

ارتباط آماری برخی شاخص‌های اساسی کتابسنجی در پژوهش‌های علم‌سنجی^۱

نوشته علی اوزن^۲
ترجمه فرشید دانش^۳
مریم ریاضی‌پور^۴

چکیده

مقاله حاضر نتایج پژوهشی در مورد ارتباطات آماری میان پدیدآورندگی برحسب تعداد مؤلفان، تعداد واژگان/اصطلاحات مهم در عنوان مقالات، تعداد منابع استناد شده، و تأثیر استنادی مجموعه‌ای متشکل از ۴۶۷ مقاله منتشر شده در مجله بین‌المللی <علم‌سنجی>، بین سال‌های ۱۹۹۹ تا ۲۰۰۳، را ارائه می‌کند. تحلیل‌ها نشان می‌دهد که کنترل رشد پویای استنادها و میانگین تأثیر استنادی مقالات به شدت به پدیدآورندگی بستگی دارد که خود از طریق تعداد مؤلفان به ازاء هر مقاله سنجیده می‌شود. آزمون مجذورکی مشخص کرد که تعداد منابع استناد شده در مقالات و تأثیر استنادی آنها به هم وابسته‌اند. همچنین درجه پیوستگی خطی نسبتاً بالایی بین تعداد دفعات استناد شده به مقاله و تعداد منابع موجود در آن دیده می‌شود. در ادامه، تحلیل‌ها نشان داد که در هیچ‌یک از مراحل تجمع، پیوستگی آماری معنی‌داری بین تأثیر استنادی و تعداد واژگان مهم در عناوین مقالات وجود ندارد.

کلیدواژه‌ها

کتابسنجی، علم‌سنجی، تأثیر استنادی، واژگان عنوان، پدیدآورندگی

مقدمه

این موضوع بود که میانگین تأثیر استنادی مقالات با افزایش تعداد منابع سیاهه شده افزایش می‌یابد (۸). عموماً فرض بر این است که پژوهش‌های مشارکتی (گروهی) مشهودتر و چشمگیرتر از پژوهش‌های تک نویسنده‌ای

نخستین بار در دهه ۱۹۶۰ این پرسش مطرح شد که آیا میان دفعات استناد به مقاله و تعداد منابع آن مقاله همبستگی وجود دارد؟ (۳). در پژوهشی ارقام ارائه شده منعکس‌کننده

1. "Statistical relationship of some basic bibliometric indicators in scientometric research". In *Proceedings of International Workshop on Webometrics, Infometrics and Scientometric & seventh COLLNET Meeting*, Nancy (France), 2006.

2. Ali Uzun

۳. عضو هیئت علمی گروه کتابداری و اطلاع‌رسانی پزشکی دانشگاه علوم پزشکی اصفهان
۴. دانشجوی کارشناسی کتابداری و اطلاع‌رسانی پزشکی دانشگاه علوم پزشکی اصفهان

است، اما این موضوع از طریق تأثیر استنادی یا سایر شیوه‌ها سنجیده نشده بود (۲؛ ۷). بیشتر پژوهش‌ها پیرامون طول واژگان عنوان مقالات مجلات تدوین شده‌اند. اما هیچ پژوهشی در مورد ارتباط میان عناوین و استنادهای مقالات انجام نشده بود (۵؛ ۶). انگیزه نوشتن این مقاله پاسخ به سه سؤال زیر است:

۱. آیا ارتباط آماری معنی‌داری میان تأثیر استنادی و تعداد نویسندگان مقالات وجود دارد؟
۲. آیا ارتباطی میان تأثیر استنادی و تعداد منابع استناد شده در مقالات پژوهشی وجود دارد؟
۳. آیا ارتباطی میان تأثیر استنادی و تعداد کلمات مهم در عناوین مقالات وجود دارد؟

روش پژوهش

با استفاده از وب آو ساینس^۵، مجموع ۴۶۷ مقاله از مجله <علم‌سنجی>، در یک دوره ۵ ساله (۱۹۹۹-۲۰۰۳) جمع‌آوری شد و مورد بررسی قرار گرفت. در هر مقاله، تعداد نویسندگان، تعداد منابع آن، تعداد واژگان یا اصطلاحات- عبارات اسمی در عناوین- و تعداد دفعاتی که بین سال انتشار و سال ۲۰۰۵ به مقاله استناد شده، در نظر گرفته شد. باید خاطر نشان کرد که مقالات انتشار یافته در سال ۲۰۰۳، جدیدترین مقالاتی است که در نظر گرفته‌ایم و از آنجاکه معمولاً بیشترین تعداد استنادهای سالانه به اثری حدوداً ۳ سال پس از انتشار آن روی می‌دهد، انتظار می‌رود که بیشترین تعداد استنادها در سال ۲۰۰۵ به‌دست آید (۱).

باید توجه داشت که ارتباط آماری برای گروه‌های انتشارات قابل اجراست. ۴۶۷ مقاله براساس تعداد نویسندگان، منابع استناد شده، و واژگان عنوان دسته‌بندی شده‌اند. میانگین تأثیر استنادی مقالات در هر گروه به‌منظور تعیین پویایی رشد از سال ۱۹۹۹ تا ۲۰۰۵ جمع و تنظیم گردیده است. آزمون‌های مجذور کی و روش حداقل مربعات به‌منظور یافتن ارتباطات آماری میان تأثیر استنادی، پدیدآورندگی، منابع استناد شده، و واژگان عنوان مقالات به‌کار گرفته شده‌اند. این بررسی دو محدودیت عمده دارد: نخست از نظر پوشش، زیرا فقط مقالات پژوهشی (از مقالات مروری، نامه‌ها، و یادداشت‌ها صرف‌نظر شده است) مجله بین‌المللی <علم‌سنجی> را پوشش می‌دهد؛ و دوم دوره زمانی مورد بررسی است، زیرا پژوهش حاضر به یک دوره ۵ ساله از سال ۱۹۹۹ تا ۲۰۰۳ به منظور شمارش مقالات، و از ۱۹۹۹ تا ۲۰۰۵ برای شمارش تعداد استنادهای آنها محدود شده است. عموماً این موضوع قابل پذیرش است که آمارهای استنادی که در بازه زمانی کمتر از سه سال تهیه شده باشند به اندازه کافی معتبر نخواهند بود (۴). بازه زمانی ۱۹۹۹ تا ۲۰۰۵ برای جمع‌آوری تعداد زیادی از استنادها، حتی برای مقالات منتشر شده در سال ۲۰۰۳، کافی به نظر می‌رسد.

تأثیر استنادی و منابع مقالات

جدول ۱ دسته‌بندی ۴۶۷ مقاله، براساس تعداد منابع آنها و تعداد استنادهای دریافت شده در فاصله زمانی سال انتشار و سال ۲۰۰۵

اشاره می‌کند که تعداد اسنادها و تعداد منابع مقالات مستقل هستند و با یکدیگر ارتباط معنی‌داری ندارند. باید خاطر نشان کرد که توزیع مجذور کی تخمین خوبی

را نشان می‌دهد. فراوانی‌های قابل انتظار در داخل پراترها در جدول ۱ مشاهده می‌شود. این فراوانی‌ها با استفاده از قانون مجموع سلول‌ها (خانه‌ها) به دست آمده است:

$$\text{مجموع سلول‌های ستون} \times \frac{\text{مجموع سلول‌های سطر}}{\text{مجموع کل}} = \text{مجموع سلول‌های مورد انتظار}$$

جدول ۱. فراوانی مقالات مشاهده شده و مورد انتظار در دوره زمانی ۱۹۹۹-۲۰۰۵

تعداد مقالات	تعداد اسنادها				تعداد منابع
	۵ یا بیشتر	۴-۳	۲-۱	۰	
۱۳۹	۲۹ (۳۸/۳)	۱۶ (۲۱/۴)	۳۶ (۴۲/۹)	۵۸ (۳۶/۴)	۱۰-۰
۱۶۱	۴۴ (۴۴/۵)	۲۹ (۲۴/۹)	۵۰ (۴۹/۶)	۳۸ (۴۲)	۲۰-۱۱
۸۵	۲۹ (۲۳/۵)	۱۱ (۱۳/۱)	۳۱ (۲۶/۲)	۱۴ (۲۲/۲)	۳۰-۲۱
۸۲	۲۷ (۲۲/۷)	۱۶ (۱۲/۶)	۲۷ (۲۵/۳)	۱۲ (۲۱/۴)	۳۱ یا بیشتر
*۴۶۷	۱۲۹	۷۲	۱۴۴	۱۲۲	تعداد مقالات

* ارقام درون پرانتز فراوانی‌های مورد انتظار هستند.

برای محاسبه داده‌های جمع شده می‌باشد. مقدار هر سلول (خانه) حداقل ۵ است. در این نمونه کمترین مقدار سلول ۱۱ است و بنابراین قانون مجموعه سلول‌ها گره‌گشاست. درجه آزادی مجذور کی طبق فرمول برابر $v = (r-1)(c-1) = (4-1)(4-1) = 9$ می‌باشد که r و c به ترتیب تعداد سطرها و ستون‌ها در جدول احتمال 4×4 است. با استفاده از ارقام موجود در جدول ۱، ارزش آزمون آماری $\chi^2 = 24/6$ به دست آمده است. نقطه بحرانی برای سطح معنی‌داری $0/01$ با درجه آزادی ۹، معادل

آزمون مجذور کی برای گروه‌هایی که در سطرهای مستقل بودند (تعداد منابع) و گروه‌های ستونی (تعداد اسنادها در سال) انجام شده است. آزمون آماری مذکور توسط این معادله به دست می‌آید:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

برای هر سلول ویژه، f_o فراوانی مشاهده شده و f_e فراوانی مورد انتظار را نشان می‌دهد. میزان بالای مجذور کی فرضیه صفر را رد می‌کند. فرضیه صفر به این نکته

۲۱/۷ (۲۱/۷=۰/۰۱,۹ χ^2) می‌باشد. بنابراین تأثیر استنادی مقالات گروه‌بندی شده در جدول ۱ به‌طور قابل توجهی با تعداد منابع موجود در آنها مرتبط است.

بر اساس نتیجه‌گیری اخیر فرض می‌کنیم که بین میانگین تعداد منابع در هر مقاله (متغیر مستقل X) و تعداد دفعاتی که به آنها استناد شده است (متغیر وابسته Y) ارتباط خطی تقریبی وجود دارد که در جدول ۲ آمده است. با نگاهی به متغیرهای X و Y در جدول مشاهده می‌کنیم که فرض ارتباط خطی میان X و Y به نظر منطقی می‌رسد. معادله خط را با داده‌های جدول ۲ تطبیق دادیم.

$Y = A + BX$ (B و A) برآورد حداقل مربعات بودند که با استفاده از مقادیر X و Y محاسبه شدند. در نتیجه معادله‌ای که بهترین امکان تطبیق را فراهم می‌آورد، چنین بود:

$$Y = 0.394 + 5/1 \times 10^{-3} X$$

باید خاطر نشان کرد در این محاسبه آماری به هیچ الگوی علت و معلولی ویژه‌ای میان

متغیرهای X و Y اشاره نشده است. تخصیص «متغیر وابسته» و «متغیر مستقل» هیچ دلالت ضمنی مبنی بر اینکه تغییرات X موجب تغییرات Y شود ندارد. ضریب همبستگی که میزان ارتباط خطی را بین متغیرها اندازه می‌گیرد 0.79 ($r = 0.79$) به دست آمد.

تأثیر استنادی و پدیدآوردگی

به منظور بررسی ارتباط آماری میان تأثیر استنادی و تعداد نویسندگان مقالات و یا امکان وجود ماهیت این رابطه، ۴۶۷ مقاله از طریق پدیدآوردگی و میزان استنادهای آنها دسته‌بندی شده‌اند. نتایج این دسته‌بندی در جدول ۳ ارائه شده است.

با نگاهی به ارقام ستون‌های اول و آخر جدول ۳، مشاهده می‌شود که ارتباط خطی مثبتی میان دو متغیر (تعداد نویسندگان به ازاء هر مقاله X) و میانگین تعداد استنادها به ازاء هر مقاله در سال Y وجود دارد. به منظور سنجش ارتباط و شکل عملکرد آن از روش حداقل مربعات، همانند بخش قبلی، استفاده

جدول ۲. دسته‌بندی منابع و استنادهای مقالات در دوره زمانی ۱۹۹۹ - ۲۰۰۵

تعداد منابع	تعداد مقالات	میانگین تعداد منابع در هر مقاله (X)	میانگین تعداد استنادها در هر مقاله (Y)
۱۰-۰	۱۳۷	۴	۰/۴۱*
۲۰-۱۱	۱۶۶	۱۴	۰/۴
۳۰-۲۱	۸۶	۲۲	۰/۶
۳۱ یا بیشتر	۷۸	۴۵/۴	۰/۶
در مجموع	۴۶۷	۱۷/۸**	۰/۵۴**

* میانگین‌ها در این ستون برای تعداد مقالات در هر دسته از منابع اندازه‌گیری شده، و به منظور تعیین پویایی رشد استنادها تطبیق داده شده‌اند. این تطبیق با استفاده از میانگین نرخ رشد سالانه حدود ۰/۲ عامل تأثیر مجله «علم‌سنجی» از سال ۱۹۹۹ تا ۲۰۰۵ بوده است.

** میانگین مجموع نیز اندازه‌گیری شده است.

جدول ۳. تعداد نویسندگان و تعداد استنادها به مقاله به طور سالانه برای سال‌های ۱۹۹۹-۲۰۰۵

تعداد نویسندگان به ازاء هر مقاله	تعداد مقالات	میانگین تعداد استنادها به ازاء هر مقاله در سال (تعدیل شده)
۱	۱۹۵	۰/۴۶
۲	۱۴۵	۰/۵۵
۳	۸۹	۰/۶۰
۴ یا بیشتر	۳۸	۰/۶۷
مجموع	۴۶۷	* ۰/۶۷

* نگاه کنید به یادداشت ** و *** در جدول ۲

شد. نتیجه، معادله $Y = 0/4 + 6/8 \times 10^{-2} X$ است. تحلیل‌هایی که درباره رابطه علت و معلولی بین تعداد منابع و میزان استنادها در قسمت قبل ارائه شد، با پدیدآوردگی و استنادها در این بخش نیز صدق می‌کند.

تأثیر استنادی و واژگان عنوان

جدول ۴ طبقه‌بندی دوسویه‌ای از ۴۶۷ مقاله، با در نظر گرفتن تعداد واژگان مهم در عناوین و تعداد استنادهای دریافت شده از شروع سال

انتشارشان تا سال ۲۰۰۵ را نشان می‌دهد. آزمون مجذورکی برای عدم وابستگی گروه‌بندی واژه و استناد با استفاده از داده‌های جدول ۴ انجام شد. مقدار آماری مجذورکی محاسبه شده برابر $(\chi^2 = 16)$ است و ارزش آماری جدول نیز با سطح معناداری ۰/۰۱ و درجه آزادی ۶، برابر $16/81 = 0/1975$ می‌باشد. با توجه به مقادیر فوق و داده‌های جدول ۴ می‌توان نتیجه گرفت که تعداد واژگان مهم عناوین و تأثیر استنادی مقالات مجله از نظر آماری مستقل هستند.

جدول ۴. تعداد واژگان مهم در عناوین و میزان استنادهای مقالات در دوره زمانی ۱۹۹۹-۲۰۰۵

تعداد مقالات	تعداد استنادها				تعداد واژگان
	۷ یا بیشتر	۴-۶	۱-۳	۰	
۱۱۶	۲۵ (۲۱)	۲۰ (۲۴)	۲۷ (۴۲)	۴۴ *(۲۹)	۵-۱
۲۶۶	۴۴ (۴۸)	۵۵ (۵۶)	۱۱۱ (۹۶)	۵۶ (۶۶)	۱۰-۶
۸۵	۱۶ (۱۶)	۲۳ (۱۸)	۳۰ (۳۰)	۱۶ (۲۱)	۱۱ یا بیشتر
۴۶۷	۸۵	۹۸	۱۶۸	۱۱۶	تعداد مقالات

* نگاه کنید به یادداشت جدول ۱

بیان نتایج

بررسی‌ها مشخص کرد که تعداد منابع موجود در مقالات پژوهشی و میانگین تأثیر استنادی آنها مستقل نیستند. تحلیل با استفاده از روش حداقل مربعات انجام گرفت و نتایج نشان داد که تعداد منابع به عنوان متغیر مستقل و تأثیر استنادی (تعدیل شده) به عنوان متغیر وابسته به طور خطی و با ضریب همبستگی $(r = 0.799)$ ارتباط دارند. علاوه بر این، تعداد نویسندگان مقالات و تأثیر استنادی آنها همبستگی مثبت قوی‌تری را نشان داد $(r = 0.994)$. دامنه تعداد واژگان مهم در عناوین مقالات بین ۲ تا ۲۳ می‌باشد و دامنه تعداد دفعاتی که آنها مورد استناد قرار گرفته‌اند بین ۰ (صفر) تا ۵۰ است. جالب اینکه عنوان مقاله دو کلمه‌ای در مجموع ۵۰ استناد دارد، در حالی که تعداد زیادی از مقالات با ۱۵ تا ۲۳ واژه مهم در عنوان، هیچ استنادی نداشته‌اند. از آزمون مجذور کی برای فرضیه مستقل بودن واژگان عنوان و میزان استنادها استفاده شد و در نتیجه این فرضیه مورد پذیرش قرار گرفت. پس براساس تعداد ۴۶۷ مقاله می‌توان نتیجه گرفت که متغیرهای بعدی وابسته نبوده و از این رو هیچ همبستگی بین آنها امکان‌پذیر نیست.

این نکته بسیار مهم است که ارتباطات آماری بررسی شده فوق در این تحلیل، صحت ارتباطات گذشته را تأیید کرد. این ارتباطات می‌توانند به عنوان پایه‌ای برای پیش‌بینی‌ها از طریق قیاس باشد. فرض بر این است که انتشارات و استنادهای علمی نویسندگان در آینده پایدار خواهد ماند.

همچنین باید توجه داشت که تأثیر استنادی مقالات مجله به وسیله عوامل زیادی تعیین می‌شوند. سه عامل پدیدآورندگی، ارجاعات

کتابشناختی، و واژگان عنوان که در این بررسی از مهم‌ترین عوامل اند. برای پژوهش‌های آینده محققان می‌توانند با استفاده از متغیرهای بیشتر و از طریق رگرسیون چند متغیره، واریانس‌های مشاهده شده در تأثیر استنادی را با جزئیات بیشتر مورد بررسی قرار دهند.

منابع

1. Adams, J. "Early citation counts correlate with accumulated impact". *Scientometrics*, No.63 (2005).
2. Beaver, D. deB. "Collaboration and team work in physics". *Czechoslovak Journal of Physics*, No. 36 (1968).
3. De Solla Price, Derek J. "Network of scientific papers". *Science*, No.149 (1956).
4. Erkut, E. "Measuring Canadian Business School Research Output and Impact". *Canadian Journal of Administrative Sciences*, No.19 (2002).
5. Lewison, G. Hartly, J. "What's in a title? Numbers of words and the presence of colons". *Scientometrics*, No.6 (2005).
6. Lipetz, B. A. "Aspects of JASIS authorship through five decades". *Journal of the American Society for Information Science*, No.50 (1999).
7. Rigby, J.; Edler, J. "Peering inside research networks: some observations on the effect of the intensity of collaboration on the variability of research quality". *Research Policy*, No.34 (2005).
8. Vinkler, P. "Dynamic changes in the chance for citedness". *Scientometrics*, No.54 (2002).