه اندازه نمونه افزایش می‌یابد متناسب با آن مدت‌زمان اجرای آن نیز افزایش می‌یابد، به طوری که در حجم نمونه ۲۰۰، مدت‌زمان آن به ۵۲۲۴۶۰ ثانیه می‌رسد. برای انجام ۵۰۰ بار شبیه‌سازی، با حجم نمونه ۱۲۰ حدوداً ۲۲،۴ دقیقه طول می‌کشد تا مقدار موردنظر حاصل شود و برای اندازه نمونه ۲۰۰ به ۴۳،۵۴ دقیقه زمان نیاز است. جدول ۹ برآورد تقریبی مدت‌زمان اجرای روش آزمون جایگشتی را بر اساس آلودگی‌ها و اندازه نمونه‌های مختلف نمایش می‌دهد. هر چه اندازه نمونه افزایش می‌یابد متناسب با آن مدت‌زمان اجرای آن نیز افزایش می‌یابد، به طوری که در حجم ۲۰۰ مدت‌زمان آن به ۲۰۷۹۸۲ ثانیه می‌رسد، که در مقایسه با روش پیشرو کمتر است.

جدول ۹. تقریب مدت‌زمان اجرای آزمون F جایگشتی

ε						
۱۰٪	۹٪	۸٪	۷٪	۶٪	۵٪	n_i
۰٫۸۰	۰٫۸۰	۰٫۸۰	۰٫۸۰	۰٫۸۰	۰٫۸۰	۲۰
۰٫۹۵	۰٫۹۴	۰٫۹۴	۰٫۹۴	۰٫۹۴	۰٫۹۳	۴۰
۱٫۱۰	۱٫۰۹	۱٫۰۸	۱٫۰۹	۱٫۰۸	۱٫۰۸	۶۰
۱٫۲۳	۱٫۲۳	۱٫۲۲	۱٫۲۳	۱٫۲۲	۱٫۲۳	۸۰
۱٫۳۸	۱٫۳۷	۱٫۳۷	۱٫۳۷	۱٫۳۷	۱٫۳۷	۱۰۰
۱٫۵۲	۱٫۵۱	۱٫۵۰	۱٫۵۱	۱٫۵۰	۱٫۵۰	۱۲۰
۱٫۶۶	۱٫۶۵	۱٫۶۴	۱٫۶۵	۱٫۶۴	۱٫۶۴	۱۴۰
۱٫۸۱	۱٫۷۹	۱٫۷۹	۱٫۷۹	۱٫۷۸	۱٫۷۸	۱۶۰
۱٫۹۶	۱٫۹۴	۱٫۹۴	۱٫۹۵	۱٫۹۴	۱٫۹۴	۱۸۰
۲٫۱۰	۲٫۰۹	۲٫۰۸	۲٫۰۹	۲٫۰۹	۲٫۰۸	۲۰۰

برای حجم نمونه ۱۲۰ مدت‌زمان آن ۱۲،۵۴ دقیقه به طول می‌انجامد. برای حجم نمونه ۲۰۰ تقریباً ۱۷،۳۳ دقیقه طول می‌کشد تا نتیجه موردنیاز حاصل شود، که نسبت به روش پیشرو کمتر است. نمودار شکل ۵ زمان اجرای برنامه روش‌های پیشرو و جایگشتی را نشان می‌دهد. همان‌طور که نتایج شبیه‌سازی در جداول ۸ و ۹ نشان می‌دهد و به آن اشاره شد، کیفیت داده‌ها و میزان آلودگی آن‌ها در زمان محاسبه روش‌های موردبحث تأثیری ندارد. ملاحظه می‌شود که روش جایگشتی با پیراستگی‌های متفاوت سرعت یکسانی را نمایش می‌دهند و همگی در یک راستا قرار دارند. همچنین زمان اجرا برای اندازه نمونه‌های ۲۰ و ۴۰ در روش جستجوی پیشرو نسبت به روش جایگشتی عملکرد بهتری را نشان می‌دهد. در اندازه نمونه ۶۰ تقریباً سرعت پردازش یکسانی دارند، اما از این مرحله به بعد با افزایش اندازه نمونه مدت‌زمان اجرا در روش جایگشتی خیلی کمتر از روش پیشرو است. به طوری که از اندازه نمونه ۱۲۰ به بالاتر زمان اجرا برای روش پیشرو به دو برابر می‌رسد و هر چه اندازه نمونه افزایش می‌یابد این فاصله نیز بیشتر می‌شود.



شکل ۵. نمودار مدت زمان اجرای روش‌ها (خطوط با علامت دایره برای روش جایگشتی با پیراستگی ۵٪، علامت مثلث پیراستگی ۱۵٪، علامت بعلاوه پیراستگی ۲۰٪ و ضربدر جستجوی پیشرو است که به غیر از روش جستجوی پیشرو همگی عملکرد یکسانی دارند)

۴.۵ تحلیل داده‌های واقعی

در این بخش با ارائه یک مجموعه داده واقعی، به ارزیابی عملکرد روش‌های جستجوی پیشرو و توزیع جایگشتی آماره F بر اساس میانگین پیراسته پرداخته می‌شود. داده‌ها مربوط به تعداد شاخه پروفیل تولیدی برای چارچوب^{۱۱} پنجره برحسب شاخه‌های ۶ متری از کارخانه گروه صنعتی آریا در سال ۱۳۹۰ است. علاقه‌مند به بررسی صحت یا سقم ادعای تساوی میانگین تعداد شاخه‌های پروفیل تولیدشده در چهارفصل سال هستیم. تعداد شاخه‌های تولیدی در فصول مختلف به ترتیب از ۵۵، ۸۲، ۸۹ و ۸۸ مشاهده تشکیل شده است.

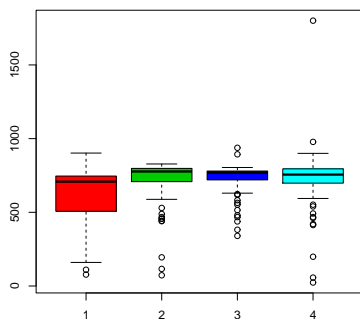
نتایج آزمون کلاسیک ANOVA در جدول ۱۰ نشان داده شده است. با توجه به مقدار F و p -مقدار به دست آمده، فرضیه صفر رد می‌شود، یعنی میانگین تعداد پروفیل تولیدشده در ۴ فصل سال متفاوت است.

جدول ۱۰. جدول تحلیل واریانس تک عاملی روی داده‌های تولید پروفیل در ۴ فصل سال ۱۳۹۰

منبع تغییر	مجموع مربعات	میانگین مربعات	درجه آزادی	F	p -مقدار
بین تیمارها	۴۹۶۸۹۰	۱۶۵۶۳۰	۳	۶۳۸۶	۰٫۰۰۰۰۳۲۸
خطا (درون تیمارها)	۸۰۴۰۷۱۲	۲۵۹۳۸	۳۱۰		

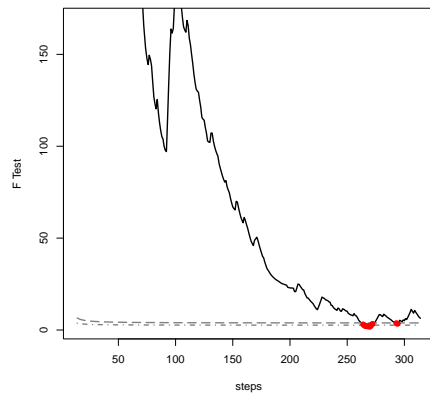
¹¹Frame

نمودار جعبه‌ای در شکل ۶ برای چهار فصل حضور مشاهداتی را در همه گروه‌ها نمایش می‌دهد که به نظر می‌رسد از بخش عمده داده‌ها تفاوت دارند و می‌توانند داده دورافتاده در نظر گرفته شوند. حال روش استوار جستجوی پیشرو روی این مجموعه داده پیاده‌سازی می‌شود. برآورد ضرایب در هر مرحله، روی نمودار (الف) شکل ۷ نشان داده شده است. در مراحل اولیه، مقادیر مربوط به فصل‌های تابستان، پاییز و زمستان نسبت به فصل بهار به هم‌دیگر نزدیک‌تر هستند، اما در ادامه این فاصله کمتر می‌شود. تقریباً از مرحله ۲۱۰ به بعد ضرایب این سه فصل یک‌روند نزولی را پیش می‌گیرند تا اینکه تقریباً در مراحل ۲۶۵ تا ۲۷۰ به نزدیک‌ترین مقادیر خود با μ_1 می‌رسند. از این مرحله به بعد ضریب μ_1 شدیداً کاهش می‌یابد. این کاهش شدید و همچنین ثابت نبودن ضرایب از مرحله‌ای به بعد، گواه بر وجود داده‌های دورافتاده در مجموعه داده‌ها هستند.

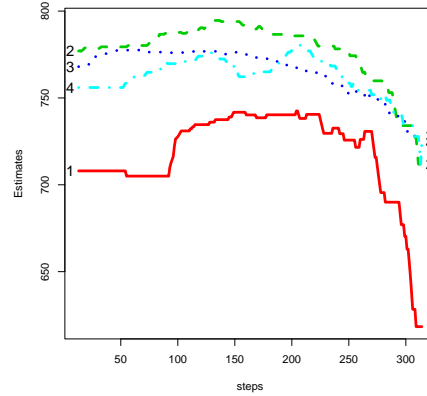


شکل ۶. نمودار جعبه‌ای تعداد شاخه‌های پروفیل تولیدی در طی فصول سال ۱۳۹۰

آماره F در مراحل ۲۶۴ تا ۲۷۰ در سطح معنی‌داری ۵٪ و در مراحل ۲۶۳ تا ۲۷۱ در سطح معنی‌داری ۱٪ در ناحیه پذیرش افتاده است. از این مرحله به بعد به دلیل حضور داده‌های دورافتاده از ناحیه پذیرش خارج می‌شود و دوباره در مراحل ۲۹۲ و ۲۹۳ در سطح معنی‌داری ۱٪ در ناحیه پذیرش می‌افتد، سپس در ادامه دوباره از ناحیه پذیرش خارج می‌شود. حال نتایج روش آزمون جایگشتی آماره SSX بر اساس میانگین پیراسته مورد بررسی قرار می‌گیرد. جدول ۱۱ نتایج روش توزیع جایگشتی را با میزان پیراستگی‌های (۰٪، ۵٪، ۱۰٪، ۱۵٪ و ۲۰٪) نمایش می‌دهد. همان‌طور که ملاحظه می‌شود با در نظر گرفتن پیراستگی ۵٪ به بالا، p -مقدار در ناحیه پذیرش واقع می‌شود.



(ب)



(الف)

شکل ۷. (الف) نمودارهای پیشرو ضرایب $\hat{\mu}_1$ تا $\hat{\mu}_4$ ، به ترتیب با خطوط شماره ۱ تا ۴ و (ب) آزمون F (خطوط خط چین، ناحیه پذیرش در سطوح معنی‌داری ۱ درصد و ۵ درصد هستند)

جدول ۱۱. جدول p -مقدارهای آزمون جایگشتی بر اساس ۵ مقدار مختلف پیراستگی

۰/۲۰	۰/۱۵	۰/۱۰	۰/۰۵	۰/۰۰	میزان پیراستگی
۰/۸۷۴	۰/۸۷۳	۰/۸۹۰	۰/۶۹۲	۰/۵۰۰	p -مقدار

بحث و نتیجه‌گیری

حضور مشاهدات دورافتاده در بین داده‌ها می‌تواند منجر به نتایج اشتباه در مدل‌سازی و آزمون فرض شود. همان‌طور که نشان داده شد استنباط در تحلیل واریانس کلاسیک نیز در حضور این مشاهدات می‌تواند گمراه‌کننده باشد. در این مقاله نشان داده شد که روش استوار تحلیل واریانس بر اساس توزیع جایگشتی تابعی از میانگین پیراسته دارای کارایی قابل‌رقابتی با رهیافت مبتنی بر جستجوی پیشرو است. خصوصاً علاوه بر اینکه روش پیشنهادی از سرعت بالاتری در حجم نمونه‌های بالا برخوردار است برخلاف رهیافت مبتنی بر جستجوی پیشرو به فرض نرمال بودن داده‌ها نیاز ندارد. این خصوصیت توزیع آزاد آن باعث می‌شود بتوان از آن برای تحلیل داده‌هایی که دارای توزیع‌های غیر نرمال با دم‌های کلفت هستند نیز استفاده کرد.

تقدیر و تشکر

نویسندگان مقاله از نظرات و پیشنهادات داوران، هیئت تحریریه و ویراستار مجله علوم آماری که باعث ارتقای مقاله شد، کمال تشکر و قدردانی را دارند.

مراجع

- Bertaccini B. and Varriale R., (2007), Robust ANalysis Of VAriance: An Approach based on the Forward Search, *Computational Statistics and Data Analysis*, **51**, 5172-5183.
- Box G. E. P. and Andersen S. L., (1955), Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumption (with discussion), *Royal Statistical Society*, **17**, 1-34.
- Butar B. F., (2008), Permutation Tests for Comparing more than Two Samples, *Mathematical Sciences and Mathematics Education*, **2**, 20-29.
- Fisher R. A., (1918), The Correlation Between Relatives on the Supposition of Mendelian inheritance, *Transactions of the Royal Society of Edinburgh*, **52**, 399-433.
- Fisher R. A. and Mackenzie W. A., (1923), Studies in Crop Variation. II. The Manurial Response of Different Potato Varieties, *Transactions of the Royal Society of Edinburgh*, **13**, 311-320.
- Fisher R. A., (1925), *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.
- Fisher R. A., (1935), *The Design of Experiments*, 1st edition (8th edition, 1966), Oliver and Boyd, Edinburgh.
- Fisher R. A., (1936), "The Coefficient of Racial Likeness" and the Future of Craniometry, *Royal Anthropological Institute of Great Britain and Ireland*, **66**, 57-63.
- Hawkins D. M., (1980), *Identification of Outliers*, Chapman and Hall, London.
- Higgins J. J., (2004), *An Introduction to Modern Nonparametric Statistics*, Brooks/Cole.
- Hoeffding W., (1952), The Large-Sample Power of Tests based on Permutations of Observations, *Annals of Mathematical Statistics*, **23**, 169-192.

- Lehmann E. L. and Stein C. M., (1949), On the Theory of Some Non-parametric Hypotheses, *Annals of Mathematical Statistics*, **20**, 28-45.
- Good P. I., (2000), *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, 2nd ed. Springer, New York.
- Mendeleev D. I., (1895), Course of Work on the Renewal of Prototypes or Standard Measures of Lengths and Weights (in russian), *Vremennik Glavnoi Palaty Mer i Vesov*, **2**, 157-185.
- Montgomery D. C., (2012), *Design and Analysis of Experiments*, 8th Edition, John Wiley and Sons, Incorporated.
- Pitman E. J. G., (1937), Significance Tests Which May be Applied to Samples from any Populations, *Royal Statistical Society*, **4**, 119-130.
- Pitman E. J. G., (1938), Significance Tests which may be Applied to Samples from any Populations III. The Analysis of Variance Test, *Biometrika*, **29**, 322-335.
- Stigler S. M., (1976), The Anonymous Professor Gergonne, *Historia Mathematica*, **3**, 71-74.
- Tukey J. W., (1962), The Future of Data Analysis, *Annals of Mathematical Statistics*, **33**, 1-67.