

قانون برادفورد در مجموعه‌های نامعین: استنتاج‌های آماری برای تحلیل کتابخانه¹

استفان جی. بنسمن

ترجمه حیدر مختاری و آنسه حسینی‌زاده²

چکیده

در این نوشته زیربنای فکری و جنبه‌های عملی قانون برادفورد تشریح شده است. تفاوت مقوله‌بندی سنتی با مجموعه‌های نامعین «لطفی‌زاده» شرح داده شده، اندیشه‌های «کارل پیرسون» درباره «مسئله بنیادی علم» آورده شده، به «داده‌های پرت» از دید «بارنت» و «لوئیس» پرداخته شده و با تکیه بر این موارد، گزارشی از یک بررسی که در دانشگاه لویزیانا در پیوند با انتخاب مجلات شیمی توسط اعضای هیئت علمی و مقایسه آن با «آی‌اس‌ای» انجام شد، ارائه گردیده و نتیجه گرفته شده است که مجموعه‌های برادفورد نیز نامعین هستند و سرانجام چند شیوه برای مهار کردن «داده‌های پرت» آورده شده و اعلام شده است که نظریه احتمالات و نظریه مجموعه‌های نامعین متناقض نیستند، اما قالب‌هایی مکمل برای تجزیه و تحلیل به حساب می‌آیند.

کلیدواژه‌ها: قانون برادفورد، مجموعه‌های نامعین، تحلیل‌های آماری، گزینش نشریه‌های ادواری

از وقتی که «اس.سی.برادفورد»³ اولین مقاله خود را در حوزه قانون کتابسنجی منتشر ساخت، بیش از شصت سال می‌گذرد. از آن زمان تا به حال درباره قانون پراکندگی برادفورد بسیار نوشته‌اند؛ اما هنوز ناگفته‌های بسیاری وجود دارد. من پس از انجام تحقیقات نظری و عملی دامنه‌دار در ارزیابی و کاربرد مجلات علمی، چنین نتیجه‌گیری کرده‌ام که قانون برادفورد اساساً ملغمه‌ای از دو مفهوم بسیار مهم در کاربرد روش‌های آماری در امر تجزیه و تحلیل کتابخانه - یعنی توزیع احتمال و مجموعه‌های نامعین - است. در واقع، این قانون را می‌توان توصیف ریاضی الگوی احتمال برای تشکیل مجموعه‌های نامعین دانست. در مقاله حاضر به استنتاج‌های آماری قانون برادفورد در ایجاد مجموعه‌های نامعین پرداخته شده است.

قانون برادفورد

«برادفورد» کتابدار ارشد کتابخانه موزه علوم در کنسینگتون⁴ جنوبی لندن بود. هدف عمده وی از این تحقیق که منجر به ایجاد قانون پراکندگی شد، بهینه‌سازی میزان پوشش نشریات علمی در خدمات نمایه‌سازی و چکیده‌نویسی بود. نقص در این پوشش که با برآورد و مشاهده حدود 500,000 مقاله علمی منعکس نشده در خدمات نمایه‌سازی و چکیده‌نویسی از مجموع 750,000 مقاله علمی منتشر شده در یک سال آشکار شد،

¹. Bensman, Stephen J. "Bradford's law and fuzzy sets: statistical implementations for library analyses". IFLA Journal, 27(4). 2001. P: 283-46.

². دانشجویان کارشناسی ارشد کتابداری و اطلاع‌رسانی پزشکی

³. Samuel Clement Bradford

⁴. Kensington

«برادفورد» را به اندیشه واداشت. علت این امر شاید در شیوه پراکندگی و توزیع مطالب مربوط به یک موضوع معین در نشریات ادواری حاوی آن موضوع باشد. «برادفورد» فرضیه‌ای را از طریق این تحقیق به صورت زیر مطرح کرد:

... فرضیه دیگر آن است که توزیع فراوانی ارجاعات در تمامی نشریات ادواری تا حد زیادی با حیطه موضوعی آن‌ها نسبت معکوس دارد. برطبق این فرضیه، از لحاظ ارتباط محدوده موضوعی با موضوع موردنظر، مجموع نشریات ادواری را می‌توان به مجموعه‌هایی تقسیم کرد؛ به نحوی که در این مجموعه، رده‌های دورتر به اندازه رده‌های مرتبط [با موضوع معین] در تولید مراجع نقش دارند. محدوده کل نشریات ادواری، بر این اساس، به مثابه خانواده‌ای از نسل‌های متناوب هستند که نسبت خویشاوندی آن‌ها پیوسته کم می‌شود و هر نسل از نسل قبلی خود بیشتر است و برحسب درجه دوری، تولید نسل‌ها تنزل می‌یابد.

این فرضیه با دو دسته از ارجاعات برگرفته از کتابشناسی‌های جاری (گردآوری شده در کتابخانه موزه علوم⁵) که یکی مربوط به ژئوفیزیک کاربردی و دیگری مربوط به روغنکاری بود، مورد آزمون قرار گرفت. با توجه به این که نتایج حاصل از بررسی ارجاعات ژئوفیزیک کاربردی و روغنکاری یکسان بوده‌اند، در مقاله حاضر، تحلیل خود را به ژئوفیزیک کاربردی محدود کرده‌ام. مجموعه ژئوفیزیک کاربردی شامل کل ارجاعات چهار سال (از 1928 تا 1931)، جمعاً 1332 ارجاع موجود در 326 عنوان مجله بوده است. در صورت مرتب‌شدن این مجلات به صورت نزولی، رتبه‌بندی از یک مجله حاوی 93 ارجاع تا 109 مجله حاوی یک ارجاع در نوسان می‌باشد. «برادفورد» در گزارش خود، نتایج حاصل از گروه‌بندی مجلات در هر دسته را بدین گونه در سه رده زیر جای داد:

a. مجلاتی که سالانه بیش از 4 ارجاع (به‌طور متوسط) داشته‌اند؛

b. مجلاتی که سالانه حداکثر 4 و حداقل بیشتر از یک ارجاع داشته‌اند؛

c. مجلاتی که سالانه دارای یک ارجاع یا کمتر بوده‌اند.

جدول 1 نشان‌دهنده گروه‌بندی «برادفورد» از مجلات ژئوفیزیک کاربردی و تعیین رده برای آن‌ها است. پرواضح است که این نتایج بر فرضیات آزمون‌شده مهر تأیید می‌زنند؛ بنابراین، رده a با وجود شمول بر 2/8 درصد مجلات، 32/2 درصد ارجاعات را دربرمی‌گیرد؛ رده b با شمول بر 18/1 درصد مجلات، 37/5 درصد ارجاعات را به خود اختصاص داده است و بالاخره رده c با دارا بودن 79/1 درصد مجلات، فقط 30/3 درصد ارجاعات را دارد.

جدول 1 به وضوح نشان می‌دهد که با وجود افزایش نمایی تعداد مجلات در هر رده، تعداد ارجاعات موجود در آن‌ها تقریباً ثابت می‌ماند. چشم‌انداز به‌دست آمده از این داده‌ها، برادفورد را به ارائه فرمول‌بندی شفاهی زیر از قانون پراکندگی رهنمون شد:

... در صورت رتبه‌بندی نزولی مجلات علمی برحسب میزان کاهش تولید مقالات در یک زمینه معین موضوعی، می‌توان آن‌ها را به نشریات ادواری هسته که اختصاص بیشتری به موضوع موردنظر دارند و چند گروه یا ناحیه حاوی همان تعداد مقاله موجود در گروه هسته تقسیم کرد؛ به طوری که نشریات ادواری هسته و گروه‌های بعدی دارای نسبت $1:n:n^2$ باشد.

البته برادفورد ارتباط تعداد مقالات در موضوع معین از مجلات مختلف را با اندازه آن‌ها بررسی نکرده است. چون دغدغه عمده وی همسوساختن کارگزاران خدمات نمایه‌سازی و چکیده‌نویسی از طریق ایجاد راهکارهایی برای اطمینان از دسترسی کامل آن‌ها به تمام عناوین موضوعی بوده، به این ارتباط توجهی نشان نداد. توجه «برادفورد» به مقالات یک موضوع معین و طرز توزیع آن‌ها در عناوین مختلف بود و بنابراین بحث

⁵. Science Museum Library

اصلی ما درصد مقالات است. برای دسته‌بندی مجلات به شیوه استاندارد، باید موضوع را از دید خود مجلات - نه از دید مقالات - بررسی کرد و میزان درصد مقالات در یک مجله خاص مربوط به یک موضوع خاص را تعیین نمود. می‌توان چنین اظهار نظر کرد که عنوانی با بیشترین مقاله در یک موضوع معین - بویژه یک موضوع اخص - یک مجله چندرشته‌ای با نسبت کمی از مقالات مربوط به آن موضوع است. من در اینجا از روش دسته‌بندی برادفورد بر حسب درصد مقالات مربوط به موضوع - و نه بر حسب درصد مقالات در یک مجله معین - پیروی کرده‌ام. با این حال، این فرض برگرفته از قانون برادفورد که «نشریات ادواری هسته بیشترین اختصاص را به موضوع مورد نظر دارند» در واقع نشانگر مجلات پایه حاوی بیشترین مقاله در یک موضوع معین است و بنای مقاله حاضر بر این فرض نهاده شده است.

از نظر گاه «تئوری مجموعه‌های نامعین» جنبه مهم قانون پراکنده‌گی برادفورد، ترسیم واقعیت مربوط به نقطه نهایی در فرضیه اولیه می‌باشد، مبنی بر اینکه در هر مجموعه موضوعی معین «محدوده کلی نشریات ادواری ... به صورت خانواده‌هایی از نسل‌های متوالی است که در آن، درجه خویشاوندی پیوسته کم و کمتر می‌شود، تعداد هر نسل از نسل قبلی آن بیشتر است، و هر نسل بسته به میزان دوری و کم شدن میزان خویشاوندی، تولید و زایش کم و کمتری دارد».

جدول 1. رده‌بندی مجلات از دیدگاه برادفورد در ژنوفیزیک کاربردی از سال 1928 تا 1931

ارجاعات		مجلات		رده
درصد	تعداد	درصد	تعداد	
32/2	429	2/8	9	a. بیش از چهار ارجاع در سال [17-93 ارجاع]
37/5	499	18/1	59	b. 2.4 ارجاع در سال [5-16 ارجاع]
30/3	404	79/1	258	c. 0/1.1 ارجاع در سال [4 تا ارجاع]
100	1332	100	326	کل

مجموعه‌های نامعین⁶ از دیدگاه «لطفی زاده»

تئوری کلاسیک مجموعه‌ها بر پایه این اندیشه قرار دارد که می‌توان بین گروه‌ها تفاوت‌های جزئی و مشخص قائل شد. بر اساس این تئوری، امکان تعیین عضویت / عدم عضویت قطعی یک چیز در یک گروه وجود دارد. «کوسکو»⁷ به نقل از «ارسطو»⁸، ایده‌های وی را از این مفهوم در کتابش در باب منطق نامعین این گونه بیان می‌کند:

منطق دودویی ارسطو به یک قانون تعمیم‌پذیری دارد: A یا نه A، یا این یا نه این. آسمان یا آبی است، یا آبی نیست و نمی‌تواند هم آبی باشد و هم آبی نباشد، «هم A هم نه A» امکان ندارد. «A یا نه A» بیانی ساده از قانون میانگین خروج از عضویت است. در عملکردهای مجموعه کلاسیک، به A عدد یک و به نه A عدد صفر را تخصیص می‌دهند.

«برادفورد» بی‌شک بر اساس تئوری مجموعه کلاسیک می‌اندیشیده است. این رویکرد از مقاله‌ای که وی در «انجمن بریتانیایی کتابشناسی بین‌المللی»⁹ در سال 1944 ارائه داد، کاملاً برمی‌آید. وی در این مقاله با کاوش در بنیادهای رده‌بندی دهدهی جهانی، رده را «مجموعه‌ای از موجودیت‌ها و اشیا با یک ویژگی مشترک» می‌داند. وی از نظام نمادها و منطق مطرح شده در کتاب «جورج بول»¹⁰ با عنوان قوانین اندیشه

⁶ fuzzy sets

⁷ Kosko

⁸ Aristotle

⁹ British Society of International Bibliography

¹⁰ George Boole

(1854) که منطق شکل‌گیری چنین مجموعه‌هایی در ذهن و اندیشهٔ آدمی را تبیین می‌کند، استفاده کرد. وی با به‌دست آوردن معادله‌ای که مشتق از بنیادهای حساب منطقی بول بود، به نتیجهٔ زیر رسید:

بنابراین به قانونی دست می‌یابیم که برطبق آن «هم تعلق و هم عدم تعلق یک ویژگی به چیزی واحد محال است». این همان اصل تعارض ارسطو می‌باشد که [بول] آن را به عنوان اصلی بسیار متقن و مبرهن، باز شناخت.

این که چنین معادله‌ای از درجهٔ دو با دو ریشهٔ صفر و یک است، نشان می‌دهد که ما عمل طبقه‌بندی را با انتزاع به جفت‌های متضاد، مثل مردان و غیرمردان انجام می‌دهیم و درمی‌یابیم که برای هر ردهٔ مشخص شده با نماد فرضی X ، فقط ارزش صفر و یک می‌توان در نظر گرفت.

با وجود این «برادفورد» از این امر کاملاً آگاهی داشت که «مانع‌الجمع نبودن رده‌ها، همواره عملی و ممکن نیست».

در نوشته‌ها، به منظور مشخص شدن تمایز بین مجموعه‌های کلاسیک از مجموعه‌های «نامعین»، آن‌ها را مجموعه‌های «معین»¹¹ نامیده‌اند. مفهوم اخیر را برای نخستین بار «لطفی زاده»¹² در مقاله‌ای که در سال 1965 منتشر ساخت بیان داشت. وی در این مقاله تعریف و اهمیت مجموعه‌های نامعین را این‌گونه شرح داد:

معمولاً رده‌های اشیای موجود در جهان واقعی، معیارهای دقیق و معین عضویت ندارند. مثلاً ردهٔ حیوانات مشخصاً شامل سگ‌ها، اسب‌ها، پرندگان، و غیره به عنوان اعضای آن می‌شود و اشیایی چون سنگ‌ها، مایعات، گیاهان و... بدان تعلق ندارند. با این حال، هویت‌هایی چون ستاره‌های دریایی، باکتری‌ها و... در ردهٔ حیوانات، جایگاهی مبهم دارند. چنین ابهامی را برای عددی چون 10 در رابطه با «رده» تمامی اعداد حقیقی بسیار بزرگتر از 1 می‌توان دید.

مسئلهٔ «گروه تمامی اعداد بسیار بزرگتر از یک» یا «ردهٔ زنان زیبا» و یا «ردهٔ مردان بلندقد» از نقطه‌نظر اصطلاحات معمول ریاضی، مجموعه یا رده به حساب نمی‌آیند؛ اما حقیقت آن است که چنین «رده‌هایی» با تعریفی مبهم و نامشخص در اندیشهٔ بشری، بویژه در حوزه‌های شناخت الگو، ارتباطات اطلاعات و انتزاع نقشی مهم دارند.

«لطفی زاده» مجموعهٔ نامعین را «رده‌ای از اشیای با درجات عضویت پیوستاری» می‌داند و اظهار می‌دارد که «یک چنین مجموعه‌ای با یک عامل عضویت (تخصیص) که درجه‌ای از عضویت بین صفر و یک را به هر چیزی اختصاص می‌دهد، شناخته می‌شود» (ص. 338). یک مجموعهٔ A (مثلاً ردهٔ مردان بلندقد در نمونه‌های «لطفی زاده») در نظریهٔ مجموعه‌ها زیر گروهی از مجموعهٔ مرجع جهانی X (مثلاً ردهٔ تمامی مردان) است. براساس نظر «لطفی زاده» به هر عضو X از مجموعهٔ جهانی X با استفاده از عامل عضویت، «درجهٔ عضویتی» در مجموعهٔ A اختصاص می‌یابد که از صفر (نه A) تا یک (A) در نوسان است. او خاطر نشان می‌سازد که مجموعه‌های نامعین مورد نظر وی، متفاوت از نوع متداول مجموعه‌های معین است که در آن‌ها، تابع عضویت صرفاً برحسب تعلق یا عدم تعلق X به مجموعهٔ A (که با دو ارزش یک و صفر است) شناخته می‌شود.

در عملگرهای مجموعهٔ کلاسیک با دو مجموعه، فرض بر آن است که مورد مشاهده بطور مجزایی به هر دو مجموعه تعلق دارد؛ مثلاً، انسان می‌تواند همزمان هم دختر و هم دانشجو باشد و بر این اساس، می‌تواند در هر دو مجموعهٔ همپوشان فوق جای داشته باشد. ولی در مجموعه‌های نامعین، عضویت ناواضح است و این می‌تواند تمام انواع متغیرهای بیرونی را به صحنه آورد. با توجه به مثال بالا، [از نظر تئوری مجموعه‌های نامعین]،

¹¹ crisp
¹² Lotfi Zadeh

دختر موردنظر در واقع دو جنسی است. از همین یک مورد می توان دریافت که چگونه مجموعه های نامعین، روابط را پیچیده می کند.

«بایس» و «میدو» و «کرافت»¹³ در کتاب خود در باب سنجش در علم اطلاع رسانی می نویسند: «مسئله اساسی در اندازه گیری مجموعه های نامعین، ارزش گذاری میزان عضویت در مجموعه است» از نظر این نویسندگان، چنین امری کم و بیش اختیاری است. آنان برای نمایه سازی مدارک دو روش اساسی را توصیه می کنند: نمایه سازی ذهنی توسط نمایه سازان و نمایه سازی عینی و عملی توسط نرم افزارهای رایانه ای بر اساس فراوانی کلمات در متن.

خود «برادفورد» در گزارش قانون پراکندگی، برای به دست آوردن تابع عضویت متناسب با این قانون، از نمودهای عینی استفاده کرد. این نمودها در شیوه به کار رفته برای تعیین رده های a و b و c در جدول 1، آشکارا موجودند. با بررسی این روند، می توان به شیوه وی پی برد که عبارت است از تقسیم کل ارجاعات هر مجله به تعداد سال های دربرگیرنده نمونه (چهار سال برای ژئوفیزیک کاربردی) به منظور تعیین تعداد کل ارجاعات در هر سال. بر این اساس، رده های معین شده به صورت زیر قابل تبیین اند: (a) مجلاتی که سالانه بیش از 4 ارجاع دارند، (b) مجلاتی که سالانه حداقل 4 و حداکثر بیش از 1 ارجاع دارند، و (c) مجلاتی که سالانه 1 ارجاع یا کمتر دارند. پیاده سازی این روش به خارج قسمت هایی با حداکثر دو مرتبه اعشاری می انجامد که عامل ایجاد رده هایی با مرزبندی های زیر است: بین a و b در 4/01؛ بین b و c در 1/01، و بین c و رده صفر - که من خود آن را اضافه کرده و d نامیده ام - در 0/01. اگر رده a را که شامل 2/8 درصد مجلات و 32/2 درصد ارجاعات در موضوع ژئوفیزیک کاربردی است، در نظر بگیریم و آن را «هسته نشریات ادواری و بسیار مختص و مرتبط با موضوع» بدانیم، برای مجموعه های برادفورد، تابع عضویت زیر قابل تخصیص است:

اگر تعداد ارجاعات سالانه به یک مجله از 4 بزرگتر باشد، نمره/درجه عضویت آن مجله برابر یک است. ولی اگر تعداد ارجاعات سالانه به یک مجله برابر 4 یا کمتر از آن باشد، درجه عضویت آن مجله برابر با تعداد ارجاعات به آن در سال، تقسیم بر 4/01 است.

عدد 4/01 با نشان دادن پایین ترین حد هسته به عنوان خارج قسمت در بخش دوم تابع عضویت انتخاب شده است. به کارگرفتن چنین تابع عضویتی برای داده های برادفورد، منتج به نتایج جدول 2 در موضوع ژئوفیزیک کاربردی می شود. در این جدول، رده های برادفورد به همراه رده اضافه شده صفر (d) بر اساس اصول مجموعه های نامعین که به ترتیب، تنزل عضویت مجموعه را می نمایاند، نامگذاری و نشان داده شده اند. $A=a$ ؛ $A=b$ ؛ $A=c$ نه A و A و بالاخره $d=A$ نه A . بررسی این جدول نشان می دهد که در پایین مجلات هسته یا همان رده A، نمره عضویت مجلات سریعاً کم می شود، به طوری که تعداد این مجلات همزمان زیاد می شود و در نهایت، حجم عظیمی از مجلات، در واقع صرفاً به صورت حاشیه مجموعه ژئوفیزیک کاربردی به حساب می آیند. تعداد مجلات موجود در رده صفر (d) تماماً مشخص نشده است؛ چه، این خود سؤال پیچیده ای است که «برادفورد» هم نتوانست با موفقیت به آن پاسخ گوید.

استنتاج های آماری «پیرسون»¹⁴

وجود خصیصه نامعین بودن مجموعه های برادفورد، در کاربرد روش های آماری برای تحلیل کتابخانه

¹³ Bayce, Meadow and Kraft

¹⁴ Karl Pearson

تأثیر بسزایی دارد. این تأثیرات برگرفته از اصل وحدت علمی است، همان چیزی که «برادفورد» قانون خود را براساس آن بنا نهاد. «برادفورد» اظهار داشت: «برطبق این اصل، هر موضوع علمی، کم و بیش با دیگر موضوعات علمی پیوستگی دارد». بر مبنای این اصل، هر اندازه نمره عضویت مدارک یا مجلات در یک مجموعه معین برادفورد کم شود، راه برای دیگر موضوعات علمی هموار می شود و بدین شکل عوامل بیرونی به این مجموعه نفوذ می کنند. این امر منتج به ناهمگنی - و چه بسا ناهمگنی شدید - در مجموعه های برادفورد می شود.

جدول 2. قانون برادفورد از دید تئوری مجموعه های نامعین در موضوع ژئوفیزیک کاربردی، سال های 1928-1931

نمره عضویت	مجلات دارای ارجاعات	تعداد ارجاعات در سال	رده ها
1/000	1	23/25	ژئوفیزیک کاربردی (a)
1/000	1	21/50	
1/000	1	14/00	
1/000	1	12/00	
1/000	1	11/50	
1/000	1	8/75	
1/000	1	7/00	
1/000	1	5/00	
1/000	1	4/25	
1/000		4/01	
0/988	4	4/00	ژئوفیزیک کاربردی و نه ژئوفیزیک کاربردی (b)
0/935	1	3/75	
0/873	5	3/50	
0/748	1	3/00	
0/686	2	2/75	
0/62	5	2/50	
0/591	3	2/25	
0/499	8	2/00	
0/436	7	1/75	
0/374	11	1/50	
0/312	12	1/25	
0/252		1/01	
0/249	17	1/00	نه - ژئوفیزیک کاربردی و ژئوفیزیک کاربردی (c)
0/187	23	0/75	
0/125	49	0/50	
0/062	169	0/25	
0/002		0/01	مرز رده های c و d
0/000	؟	0/00	نه - ژئوفیزیک کاربردی (d)

این ناهمگنی آنچه را که «کارل پیرسون» مسئله بنیادی علم خواند، با پیچیدگی مواجه می سازد. «پیرسون» اساس این مسئله را در ویراست سال 1911 دستور زبان علم، در مبحث عمومی مربوط به مفاهیم نوین احتمال و همبستگی مطرح ساخت و در رویکرد علمی خود با اشاره به اصل رده بندی نوشت: «رده بندی اشیا و

شکل‌گیری قضاوت‌های محض بر مبنای این رده‌بندی - قضاوت‌های مستقل از اندیشه‌های ذهنی افراد - اساساً هدف و شیوه علم نوین است». به زعم وی، تغییر، خصیصه اصلی واقعیت است و در ایجاد شرایط برای کارکردهای علمی نقشی شایان توجه دارد. وی می‌نویسد:

... اساس نتایج فیزیکدان یا شیمیدان، تجارب متوسط‌اند که دقیقاً منطبق بر هم نیستند. رویه اداری آن‌ها، حتی در مهم‌ترین شرایط هم تفاوت و تغییر معنی دارا است. هر چند آنان ممکن است این تغییرپذیری را ناشی از خطاهای مشاهده، ناخالص بودن نمونه‌ها، عوامل فیزیکی حاکم بر محیط و شرایط قلمداد کنند، اما این تغییر وجود دارد و هرگاه این تغییرپذیری را با رویه‌هایی مثل معدل‌گیری منتفی نماییم، از عرصه شهود به عرصه ذهن قدم می‌گذاریم و به جای جهان واقعی، جهانی ساختگی بنامی کنیم.

«پیرسون» بر اساس تغییرپذیری پدیده‌ها، تئوری جدیدی را طرح کرد که بر طبق آن، مقوله پیوستگی و ارتباط را جایگزین ایده مرسوم علیت ساخت. وی نظریه جدید خود را با عبارات زیر بیان داشت:

اگر به تشخیص فردیت در ارکان کل هستی دست یابیم و همسانی را وابسته به حدود و ظرافت طبقه‌بندی بدانیم، درمی‌یابیم که علت و معلول... صرفاً حامل درجه‌ای از بار معنایی تشابه‌اند و نه تکرار مطلق. قانون علیت ساخته و پرداخته ذهن و مستخرج از پدیده‌ها است و بیانگر ماهیت واقعی آن‌ها نیست. مسئله اساسی فراروی بشر، چیزی بیش از «علیت» است و می‌توان آن را چنین خلاصه کرد: اگر «علت‌ها» مشابهتی با درجات مختلف دارند، «معلول‌ها» تا چه حد همسانی خواهند داشت؟ از دیدگاهی وسیعتر، چیزی را می‌توان علت یک پدیده دانست که پیش از آن یا همراهش می‌آید. اگر ما این علت را تغییر دهیم، تا چه حد به تغییر و تبدیل پدیده منجر می‌شود؟ اگر این تغییر، اثری بر آن پدیده نداشته، در این صورت استقلال مطلق حاکم است. اگر چنین برداشت کنیم که صرفاً تغییرپذیری علت، عامل دگرگونی پدیده بوده، در این صورت وابستگی مطلق حاکم است. این گونه از وابستگی مطلق یک پدیده به یک علت مشخص و سنجش‌پذیر، از استثنایا است... چنین تعامل و ارتباطی صرفاً در محدوده ذهنی مصداق دارد و در عینیت آن در جهان واقعی جداً شک داریم؛ اما نمود تمامی درجات ارتباط و پیوستگی بین دو محدوده - یعنی استقلال مطلق و وابستگی مطلق - امکان‌پذیر است.

«پیرسون» با مطرح کردن «مسئله بنیادی علم» در حیطه این نوع رده‌بندی و تغییرپذیری چنین نوشت: جهان از بشمار شئی تشکیل شده که هویت مجزا و غیر پایای فرضی دارند. همه می‌توانند با اندازه‌گیری و مشاهده خصیصه‌های این اشیا آن‌ها را در رده‌هایی با هویت‌های مشابه جای دهند. تغییرپذیری را در این رده‌ها می‌توان مورد توجه قرار داد و مسئله اساسی علم، یافتن چگونگی همبستگی و وابستگی در تغییر یک رده به رده دیگر است.

«پیرسون» این نظریه جدید را با ترسیم متغیر A در مقابل متغیر B در یک نمودار پراکنندگی نشان داد. نقاط روی نمودار به شکل عمومی یک منحنی پراکنده شده‌اند. بر طبق نظر وی، با عسکبرداری نمودار فوق از فاصله 50 یاردی یا با مشاهده آن از طریق تلسکوپ وارونه، یک فیزیکدان می‌تواند به بررسی آن پردازد. با کاربرد چنین روش‌هایی، نقاط پراکنش به صورت منحنی منظم درمی‌آیند و بدین گونه، مشاهده واقعی، جایگزین تابع ریاضی می‌شود. «پیرسون» ارتباط علیت با همبستگی را با استفاده از نمودار پراکنش چنین خلاصه کرد:

دو رده قابل سنجش از اشیا را در حیطه فیزیک، شیمی، اجتماعی یا اقتصاد به صورت ادراکی در نظر بگیرید. در این موارد هم، این نقاط و نمودار پراکنندگی وجود دارد. در بعضی موارد، این نقاط در سراسر صفحه پراکنده می‌شوند که در این صورت، همبستگی بین A و B وجود ندارد. در مواردی هم شکل نواری

گسترده دارند، که نشانگر ارتباطی متوسط است. وقتی نقاط به صورت ستاره دنباله دار درمی آیند، ارتباط بیشتر می شود. با این حال، تمام ردیف های نمودارها، پیوستاری است که نمی توان خطی به نشانه پایان همبستگی و شروع علیت در جایی از آن ها مشخص کرد. علیت صرفاً محدوده ذهنی همبستگی است که در حین باریک شدن بیش از حد نوار نمودار ظاهر می شود و شبیه منحنی به نظر می رسد.

«پیرسون» با انعکاس دیدگاه های خود در نظریه جدیدش، رویکرد قدیمی علت و معلول را این گونه نکوهش کرد:

... هر نوع تغییری در اجزای موجود یک رده با تغییری متناظر در بین اجزای رده دیگری مرتبط است. علم باید بتواند میزان شدت و ضعف این تغییرات همگام را اندازه گیری کند. در این محدوده ارتباط ذهنی، استقلال مطلق در یک انتها و وابستگی مطلق در انتهای دیگر آن قرار دارند. دیدگاه قدیمی علت و معلول در پی گنجاندن جهان تحت این دو محدوده ذهنی است. چنین چالشی محکوم به شکست است. در مشاهدات عینی ما، اشیا یا مستقل یا علی نمی توانند باشند.

با توجه به اینکه «پیرسون» هم در ضریب احتمال خود و هم در نسبت همبستگی خود به استقلال مطلق، عدد صفر و به وابستگی مطلق، عدد یک را تخصیص داده است، مسلماً به نامعین بودن محدوده بین صفر و یک، یعنی ناحیه نامعین میانگین خروج از عضویت توجه داشته است.

«داده های پرت» از دید «بارنت» و «لوئیس»¹⁵

نامعین بودن مجموعه های برادفورد به خاطر وجود سازوکار داده های پرت،¹⁶ روند استنباط آماری «پیرسون» را از طریق اندازه گیری تأثیر یک مجموعه بر دیگری پیچیده کرده است. «بک من»¹⁷ و «کوک»¹⁸ داده های پرت را یک مفهوم ذهنی و پس داده ای می دانند و «بارنت» و «لوئیس» نیز در کتابی آن را تایید کرده اند. آنان در این کتاب، داده های پرت را در مجموعه ای از داده ها چنین تعریف کرده اند: «مشاهده ای (یا زیرمجموعه ای از مشاهداتی) که با بقیه آن مجموعه از داده ها ناسازگار باشد». سپس آنان موضوع بحرانی مربوط به داده های پرت را این گونه پیش کشیدند:

عبارت «ناسازگار باشد» مهم و قابل توجه می باشد و مقوله ذهنی مربوط به مشاهده کننده است، چه یک مشاهده (یا مجموعه ای از مشاهدات) برای واریسی دقیق مشاهده مشخص شده باشند و چه نشده باشند. اما این که آیا بعضی مشاهدات، اعضای واقعی جمعیت اصلی اند، یا آلوده کننده¹⁹ (یعنی برخاسته از بعضی توزیع های دیگر) هستند، مهم می باشد. مورد اخیر ممکن است استنتاج های برگرفته از جمعیت اصلی را بی نتیجه سازد.

«بارنت» و «لوئیس» مسئله داده های پرت را به فرضیه های احتمال توزیع واقعی جمعیت، کاملاً مرتبط دانستند. مشاهده ای که ممکن است بر طبق فرضیه توزیع نرمال جزو داده های پرت باشد، در صورتی که مشاهده گر انتظار تویع کاملاً منحنی وار در داده های زیستی، اجتماعی و اطلاعاتی را داشته باشد، مسئله خاصی پیش نمی آورد. بنابراین آنان فرضیه های صفر²⁰ و کاری²¹ هر آزمون ناسازگاری برای داده های پرت را نوعی مدل اساسی احتمال برای ایجاد تمام داده ها، بدون توجه به داده های پرت می دانند. اگر شواهدی مهم برای رد فرضیه کاری پیدا شود، «بارنت» و «لوئیس» الگوهای «آلودگی» یا الگوهای ناشی از «داده های پرت»²² را تبیین

¹⁵ Barnet and Lewis

¹⁶ outliers

¹⁷ Beckman

¹⁸ Cook

¹⁹ contaminant

²⁰ null hypothesis

²¹ working hypothesis

²² outlier-generating

می‌کنند که ممکن است به عنوان فرضیه‌های جانشین وارد عمل شوند.

دو تا از این فرضیه‌ها بیشترین ارتباط را با این مقاله دارند. به اولین نوع این فرضیه‌ها نام «جانشین جبر گرایانه»²³ دادند که شامل مواردی از داده‌های پرت می‌باشد که از اشتباهات فاحش انسانی در اندازه‌گیری، ثبت و ضبط و ... ناشی شده‌اند. دومین واژه «بارنت» و «لوئیس» «جانشین ترکیبی»²⁴ نام دارد و بر آن است که نمونه تحت بررسی، دارای آلودگی‌هایی جدا از مدل اساسی جمعیت است و این اعضای «خارجی» نمونه یا همان آلاینده‌ها خود را در قالب داده‌های پرت می‌نمایانند. با تنزل سریع درجه عضویت در اعضای مجموعه برادفورد که به سرعت به سوی چنین آلودگی‌هایی پیش می‌روند، جانشین ترکیبی اهمیت بیشتری در تحلیل آماری داده‌های کتابخانه پیدا می‌کند.

تصویر عملی

اکنون در قالب داده‌های منتج از پروژه تغییر ساختار موجودی پایاندها در دانشگاه ایالتی لوئیزیانا،²⁵ تصویری از مفاهیم فوق را ارائه می‌کنم. در بخشی از مقدمات این پروژه، نیاز هیئت علمی بخش شیمی دانشگاه لوئیزیانا در آوریل 1993 به پایاندها مورد بررسی قرار گرفت. لازم به ذکر است که فقط نیاز هیئت علمی بخش شیمی مورد بررسی قرار گرفت و بخش‌های بیوشیمی و مهندسی شیمی بررسی نشدند. از هیئت علمی این دانشگاه خواسته شد بدون این که خود را محدود به پایاندهای مورد اشتراک دانشگاه کنند، پایاندهای مهم در امر تحقیق و تدریس خود را مشخص سازند. نشریات منتخب آنان بر اساس موضوعات مندرج در Science Citation Index Journal Citation Reports (SCIJCR)، در سال 1993 تقسیم‌بندی گردید (این نمایه‌نامه را «مؤسسه اطلاعات علمی»²⁶ منتشر می‌کند).

پایاندهایی که هیئت علمی بخش شیمی دانشگاه لوئیزیانا انتخاب کرده بودند و همراه با قانون پراکندگی برادفورد برطبق موضوعات متعدد (موجود در مؤسسه اطلاعات علمی) مرتب شده بودند، از جمله این موارد را شامل بودند: مهندسی برق و الکترونیک، علم محیط، علم زمین‌شناسی، علم مواد (سرامیک)، تغذیه و رژیم غذایی، فیزیک، رادیولوژی و پزشکی هسته‌ای.

با توجه به این امر، لازم بود نمونه را به پایاندهایی که «مؤسسه اطلاعات علمی» آن‌ها را در شاخه‌های مختلف شیمی طبقه‌بندی کرده بود (شامل مهندسی شیمی و کریستالوگرافی) محدود ساخت. استثنائاً موضوع اسپکتروسکوپی نیز به خاطر تأکید بخش شیمی دانشگاه، به این فهرست افزوده شد. هر چند این رشته عمدتاً بخشی از مبحث نور در فیزیک است. نتیجه نهایی نمونه‌ای شامل 154 مجله بود.

برای تعیین ارزش علمی این 154 مجله، سه متغیر کمی به کار گرفته شد: نمرات هیئت علمی دانشگاه لوئیزیانا، استنادهای کلی «نمایه‌نامه استنادی علوم» در سال 1993 و ضریب تأثیر این نمایه‌نامه در سال 1993. از این مقادیر فقط دو مورد اول معتبر شناخته شدند. امتیازات علمی دانشگاه لوئیزیانا به عنوان معیار اصلی اندازه‌گیری ارزش علمی به کار رفت، چرا که منطق مجموعه مجلات با ارزیابی بخش شیمی و نیز با دلایل منطقی معین شده بود. در این روش از مدرسین شیمی خواسته شد 10 عنوان مجله را نام ببرند و بیان کنند که آیا لازم است این مجلات مورد اشتراک دانشگاه باشند یا از طریق تحویل مدارک از راه دور در دسترس قرار گیرند؛ سپس عناوین مجلات را به ترتیب نزولی از 10 تا 1 بر اساس رتبه داده شده به آن‌ها مرتب کنند. یک عنوان بدین ترتیب رتبه‌بندی شد: 10 نمره به هر بار انتخاب یک عنوان، 10 نمره به این که آن مجله در آن دانشگاه وجود داشته باشد، و از 10 تا یک نمره بسته به رتبه‌ای که استاد به آن مجله داده بود. 25 استاد شیمی به

²³ uterministic alternative

²⁴ mixture alternative

²⁵ Louisiana State University (LSU)

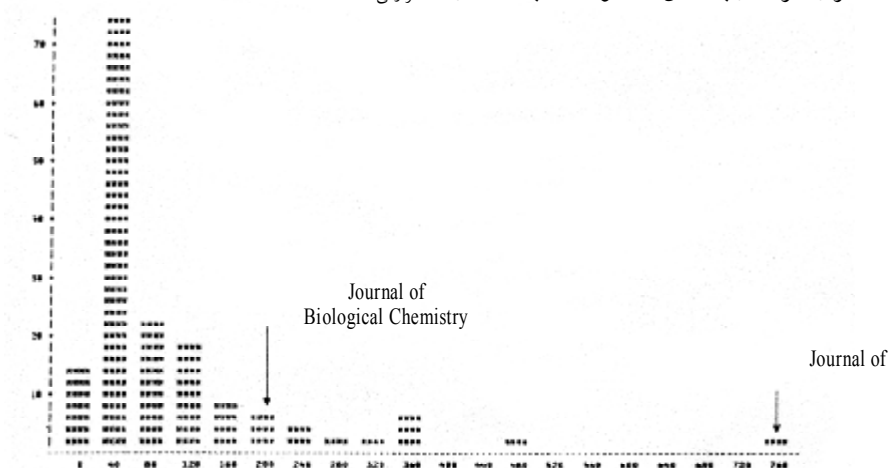
²⁶ Institute for Scientific Information (ISI)

این تحقیق پاسخ دادند و 154 مجله در نمره‌دهی هیئت علمی از 10 تا 755 امتیاز گرفتند (که 755 متعلق به کلی «نمایه‌نامهٔ استنادی علوم» تطبیق دادند تا میزان همخوانی نمره‌دهی اساتید دانشگاه لویزیانا با نظرات بخش انتشارات جامعهٔ علمی مشخص شود. در حین انجام این کار، با مشکل جدی «داده‌های پرت» روبرو شدیم. جدول 3، پراکنندگی دو متغیر پیشگفته را در دوره‌های مشخص شده بر حسب چارک نشان می‌دهد. برطبق این جدول با نوعی توزیع برادفورد سروکار داریم که در آن، انتهای فوقانی توزیع بیشترین ارزش را دارد. بنابراین در مجموعه ژئوفیزیک کاربردی برادفورد، 2/8 درصد از مجلات مختص 32/2 درصد از مراجع هستند و در اینجا، ردهٔ مربوط به چارک فوقانی، 62/5 درصد از 154 عنوان مجله را شامل می‌شود که ترکیبی از نمره‌دهی هیئت علمی و 80/2 درصد از استنادهای کلی آن‌ها است.

جدول 3. توزیع 154 عنوان مجله شیمی با ترتیب نزولی برطبق نمره‌دهی هیئت علمی و استناد کلی «نمایه‌نامهٔ استنادی علوم» در ردیف‌های مشخص‌شده بر حسب چارک

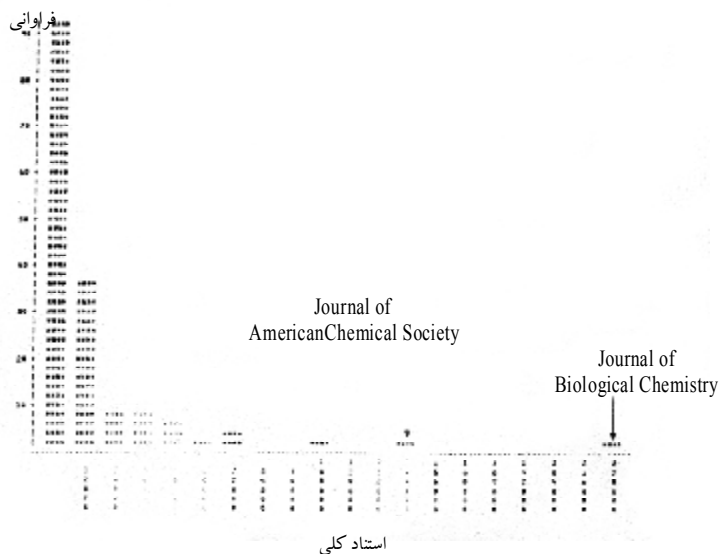
اسنادهای کلی «نمایهٔ استنادی علوم»		نمرات هیئت علمی		
درصد استنادهای کلی «نمایهٔ استنادی علوم» برای مجلات	محدودهٔ گروه چارک	درصدنمرات هیئت علمی	محدودهٔ گروه چارک	
80/2	11685-231324	62/5	111-755	بالا
13/4	3303-11586	20/6	50-110	بالای میانی
4/6	1533-3285	11/1	33-50	پایین میانی
1/8	255-1526	5/8	10-32	پایین

توزیع فراوانی نمرات هیئت علمی و استنادهای کلی، به صورت ترسیمی در نمودار 1 و 2 نمایش داده شده و به وضوح، شاهدی بر حضور آلایندهٔ جمعیتی غیر از آنچه براساس امتیازات اعضای هیئت علمی ترسیم شده است را نشان می‌دهد. آلاینده‌ها که یک پدیدهٔ طبیعی در توزیع برادفورد هستند، با مشاهدهٔ حد انتهای راست مشخص نیستند، اما در موقعیت‌های نسبی مشابه مثل Journal of the American Chemical Society (با نمرهٔ هیئت علمی برابر 755 و استناد کلی «نمایه‌نامهٔ استنادی علوم» برابر 148900) و Journal of Biological Chemistry (با نمرهٔ هیئت علمی برابر 197 و استناد کلی «نمایه‌نامهٔ استنادی علوم» برابر 231324) چنین نیست. در نمودار 1 که فراوانی توزیع 154 مجله بر حسب رتبه‌بندی هیئت علمی دانشگاه را نشان می‌دهد، موقعیت مجله Journal of the American Chemical Society در سمت راست با منطق مجموعه همسان است. هرچند در نمودار 2 نشان‌دهندهٔ توزیع فراوانی این مجلات بر حسب استناد کلی «نمایه‌نامهٔ استنادی علوم» است، مجله Journal of Biological Chemistry در سمت راست نشریه Journal of the American Chemical Society قرار گرفته و با منطق مجموعه مطابقت ندارد. فراوانی



نمرات اعضای هیئت علمی

نمودار 1. توزیع فراوانی 154 عنوان مجله در موضوع شیمی بر حسب نمره‌دهی اعضای هیئت علمی دانشگاه ایالتی لویزیانا

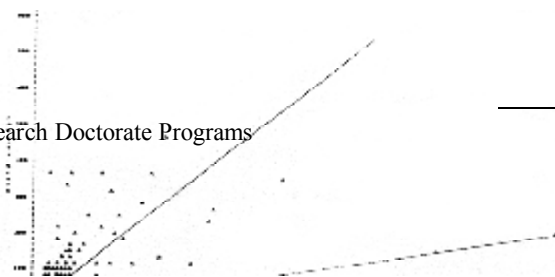


نمودار 2. توزیع فراوانی 154 عنوان مجله در موضوع شیمی بر حسب استنادهای کلی «نمایه‌نامه استنادی علوم»

مجله Journal of Biological Chemistry جزو «داده‌های پرت» محسوب می‌شود. این پیش‌بینی در نمودار 3 مورد تأیید قرار گرفته است که توزیع پراکندگی نمرات هیئت علمی را در برابر استنادهای کلی «نمایه‌نامه استنادی علوم» نشان می‌دهد. در این شکل یک خط رگرسیون فرضی از مجله Journal of the American Chemical Society تا مبدأ ترسیم شده که از نقاط میانی می‌گذرد، در حالی که خط ترسیم شده از مجله Journal of Biological Chemistry تا مبدأ در قسمت پایین و سمت راست تمام نقاط قرار دارد. وقتی که این مجموعه را ساختم، تصمیم گرفتم که بیوشیمی را زیرشاخه‌ای از شیمی قلمداد کنم. تصمیم من در این زمینه از طرح رده‌بندی کتابخانه کنگره تأثیر پذیرفته بود که در آن، بیوشیمی به عنوان زیرمجموعه‌ای از شیمی آلی در شیمی است. با وجود این، موقعیت مجله Journal of Biological Society مرا به شک واداشت. پژوهش بعدی این سوءظن را تأیید کرد. برخلاف طرح رده‌بندی کنگره، در رده‌بندی دهدهی دیوئی، بیوشیمی نه تنها شاخه‌ای از شیمی نیست، بلکه به عنوان زیرشاخه زیست‌شناسی و علوم زیستی تقسیم‌بندی شده است. نه تنها هیئت علمی دانشگاه لویزیانا بخش‌های شیمی و بیوشیمی را از هم جدا کرده‌اند، حتی در آخرین رتبه‌بندی برنامه‌های تحقیقاتی دکترای در انجمن ملی تحقیقات آمریکا²⁷ نیز، شیمی زیر سرفصل علوم فیزیکی و ریاضیات رده‌بندی شده است و بیوشیمی ضمن ترکیب با زیست‌شناسی مولکولی، زیر سرفصل علوم زیستی رده‌بندی می‌شود. بنابراین، Journal of Biological Society همراه با تعدادی از مجلات بیوشیمی دیگر در مجموعه موردنظر من، نتیجه نامعین مجموعه‌های برادفورد است. نشریه فوق هم A و هم نه A، هم شیمی و هم بیوشیمی است.

Journal of
AmericanChemical Society

²⁷. US Research Doctorate Programs



روش‌های مهار کردن داده‌های پرت

«بارنت» و «لوئیس» روش‌های مهار کردن داده‌های پرت را در چهار گروه کلی بیان کرده‌اند: قوانین محکم و صریحی مبنی بر این که از کدام گروه از روش‌ها استفاده شود، وجود ندارد؛ بلکه این امر به چگونگی پدید آمدن داده‌های پرت و هدف فرد بستگی دارد. «بارنت» و «لوئیس» یکی از مقوله‌هایشان را «رد»²⁸ نامیدند. منظور آن دو، عبارت بود از این که اگر داده‌های پرت قابل تصحیح نباشد، فرد آن‌ها را از دور خارج می‌کند و سپس باقی نمونه را تحلیل می‌نماید. هنگامی که من پنج مورد از داده‌های پرت را در حین تعیین همبستگی تولید - زمان «پیرسون» بین رتبه‌بندی هیئت علمی و استنادهای کلی «نمایه‌نامه استنادی علوم» پیدا کردم، چنین موردی پیش آمد. تحلیل بقیه موارد، پنج داده پرت را (که از بین آن‌ها چهار مورد از نظر هیئت علمی نسبت به استنادهای کلی، رتبه پایین داشتند) ایجاد نمود که دو مورد از داده‌های پرت اخیر را «مؤسسه اطلاعات علمی» در بیوشیمی مولکولی رده‌بندی کرده است. ضریب همبستگی اولیه 66 درصد بود که پس از حذف داده‌های پرت از نمونه، این ضریب به 72 درصد رسید.

دومین گروه در روش‌های مهار کردن داده‌های پرت برطبق نظر «بارنت» و «لوئیس»، «تعیین هویت»²⁹ نامیده شده. منظور آن‌ها این بود که داده‌های پرت ناسازگار، علائمی از بعضی عوامل مفروض در حین کار بر روی جمعیت تحت تجزیه و تحلیل هستند. من هم این کار را انجام دادم و به نتایج جدیدی در تبیین رابطه بیوشیمی با شیمی رسیدم. هر چند ابتدا فکر کرده بودم بیوشیمی شاخه یا زیرمجموعه‌ای از شیمی است، اکنون آن را به عنوان مجموعه‌ای جدا با الگوهای آماری مربوط به خود محسوب می‌کنم.

گروه دیگر در روش‌های مهار داده‌های پرت از نظر «بارنت» و «لوئیس»، «الحاق»³⁰ نام دارد. در این روش، هدف آن است که یک الگوی همگن را با الگوی همگن دیگر برای کل نمونه‌ها جایگزین کنیم (یکپارچه‌سازی داده‌های پرت) تا هیچ مشاهده‌ای دارای تناقض نباشد.

چهارمین و آخرین مقوله در روش‌های مهار کردن داده‌های پرت که «بارنت» و «لوئیس» آن را پیشنهاد کردند، «سازگاری» نام دارد. آنان این مقوله را به دو جزء تقسیم کردند: اولین جزء روش‌های باثبات هستند که ارزش منطقی را در برابر داده‌های پرت حفظ می‌کنند. یک نمونه از این روش‌ها را می‌توان برای آزمون استقلال مربع کای³¹ - به جای استفاده از روش‌های همبستگی - برای تحقیق درباره ارتباط نمره‌دهی هیئت علمی دانشگاه لویزیانا با استناد کلی «نمایه‌نامه استنادی علوم» به کار برد. آزمون استقلال مربع کای را، «کارل پیرسون» اساساً برطبق احتمال پیشنهاد کرد. از آنجا که روش‌های همبستگی مستلزم کاربرد ریاضیات است که هم تناسب نقاط داده‌ها را با خط رگرسیون در رابطه همبستگی ضریب گشتاوری پیرسون و هم رابطه یک رتبه با رتبه دیگر را در همبستگی رتبه‌بندی اسپیرمن، تعیین می‌کند، بنابراین وابستگی متغیرها به یکدیگر در گروه‌های بزرگ با استفاده از آزمون استقلال مربع کای امکان‌پذیر است. این امکان در جدول 3 نشان داده

²⁸ re-jection
²⁹ identification
³⁰ incorporation
³¹ chi-square

شده است که در آن، دو مجله مذکور برحسب هر دو اندازه کیفیت علمی در چارک بالایی قرار گرفته‌اند. جزء دیگر الحاق، حاوی آن روش‌هایی است که با تخصیص ارزش کمتر به مقادیر انتهایی در مقایسه با دیگر اعضای نمونه، داده‌های پرت را مهار می‌کنند. یک چنین روشی می‌تواند winsovization باشد، که در آن مشاهدات با نزدیکترین مجاور آن جایگزین می‌شود. در این روش اسنادهای کلی «نمایه‌نامه استنادی علوم» از مجله Journal of Biological Chemistry به اندازه اسنادهای کلی همین نمایه‌نامه از مجله Journal of American Chemical Chemistry است.

با این حال، «الحاق» شاید روش بهتری باشد؛ چرا که نظریه مجموعه‌های نامعین را برای مهار داده‌های پرت به کار می‌گیرد. کاربرد این نظریه در طبیعت عملی است و به منطق مجموعه تحت بررسی و نیز هدف پژوهش بستگی دارد. هدف عمده از این کار، سازگار کردن داده‌های پرت به تناسب عضویت آن‌ها در مجموعه است. در مثال فوق یکی از مصادیق نشان‌دهنده این امر آن است که نمره‌دهی هیئت علمی را به صورت تجربی، استخراج و از نتایج میزان عضویت برای انطباق با اسنادهای کلی «نمایه‌نامه استنادی علوم» استفاده کنیم. یک روش دیگر، تجزیه و تحلیل پایگاه «مؤسسه اطلاعات علمی» و محدود ساختن اسنادها به انواع متناسب با منطق مجموعه است.

بر این اساس، نظریه احتمالات و نظریه مجموعه‌های نامعین متناقض نیستند، اما قالب‌هایی مکمل برای تجزیه و تحلیل به حساب می‌آیند.

منابع و یادداشتها

Bradford, S. C. "Sources of Information on Specific Subjects". **Engineering** 137:85-86 (1934).

A full treatment of the various formulations of Bradford's Law is outside the scope of this paper.

Kosko, B. **Fuzzy Thinking: The new Science of fuzzy Logic**. London: Flamingo, 1994, p.6.

Bradford, S. C. "Some General Principles of a Bibliographical Classification Scheme, with Application to the Universal Decimal Classification". **Proceedings of the British Society for International Bibliography** 6 (3): 57-69 (1944).

Zadeh, L.A. "Fuzzy Sets". **Information and Control** 8 (3): 338-353 (1965).

Boyce, B.R., C.T. Meadow. And D.H. Kraft. **Measurement in Information Science**. San Diego: Academic Press, 1994, p. 95.

Bradford, S.C. (1948). "Complete Documentation" in **The Royal Society Empire Scientific Conference, June-July 1946**: Report. London: The Royal Society, 1948, Vol. 1, pp. 729-748.

Pearson, K. **The Grammar of Science**. 3rd ed., rev. and enl. London: A. and C. Black, 1911. Pt. 1, p. 165.

Beckman, R. J. and R.D. Cook. "Outliers". **Technometrics** 25 (2): 119-149 (1983).

Barnett, V. and T. Lewis. **Outliers in Statistical Data** 3rd ed. Chichester: J. Wiley, 1995. p. 7. A short, general review of outliers and the procedures for handling them is: Barnett, V. "The Study of Outliers: Purpose and Model". **Applied Statistics** 27 (3): 242-250 (1978)

Bensman, S.J. "The Structure of the Library Market for Scientific Journals: The Case of Chemistry". **Library Resources & Technical Services** 40 (2): 145-170 (1996): Bensman, S.J. And S.J. Wilder "Scientific and Technical Serials Optimization in an Inefficient Market: A LSU Serials Redesign Project Exercise". **Library Resources &**

Technical Services 42 (3): 147-242.

Science Citation Index Journal Citation Reports: A Bibliometric Analysis of Science Journals in the ISI Database: 1993. Philadelphia: Institute for Scientific Information, 1994.

Glodberger, M.L., B.A. Maher, and P.E. Flattau, eds. **Research Doctorate Programs in the United States: Continuity and Change.** Washington, D.C.: National Academy Press, 1995.