مجله تحقیق در عملیات در کاربردهای آن، سال سیزدهم، شماره دوم (پیاپی ۴۹)، تابستان ۹۵، ص ص ۱۲۱-۱۱۳
شاپا ۷۲۸۶-۲۲۵۱

دو آزمون آماری برای شناسایی نقاط پرت در اندازه‌گیری عملکرد ناپارامتری

محسن خون سیاوش*[۱]، رضا کاظمی متین[۲]، مرتضی خدابین[۲]
۱- استادیار، دانشگاه آزاد اسلامی واحد قزوین، گروه علوم پایه، قزوین، ایران
۲- دانشیار، دانشگاه آزاد اسلامی واحد کرج، گروه ریاضی کاربردی، کرج، ایران

رسید مقاله: ۹ دی ۱۳۹۴
پذیرش مقاله: ۵ خرداد ۱۳۹۵

چکیده
در تحلیل پوششی داده‌ها که از مقایسه‌ی عملکرد نسبی یک واحد در قیاس با مجموعه‌ی مرجع برای تشخیص ناکارایی نسبی و ارایه‌ی الگوی بهبود استفاده می‌شود، تشخیص درست واحدهای پرت برای دستیابی به نتایج دقیق بسیار مهم است. در این نوع مدل‌های مرز قطعی، امروزه به‌کارگیری آزمون‌های آماری در تشخیص داده‌های پرت بسیار مرسوم هستند. این مقاله به معرفی دو آزمون آماری برای تشخیص نقاط پرت در تحلیل پوششی داده‌ها می‌پردازد. در هر دو روش ارایه شده، هر مشاهده یک‌بار از نمونه حذف‌شده و نتایج مدل‌های برآورد کارایی مربوط به حذف این واحد برای تولید به توزیع برآورد کارایی قبل و بعد از حذف استفاده می‌شود. بر اساس توزیع به‌دست‌آمده، دو آزمون آماری طراحی و معرفی می‌شوند تا نقاط پرت بالقوه را شناسایی کنند. نتایج اجرای این روش را از طریق مجموعه‌ای از داده‌های واقعی نشان داده‌ایم. درمجموع، روش معرفی‌شده می‌تواند در اولین گام قبل از استفاده از هر برآورد مرزی جهت تشخیص و حذف داده‌های پرت استفاده شود.

کلمات کلیدی: تحلیل پوششی داده‌ها، نقاط پرت، کارایی، آزمون نابرابری چبیشف، آماره‌ی کوک.

۱ مقدمه

تحلیل پوششی داده‌ها (DEA) توسط چارنز و همکاران [۱] به‌عنوان یک روش نا پارامتری قطعی در برآورد تابع تولید معرفی شد. در اجرا، این روش به خاطر تکیه بر چند فرض ساده بسیار جذاب بود؛ اما با این ساختار، نتایج حاصل از مدل‌های DEA در ارزیابی کارایی بسیار حساس به مقادیر دورترین نقاط یا مشاهدات پرت هستند. چون‌که مشاهدات بالاتر معمولاً مرز تولید DEA را تعیین می‌کنند، برآورد مرز کارایی ممکن است حساس به خطای اندازه‌گیری در داده‌های نمونه باشد. در نتیجه‌ی آلوده شدن یک مشاهده به خطا، خروجی ممکن است افرایش و یا ورودی کاهش یابد وآن واحد کارا تلقی شود. درنتیجه امکان ورود این واحد به مجموعه‌های مرجع

* عهده‌دار مکاتبات
آدرس الکترونیکی:mfsiavash@gmail.com

[DOR: 20.1001.1.22517286.2016.13.2.8.8]

دیگر واحدها و به دنبال آن تغییر نتایج کارایی محاسبه‌شده وجود دارد؛ بنابراین، تشخیص نقاط پرت به‌عنوان اولین گام در برآورد DEA از تابع تولید مهم است.

اساساً، همواره مشکلاتی با داده‌های تجربی وجود دارد؛ زیرا که برخی از واحدهای تصمیم‌گیرنده (DMUs) نقاط پرت هستند و یا به مجموعه داده‌ها تعلق ندارند. واضح است که گام اول پیش از ارزیابی عملکرد نسبی این واحدها می‌بایست معطوف به بررسی و تشخیص این واحدهای مشاهده‌شده‌ی با عملکرد بسیار خوب و یا بسیار بد باشد. رویکرد پیشنهادی ما به‌کارگیری مدل‌های DEA و تشخیص واحدهای ابر–کارا است که می‌توانند در اندازه‌گیری کارایی سایر واحدها و تولید وزن‌های مطلوب محاسبه‌ی کارایی بسیار تاثیرگذار باشند. این واحدها می‌توانند به‌عنوان نقاط پرت بالقوه در نظر گرفته شوند.

ادامه‌ی مطالب این مقاله به شرح زیر است: در بخش ۲ مروری بر مقالات و خلاصه‌ای از روش‌ها در مورد تشخیص نقاط پرت در ادبیات DEA خواهیم داشت. بخش ۳ به معرفی روش جدید پیشنهادی با استفاده از دو آزمون آماری در شناسایی نقاط پرت ابر-کارا اختصاص داده‌شده است. یک کاربرد تجربی مربوط به ۴۲ واحد آموزشی در دانشگاه آزاد اسلامی، واحد کرج در بخش ۴ جهت اجرای روش‌های پیشنهادی و تشخیص داده‌های پرت تشریح شده است. بخش ۵ نتیجه‌گیری مباحث ارایه می‌گردد.

۲ تشخیص نقاط پرت در DEA

مطالعات بسیاری برای اندازه‌گیری حساسیت و استحکام نتایج DEA در حضور خطاهای داده با استفاده از انواع روش‌ها برای شناسایی نقاط پرت در ادبیات DEA انجام‌شده است. در اینجا ما به بررسی مختصری از آثاری که در این زمینه بیشتر تاثیرگذار هستند می‌پردازیم.

بنکر و گیفورد [۲] استفاده از مدل ابر کارایی برای شناسایی مشاهدات با خطاهای داده را پیشنهاد دادند و برآورد کارایی قابل‌اطمینانی را بعد از شناسایی نقاط پرت به دست آوردند . بنکر و داتار [۳] این روش را برای شناسایی نقاط پرت در تجزیه‌وتحلیل کارایی هزینه‌ی ۱۱۷ بیمارستان به کاربردند. ویلسون [۴] و [۵]، با استفاده از توابع موثر، روش‌هایی برای تشخیص نقاط پرت در چارچوب DEA معرفی نمودند اما روش پیشنهادی ایشان از حیث حجم محاسباتی به دلیل افزایش تعداد مشاهدات پرهزینه است. اوندریچ و روگریو [۶] به تجزیه‌وتحلیل با استفاده از روش نمونه‌گیری مجدد به کمک فن جکنایف (Jackknifing) و توانایی آن در تشخیص نقاط پرت در مدل‌های DEA پرداختند. سیمار [۷] استفاده از یک روش آماری مرزی m مرتبه‌ای را شرح داد که توسط کازالز و همکاران [۸] برای تشخیص نقاط پرت معرفی‌شده بود. بنکر و چانگ [۹] روشی مبتنی بر ابر-کارایی ارایه کردند که نقاط پرت در برآورد کارایی مدل‌های DEA مشخص و حذف می‌شد. جانسون و مک گینیس [۱۰] یک روش شناسایی برای نقاط پرت با استفاده از هر دو مرز کارا و ناکارا معرفی کردند. چن و جانسون [۱۱] یک مدل متحد و یکپارچه‌ای را در تشخیص هر دو نوع نقاط پرت کارا و ناکارا توسعه بخشیدند. اخیراً، بلینی [۱۲] نیز یک روش جستجویی برای کشف نقاط پرت با یکپارچه‌سازی رگرسیون خطی در چارچوب DEA و ابر-کارایی معرفی کرد. بیشتر این فعالیت‌های علمی اشاره‌شده از آزمون فرض آماری در تشخیص نقاط پرت

[DOR: 20.1001.1.22517286.2016.13.2.8.8]

بهره‌مند شده‌اند. توجه کنید که در بحث DEA یک واحد پرت به‌طور مستقیم توزیع نمرات کارایی و ارزیابی عملکرد واحد مشاهده‌شده را تحت تاثیر قرار خواهد داد؛ بنابراین، یکی از راه‌های تشخیص نقاط پرت بالقوه این است که محکی برای میزان بزرگی این تاثیر ارایه کنیم. در این مقاله از مدل ابر-کارایی DEA برای دستیابی به توزیع کارایی بعد از حذف یک واحد کارا و مقایسه‌ی آن با توزیع کارایی قبل از حذف استفاده خواهیم نمود. از برخی آزمون‌های آماری در دسترس برای تعیین میزان بزرگی تاثیر اشاره‌شده استفاده خواهد شد. در بخش بعدی، دو روش پیشنهادی ما برای شناسایی نقاط پرت کارا بر اساس دو روش آماری بیان‌شده است.

۳ تشخیص نقاط پرت با استفاده از آزمون‌های آماری

با استفاده از نمادهای رایج در DEA، فرض کنید که n واحد تصمیم‌گیرنده داریم، واحد jام یا DMU_j ورودی‌های x_{ij} برای $i = ۱,...,m$ را مصرف می‌کند تا خروجی‌های y_{rj} که $r = ۱,...,s$ و $j = ۱,...,n$ را تولید کند . هم‌چنین فرض کنید که $(\mathbf{x}_j, \mathbf{y}_j) \in \Re^{m \times s}$ بردار ورودی و خروجی واحد jام را مشخص کند.
در مدل کلاسیک CCR- که توسط چارنز، کوپر و رودز [۱] معرفی شد، تحت فرض بازه مقیاس ثابت (CRS) و مجموعه امکان تولید (PPS) تعریف‌شده و به‌صورت $T_c = \left\{ (\mathbf{x}, \mathbf{y}) \middle| \mathbf{x} \geq \sum_{j=۱}^{n} \mathbf{x}_j \lambda_j ; \mathbf{y} \leq \sum_{j=۱}^{n} \mathbf{y}_j \lambda_j ; \boldsymbol{\lambda} \geq ۰ \right\}$ در نظر گرفته می‌شود.

در این مطالعه، از اندازه کارایی ورودی فارل تعریف‌شده به‌صورت $Eff(\mathbf{x}_0, \mathbf{y}_0) = \min\{\theta | (\theta \mathbf{x}_0, \mathbf{y}_0) \in T_c\}$ استفاده می‌کنیم که در آن بردار $(\mathbf{x}_0, \mathbf{y}_0)$ واحد مشاهده‌شده‌ی تحت ارزیابی را نشان می‌دهد. کارایی فارل ورودی محور برای هر یک از مشاهدات محاسبه می‌شود:

$$
\begin{aligned}
& e_o = \min_{\theta, \lambda} \quad \theta \\
& s.t. \\
& \sum_{j=۱}^{n} x_{ij} \lambda_j^o \leq \theta x_{io}, \quad i = ۱,...,m, \qquad (۱) \\
& \sum_{j=۱}^{n} y_{rj} \lambda_j^o \geq y_{ro}, \quad r = ۱,...,s, \\
& \lambda_j^o \geq ۰, \qquad j = ۱,...,n.
\end{aligned}
$$

این LP یک‌بار برای هر مشاهده $(o = ۱,...,n)$ محاسبه می‌شود تا نمرات کارایی برای تمام مشاهدات محاسبه شود. حل مدل (۱) یک اندازه کارایی نا پارامتری e_j برای $j = ۱,...,n$ در محدوده‌ی $۰ < e_j \leq ۱$ را فراهم می‌کند.

در این مرحله از یک روش تکراری به‌منظور برآورد تفاوت ایجادشده مربوط به حذف هر مشاهده در نمره کارایی واحدهای دیگر، در تلاش برای شناسایی نقاط پرت استفاده می‌کنیم. برای این هدف، واحد تحت ارزیابی

[DOR: 20.1001.1.22517286.2016.13.2.8.8]

از نمونه‌ها حذف‌شده و برنامه‌ی خطی برای دیگر مشاهدات حل می‌شود تا اثر DMU_o بر مرز کارایی مجموعه‌ی تولید مشخص گردد. برای واحد k که $k \neq o$ ما برنامه‌ی خطی زیر را حل می‌کنیم:

$$
\begin{aligned}
& e_k^o = \min_{\theta,\lambda} \quad \theta \\
& s.t. \\
& \sum_{j=1}^{n} x_{ij}\lambda_j^k \le \theta x_{ik}, \quad i = 1,\dots,m, \\
& \sum_{j=1}^{n} y_{rj}\lambda_j^k \ge y_{rk}, \quad r = 1,\dots,s, \\
& \lambda_j^k \ge 0, \qquad j = 1,\dots,n, \quad j \neq o, \\
& \lambda_o^k = 0.
\end{aligned}
\tag{۲}
$$

این مدل خطی، شدنی و کران‌دار است و برای واحدهای کارا، هنوز مقدار تابع هدف یک است. قید اضافی $\lambda_o^k = 0$، واحد "o" را از مشاهدات حذف می‌کند.

نتایج حاصل از این مدل منجر به توزیع $n-1$ نمره کارایی برای هر مشاهده است. برای انجام دادن آزمون آماری موردنظر از نمونه n تایی، ما پیشنهاد می‌دهیم که از نمرات کارایی e_o^o که توسط مدل بالا محاسبه‌شده است استفاده شود. توجه داشته باشید که در این مرحله‌ی محاسباتی برای یافتن ابر-کارایی سازگار با اندازه‌ی فارل (۱)، نیاز به حل n^2+n برنامه خطی داریم . بااین‌حال، بر اساس خواص مجموعه‌ی مرجع، ما فقط نیاز به حل مدل (۲) برای مشاهداتی داریم که DMU_o عضو مجموعه‌ی مرجع آن‌ها در مدل (۱) است.

۳-۱ نامساوی چبیشف

یکی از مشهورترین نامساوی‌ها در بحث آمار و احتمال نابرابری چبیشف است با این بیان که برای متغیر تصادفی X با میانگین μ و واریانس σ^2 زمانی که توزیع مشاهدات نرمال هستند داریم $p(|X-\mu| \ge k\sigma) \le \frac{1}{k^2}$. با این فرض که یک مشاهده پرت باشد، پیش‌بینی این است که درواقع دوری آن از میانگین μ بیش‌تر از 3σ است، به خاطر اینکه احتمال مشاهده‌ی چنین داده‌ای تحت فرض نرمال کم است. اهمیت نامساوی چبیشف این است که توزیع داده‌ها در ارزیابی مهم نیست [۱۳].

این فرض صفر را در نظر بگیرید که H_0 : مشاهده پرت نیست، برای انجام دادن این آزمون، ما مقدار p_value را محاسبه می‌کنیم و فرض صفر H_0 را رد می‌کنیم اگر $p\_value < \alpha$. برای بررسی امکان پرت بودن یک واحد بر اساس این آزمون برای کارایی واحد "j" ام، ما از ۲ نمونه n تایی استفاده می‌کنیم. اولی بردار کارایی نمرات $\mathbf{e}$ است که با حل مدل (۱) برای هر مشاهده به دست می‌آید و دومی $\mathbf{e^j}$ که برداری است که شامل نتیجه‌ی مدل (۲) برای واحد کارای "j" ام است، یعنی e_k^j که $k = 1,\dots,n$. تفاوت معناداری بین این ستون‌ها تاثیر واحد کارا "j" ام در نمره کارایی دیگر واحدها را نشان می‌دهد.

[DOR: 20.1001.1.22517286.2016.13.2.8.8]

۳-۲ روش رگرسیون

درروش رگرسیون، بر مبنای کوردر و فرمن [۱۴]، امکان شناسایی واحدهای پرت به‌وسیله‌ی مشخص کردن نقاط موثر با استفاده از اندازه‌ی فاصله‌ی کوک (Cook) و باقی‌مانده‌ی استاندارد میسر خواهد بود. برای شرح این روش در چارچوب DEA، ابتدا $\mathbf{e}$ را به‌عنوان متغیر مستقل و $\mathbf{e^j}$ را به‌عنوان متغیر وابسته دریک مدل رگرسیون خطی در محاسبه‌ی آماره‌ی کوک و باقی‌مانده استاندارد برای تمامی j ها در نظر می‌گیریم. اگر باقی‌مانده‌ی استاندارد یک مشاهده بزرگ‌تر از ۲ باشد یا اندازه‌ی فاصله کوک یک مشاهده خیلی دورتر از دیگر مشاهدات باشد، به‌عنوان نقطه‌ی پرت در نظر گرفته می‌شود.

٤ شناسایی نقاط پرت در یک مثال تجربی

در این بخش، ما از روش‌های آماری اشاره‌شده در بخش قبلی استفاده می‌کنیم تا واحدهای تولیدی که می‌توانند درنتیجه‌ی برآورد کارایی DEA اثرگذار باشند مشخص نماییم. این واحدها دارای عملکرد آن‌قدر بالایی هستند که بسیاری دیگر از DMU ها خوب می‌تواند مغلوب و ناکارآمد در نظر گرفته شوند، به‌عبارت‌دیگر مدل‌های DEA برآورد کم‌تری از نمرات کارایی ارایه دهند.

روش پیشنهادی را با یک مثال تجربی از مدل‌های DEA در ارزیابی کارایی ۴۲ گروه آموزشی واحد دانشگاهی کرج شرح می‌دهیم. در این مطالعه، سه ورودی x_1, x_2, x_3 و چهار خروجی y_1, y_2, y_3, y_4 داریم. جدول ۱ برخی از آمارهای توصیفی در مورد این داده‌ها را ارایه نموده است.

جدول ۱. توصیف آماری داده‌های ۴۲ واحد آموزشی

Variables	Min	Max	Mean	Median	St. Dev
# post graduate students (x_1)	۱	۴۸۴	۷۸/۵۵	۰	۱۳۷/۴
# bachelor students (x_2)	۰	۱۲۰۲	۶/۳۹۴	۳۰۴/۵	۳۶۰/۳
# master students (x_3)	۰	۵۳۵	۲۶/۸۶	۰	۸۲/۲۸
# graduations (y_1),	۳۲	۱۱۵۸	۳۸۵/۴	۳۲۷	۲۹۷/۳
# scholarships (y_2)	۰	۱۴	۲/۳۱	۱	۳/۱۸۱
# research products (y_3)	۰	۱۲	۱/۹۰۵	۱	۲/۷۸۳
manager satisfaction (y_4)	۱	۴	۲/۵۹۵	۳	۰/۸۸۵۱

نمره‌ی کارایی شعاعی ورودی محور با مدل (۱) به‌دست‌آمده و در ستون اول جدول ۲ ارایه‌شده است، که نشان می‌دهد واحدهای ۱۴، ۱۶، ۱۷، ۱۸، ۱۹، ۳۵، ۳۶ و ۳۷ در این ارزیابی کارا هستند. برای بررسی میزان تاثیر هر یک از این واحدها بر واحدهای دیگر، نتایج حاصل از حذف هر واحد کارا روی ساختار مجموعه تولید از طریق مقدار بهینه‌ی مدل (۲) به‌دست‌آمده است و در ادامه، در ستون‌های جدول ۲ ارایه‌شده است.

[DOR: 20.1001.1.22517286.2016.13.2.8.8]

جدول ۲. کارایی محاسبه‌شده، قبل و بعد از حذف واحدهای کارا

DMUj	e_j	e^{14}	e^{16}	e^{17}	e^{18}	e^{19}	e^{35}	e^{36}	e^{37}
۱	۰/۸۸۵۲	۰/۸۸۵۲	۰/۸۸۵۶	۱/۰۰۰۰	۰/۸۸۶۳	۰/۸۸۵۲	۰/۸۸۵۲	۰/۸۸۵۲	۰/۸۸۵۲
۲	۰/۹۵۶۴	۰/۹۸۶۶	۰/۹۵۶۴	۱/۰۰۰۰	۰/۹۵۶۴	۰/۹۵۶۴	۰/۹۵۶۴	۰/۹۵۶۴	۰/۹۵۶۴
۳	۰/۹۳۹۸	۰/۹۶۳۴	۰/۹۳۹۸	۰/۹۶۷۸	۰/۹۳۹۸	۰/۹۳۹۸	۰/۹۳۹۸	۰/۹۳۹۸	۰/۹۳۹۸
۴	۰/۹۴۰۵	۰/۹۶۳۳	۰/۹۴۰۵	۰/۹۷۱۸	۰/۹۴۰۵	۰/۹۴۰۵	۰/۹۴۰۵	۰/۹۴۰۵	۰/۹۴۰۵
۵	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
۶	۰/۹۱۶۸	۰/۹۱۶۸	۰/۹۱۶۸	۰/۹۱۶۸	۰/۹۱۶۸	۰/۹۱۶۸	۰/۹۴۷۸	۰/۹۱۶۸	۰/۹۱۶۸
۷	۰/۸۷۰۹	۱/۰۰۰۰	۰/۸۷۰۹	۰/۸۷۰۹	۰/۸۷۰۹	۰/۸۷۰۹	۰/۸۷۰۹	۰/۸۷۰۹	۰/۸۷۰۹
۸	۰/۵۳۷۸	۰/۵۳۷۸	۰/۵۳۷۸	۰/۶۳۳۱	۰/۵۳۷۸	۰/۵۳۷۸	۰/۵۳۷۸	۰/۵۳۷۸	۰/۵۳۷۸
۹	۰/۹۲۸۵	۰/۹۲۸۵	۰/۹۲۸۵	۰/۹۶۷۷	۰/۹۲۸۵	۰/۹۲۸۵	۰/۹۲۸۵	۰/۹۲۸۵	۰/۹۲۸۵
۱۰	۰/۹۰۱۷	۰/۹۰۱۷	۰/۹۰۱۷	۰/۹۲۶۳	۰/۹۰۱۷	۰/۹۰۱۷	۰/۹۰۱۷	۰/۹۰۱۷	۰/۹۰۱۷
۱۱	۰/۷۷۲۷	۰/۹۲۴۲	۰/۷۷۲۷	۰/۷۷۲۷	۰/۷۷۲۷	۰/۷۷۲۷	۰/۷۷۲۷	۰/۷۷۲۷	۰/۷۷۲۷
۱۲	۰/۲۷۱۱	۰/۲۷۱۱	۰/۲۷۱۱	۰/۳۰۴۶	۰/۲۷۱۱	۰/۲۷۱۱	۰/۲۷۱۱	۰/۲۷۱۱	۰/۲۷۱۱
۱۳	۰/۸۸۲۳	۰/۸۸۲۳	۰/۹۰۲۸	۰/۹۲۹۷	۰/۸۸۸۳	۰/۸۸۲۳	۰/۸۸۲۳	۰/۸۸۲۳	۰/۸۸۲۳
۱۴	۱/۰۰۰۰	۱/۱۹۶۱	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
۱۵	۰/۷۵۷۹	۰/۷۵۷۹	۰/۷۵۸۴	۰/۸۵۷۱	۰/۷۷۱۹	۰/۷۵۷۹	۰/۷۵۷۹	۰/۷۵۷۹	۰/۷۵۷۹
۱۶	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۴۵۹	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
۱۷	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۴/۷۵۱۸	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
۱۸	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۲۵۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
۱۹	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۱۸	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
۲۰	۰/۸۹۲۳	۰/۸۹۲۳	۰/۸۹۲۳	۰/۸۹۲۳	۰/۸۹۲۳	۰/۸۹۲۳	۰/۸۹۲۳	۰/۸۹۲۳	۰/۸۹۲۳
۲۱	۰/۸۷۹۶	۰/۸۷۹۶	۰/۸۸۸۸	۰/۸۷۹۹	۰/۸۸۸۵	۰/۸۷۹۶	۰/۸۷۹۶	۰/۸۷۹۶	۰/۸۷۹۶
۲۲	۰/۸۷۴۲	۰/۸۷۴۲	۰/۸۸۱۷	۰/۸۷۴۴	۰/۸۸۵۸	۰/۸۷۴۲	۰/۸۷۴۲	۰/۸۷۴۲	۰/۸۷۴۲
۲۳	۰/۸۳۹۶	۰/۸۳۹۶	۰/۸۴۲۶	۰/۸۵۶۳	۰/۸۵۶۳	۰/۸۳۹۶	۰/۸۳۹۶	۰/۸۳۹۶	۰/۸۳۹۶
۲۴	۰/۷۵۶۹	۰/۷۵۶۹	۰/۷۵۷۰	۰/۷۶۵۵	۰/۷۷۱۱	۰/۷۵۶۹	۰/۷۵۶۹	۰/۷۵۶۹	۰/۷۵۶۹
۲۵	۰/۷۳۸۸	۰/۷۳۸۸	۰/۷۴۱۹	۰/۷۳۹۱	۰/۷۴۹۶	۰/۷۳۸۸	۰/۷۳۸۸	۰/۷۳۸۸	۰/۷۳۸۸
۲۶	۰/۸۹۰۱	۰/۸۹۰۱	۰/۸۹۰۱	۰/۹۱۶۹	۰/۹۰۷۱	۰/۸۹۰۱	۰/۸۹۰۱	۰/۸۹۰۱	۰/۸۹۰۱
۲۷	۰/۹۹۹۰	۰/۹۹۹۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۰/۹۹۹۰	۰/۹۹۹۰	۰/۹۹۹۰	۰/۹۹۹۰
۲۸	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
۲۹	۰/۶۱۸۱	۰/۶۱۸۱	۰/۶۱۸۱	۰/۷۹۵۷	۰/۶۱۸۱	۰/۶۱۸۱	۰/۶۱۸۱	۰/۶۴۸۷	۰/۶۱۸۱
۳۰	۰/۶۲۱۷	۰/۶۲۱۷	۰/۶۲۱۷	۰/۹۷۷۵	۰/۶۲۱۷	۰/۶۲۱۷	۰/۶۲۱۷	۰/۶۲۱۷	۰/۶۲۱۷
۳۱	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
۳۲	۰/۷۷۶۱	۰/۷۷۶۱	۰/۷۷۶۱	۰/۷۹۴۷	۰/۷۷۶۱	۰/۷۷۶۱	۰/۷۷۶۱	۰/۷۷۶۱	۰/۷۷۶۱
۳۳	۰/۸۱۶۳	۰/۸۱۶۳	۰/۸۱۶۳	۰/۸۱۶۳	۰/۸۱۶۳	۰/۸۱۶۳	۰/۸۴۴۶	۰/۸۱۶۳	۰/۸۱۶۳
۳۴	۰/۹۷۱۱	۰/۹۷۱۱	۰/۹۷۲۲	۱/۰۰۰۰	۰/۹۷۱۱	۰/۹۷۱۱	۰/۹۷۱۱	۱/۰۰۰۰	۰/۹۷۱۱
۳۵	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۷۵۷	۱/۰۰۰۰	۱/۰۰۰۰
۳۶	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۱۷۴۴	۱/۰۰۰۰
۳۷	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	۵/۴۸۲۸
۳۸	۰/۹۲۱۵	۰/۹۲۶۷	۰/۹۲۱۵	۱/۰۰۰۰	۰/۹۲۱۵	۰/۹۲۱۵	۰/۹۲۱۵	۰/۹۲۱۵	۰/۹۲۱۵
۳۹	۰/۳۶۴۰	۰/۳۶۶۵	۰/۳۶۴۰	۰/۴۰۱۶	۰/۳۶۸۶	۰/۳۶۴۰	۰/۳۶۴۰	۰/۳۶۴۰	۰/۳۶۴۰
۴۰	۰/۹۲۲۷	۰/۹۷۱۰	۰/۹۲۲۷	۰/۹۲۲۷	۰/۹۲۲۷	۰/۹۲۲۷	۰/۹۲۲۷	۰/۹۲۲۷	۰/۹۲۲۷
۴۱	۰/۷۷۷۲	۰/۷۷۷۲	۰/۷۷۷۲	۰/۷۷۷۲	۰/۷۷۷۲	۰/۷۷۷۲	۰/۷۷۷۲	۰/۷۷۷۲	۰/۷۷۷۲
۴۲	۰/۷۲۸۸	۰/۷۲۸۸	۰/۷۲۸۸	۰/۷۲۸۸	۰/۷۲۸۸	۰/۷۲۸۸	۰/۷۲۸۸	۰/۷۲۸۸	۰/۷۲۸۸

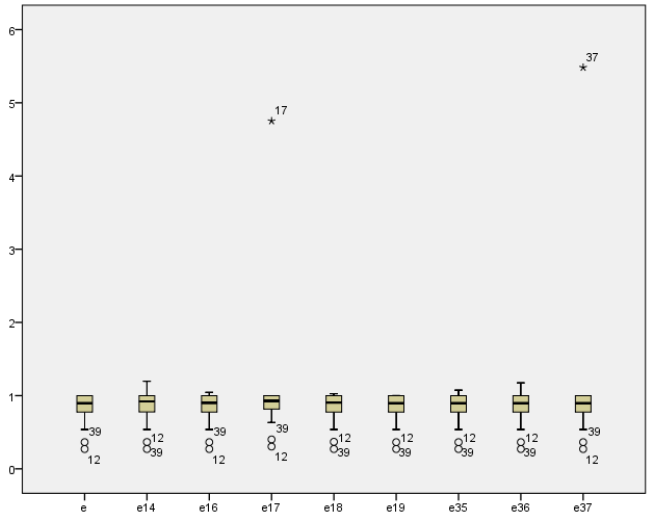
[DOR: 20.1001.1.22517286.2016.13.2.8.8]

برای تعیین نقاط پرت بالقوه، در ابتدا، ما با استفاده از روش‌های توصیفی بر روی کارایی شروع به پیدا کردن واحدهای کارا خواهیم کرد که به‌عنوان کاندیداهای نقاط پرت هستند.

آمار توصیفی نمرات محاسبه‌شده مربوط به حذف واحدهای دانشگاهی کارا در جدول ۳ آمده است. با توجه به شکل ۱، به نظر می‌رسد واحد ۱۷ و ۳۷ می‌توانند به‌عنوان نقاط دورافتاده در نظر گرفته شوند. برای تعیین معنادار بودن نتایج از روش‌های استنباطی که در بالا شرح داده شد، نابرابری چبیشف و رگرسیون خطی استفاده‌شده است. جدول ۴ و ۵ نتایج حاصله را نشان می‌دهند.

جدول ۳. توصیف آماری کارایی‌های محاسبه‌شده

	e	e^{14}	e^{16}	e^{17}	e^{18}	e^{19}	e^{35}	e^{36}	e^{37}
Mean	۰/۸۵۵۹	۰/۸۷۰۵	۰/۸۵۸۱	۰/۹۷۶۴	۰/۸۵۹۱	۰/۸۵۶۰	۰/۸۵۹۱	۰/۸۶۱۵	۰/۹۶۲۷
Variance	۰/۱۶۹۶	۰/۱۷۸۵	۰/۱۷۰۸	۰/۶۱۶۴	۰/۱۶۹۴	۰/۱۶۹۶	۰/۱۷۱۷	۰/۱۷۴۷	۰/۷۳۴۰



شکل ۱. تشخیص واحدهای پرت ۱۷ و ۳۷

در اینجا، برای e^{37} و e^{17} فرض H_0 رد می‌شود، به این دلیل که مقدار p_value آن‌ها ۰/۰۲۶۷ و ۰/۰۲۶۴ کوچک‌تر از ۰/۰۵ است و درنتیجه آن‌ها به‌عنوان نقاط پرت بالقوه در نظر گرفته می‌شوند.

جدول ٤. نتایج مقادیر p_value برای نامساوی چبیشف

	e^{14}	e^{16}	e^{17}	e^{18}	e^{19}	e^{35}	e^{36}	e^{37}
Upper bound	۰/۲۴۸۶	۰/۹۸۹۴	۰/۰۲۶۷	۱	۱	۰/۶۲۸۴	۰/۲۲۰۷	۰/۰۲۶۴

[DOR: 20.1001.1.22517286.2016.13.2.8.8]

جدول ۵. اندازه‌ی فاصله‌ی کوک و باقیمانده‌ها

	$e^{۱۴}$	$e^{۱۶}$	$e^{۱۷}$	$e^{۱۸}$	$e^{۱۹}$	$e^{۳۵}$	$e^{۳۶}$	$e^{۳۷}$
S.RE.	۰/۱۷۸۳۱	۰/۴۲۸۱	۳/۵۸۷۱	۰/۰۲۲۰۸	۰/۰۰۱۷۳	۰/۰۷۱۰۶	۰/۱۶۶۲۴	۴/۲۹۷۱۴
COO.	۰/۳۹۲۷۵	۰/۶۷۶۷۸	۰/۸۵۵۱۳	۰/۲۸۰۰۲	۰/۸۶۴۱۳	۰/۶۵۵۶۲	۰/۸۱۶۷۴	۰/۸۶۴۱۳

هم‌چنین در جدول ۵ موارد $e^{۳۷}$ و $e^{۱۷}$ هر دو باقیمانده‌ی استاندارد بزرگ‌تر از ۲ را با معیارهای فاصله‌ی بسیار بزرگ کوک دارند. پس، این دو مورد دوباره به‌عنوان نقاط پرت پیشنهادی در نظر گرفته می‌شوند.

۵ نتیجه‌گیری

در این مقاله دو آزمون آماری برای شناسایی نقاط پرت به‌عنوان اولین قدم قبل از استفاده از هر برآوردگر DEA معرفی‌شده است. نشان داده شد که روش ارایه‌شده یک ابزار قدرتمند برای حذف هرگونه واحد پرت بالقوه‌ی ابر کارا است. مطالعه‌ی تجربی مربوط به یک واحد دانشگاهی نشان‌دهنده‌ی سهولت اجرا و کاربردی بودن روش‌های پیشنهادی است. از روش‌های پیشنهادی می‌توان به‌عنوان گام نخست پالایش داده‌ها، در اغلب زمینه‌های کاربردی تحلیل پوششی داده‌ها نظیر ارزیابی کارایی، بازده به مقیاس، تخصیص مجدد منابع [۱۵] و بسیاری دیگر از زمینه‌ها بهره‌مند شد.

منابع

[۱۵] کردرستمی، س.، امیرتیموری، ع.، فاضلی سندیانی، س.، (۱۳۹۰). تخصیص مجدد منابع با حفظ پایداری مرزهای کارآ در مناطق. مجله تحقیق در عملیات در کاربردهای آن، ۸ (۴)، ۱۰۵-۹۳.

[1] Charnes, A., Cooper, W. W., Rhodes, E., (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2, 429–444.

[2] Banker, R. D., Gifford, J. L., (1988). A relative efficiency model for the evaluation of public health nurse productivity. Mellon University Mimeo, Carnegie.

[3] Banker, R. D., Datar, S. M., (1989). Analysis of cost variances for management control in hospitals. JAI Press;5, 268–91.

[4] Wilson, P. W., (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. Journal of Business and Economic Statistics, 11, 319–323.

[5] Wilson, P. W., (1995). Detecting influential observations in data envelopment analysis. Journal of Productivity Analysis, 6, 27–45.

[6] Ondrich, J., Ruggiero, J., (2002). Outlier detection in data envelopment analysis: an analysis of jackknifing. Journal of the Operations Research Society, 53, 342-346.

[7] Simar, L., (2003). Detecting outliers in frontier models: A simple approach. Journal of Productivity Analysis, 20, 391–424.

[8] Cazals, C., Florens, J. P., Simar, L., (2002). Non-parametric frontier estimation: a robust approach. Journal of Econometrics, 106, 1-25.

[9] Banker, R. D., Chang, H., (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. European Journal of Operational Research, 175, 1311–1320.

[10] Johnson, A. L., McGinnis, L. F., (2008). Outlier detection in two-stage semi-parametric DEA models. European Journal of Operational Research, 187, 629-635.

[11] Chen, W. C., Johnson, A. L., (2010). A unified model for detecting efficient and inefficient outliers in data envelopment analysis. Computers & Operations Research, 37, 417-425.

[DOR: 20.1001.1.22517286.2016.13.2.8.8]

[12] Bellini, T., (2012.). Forward search outlier detection in data envelopment analysis. European Journal of Operational Research, 216, 200-207,
[13] Gouri, K., Bhattacharyya, R., Johnson, A., (1977). Statistical concepts and methods. John Wiley & Sons.
[14] Corder, G., W., Foreman, D. I., (2009). Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach. New Jersey: Wiley.

[DOR: 20.1001.1.22517286.2016.13.2.8.8]