چگونه باید نوشته‌های مربوط به روشهای آماری را مطالعه کرد: راهنمایی برای دانشجویان

جیمز مورفی

ترجمهٔ افسانه یزدانی و ترانهٔ آبرین*

چکیده

مقاله‌های مربوط به روشهای آماری به صورت فشرده نوشته می‌شوند. نویسندگان آنها فرض می‌کنند که خوانندگان در زمینهٔ موضوع مورد بحث تبحّر دارند. به علاوه نمادگذاریهای آماری استاندارد هنوز به وجود نیامده‌اند. دانشجویانی که تکنیکی را بر اساس یک نوع نمادگذاری فرا می‌گیرند، ممکن است در خواندن مقاله‌هایی که در نمادگذاریهای دیگر نوشته شده‌اند، دچار سردرگمی شوند. این مقاله شامل پیشنهادهایی برای آسانتر کردن کار دانشجویان و پربار کردن آن است.

۱ مقدمه

راهنماهای متعددی وجود دارند که به غیر آماردانان می‌گویند چگونه نتایج آماری کاربردی را بخوانند و تعبیر و تفسیر کنند. هاف (۱۹۴۵) پنج نکته را متذکّر می‌شود که خوانندگان شکّاک باید به خاطر بسپارند. ساکت (۱۹۹۱) و کولتون (۱۹۷۹) هر یک سیاهه‌هایی را تهیه کرده‌اند که مشخص می‌کنند آیا روشهای آماری به طرز مناسبی در مقاله‌های پزشکی به‌کار رفته‌اند یا خیر. همچنین راهنماهایی برای مطالعهٔ مقاله‌هایی که شامل ریاضیات پیچیده می‌شوند، وجود دارند که کوون (۱۹۹۱)، فانسل تی‌یل (۱۹۹۰)، پارک (۱۹۵۸)، و پمبرتون (۱۹۹۶) از آن جمله‌اند. شختمان (۱۹۸۷) پیشنهاد می‌کند که آمار زیستی باید از طریق مطالعهٔ منابع پزشکی آموزش داده شود. تمام مراجع بالا، اطلاعات مفیدی در خصوص نحوهٔ مطالعهٔ مطالب تکنیکی در اختیار می‌گذارند. با این حال مسائل خاصی در مطالعهٔ نوشته‌های روشهای آماری وجود دارند. در این مقاله طرحی از یک روش کلی برای سر و سامان دادن به چنین مطالعاتی ارائه می‌شود.

در مطالعهٔ نوشته‌های آماری خواننده با چالشهای متعددی مواجه می‌شود. اول اینکه داشتن مهارتهای مختلفی برای مطالعهٔ روشهای آماری ضروری است: مهارتهای اساسی در زبان، دانستن نمادگذاریهای آماری، مهارتهای جبری، و به طور روز افزون تشخیص اینکه رایانه‌ها چگونه کار می‌کنند. دوم اینکه مقاله‌های تکنیکی به ندرت الگوهایی از سبک نگارش مقاله‌های توصیفی‌اند. این مقاله‌ها به شدت به اصطلاحات فنی و تعامل بین نوشتار کلماتی و نوشتار نمادی وابسته‌اند. مفهومی که توصیف آن به زبان عادی بسیار مشکل است، گاهی با استفاده از نمادها بسیار ساده‌تر است، اما ناهماهنگی بین آنها ممکن است فهم مقاله را مشکل کند. سوم اینکه در مقاله‌های پیشرفته فرض می‌شود که نیازی به شرح مفاهیم پایه‌ای نیست. چهارم اینکه با نمادهایی که آنها را در کلاسهای درس یاد گرفته‌ایم، احساس الفت بیشتری می‌کنیم و ممکن است نمادگذاری متفاوت موجب سردرگمی شود. سرانجام، در بسیاری از مقاله‌ها، هم نظریهٔ آماری و هم روشهای محاسباتی لازم برای به مرحلهٔ عمل درآوردن، مورد بحث قرار می‌گیرد. نظریه و محاسبات همواره به صورت روشنی از هم تفکیک نمی‌شوند.

هر آماردانی باید در مطالعهٔ مقاله‌های آماری با این مشکلات دست و پنجه نرم کند، اما راه حلهای فردی، به صورت رسمی منتقل نمی‌شوند.

*) کارشناسان مرکز آمار ایران

[Downloaded from andisheyeamari.irstat.ir on 2024-05-08]

این بدان معناست که هر گروهی از دانشجویان ناچارند راهی برای رویارویی با این مشکلات بیابند. در این مقاله من راهبردهایی را که برای مطالعهٔ مقاله‌های روشهای آماری به کار می‌برم، به طور خلاصه بیان خواهم کرد. این پیشنهادها از طریق آزمون و خطا، بحث با همکاران و پیشنهادهای پولیا (۱۹۴۵) بسط یافته‌اند. این طرح برای دانشجویان نوشته شده و می‌توان از آن در هر کلاسی که دانشجویان آن ملزم به خواندن مقاله‌های آماری باشند، استفاده کرد. پس از آنکه برای نخستین بار این طرح را در کلاسی عرضه می‌کنم از دانشجویان می‌خواهم تا از آن در موقع خواندن دو مقاله که از مجله‌هایی مانند مجلهٔ آمار در پزشکی[1] یا مجلهٔ انجمن آماری آمریکایی[2] انتخاب می‌کنم، استفاده کنند. مقاله‌های انتخابی به موضوع مورد بحث در کلاس اختصاص دارند و سطح آنها کمی بالاتر از سطح دانش دانشجوی متوسط کلاس است. دانشجویان خلاصه‌ای از هر مقاله را تهیه می‌کنند و به پرسشهای مطرح شده در طرح پاسخ می‌دهند. من تاکنون هیچ ارزشیابی رسمی از این طرح انجام نداده‌ام، اما از بحثهای غیررسمی با دانشجویان چنین برمی‌آید که این طرح آنها را در مطالعهٔ مقاله‌ها یاری می‌کند.

۲ چگونه مقاله‌های روشهای آماری را مطالعه کنیم

۱.۲ وضع و محیط مناسب (مطالعهٔ مقاله زمان می‌برد؛ آسوده باشید.)

من همواره کسی را که خواندن مقاله‌های آماری را کاری ساده می‌داند تحسین می‌کنم. برای دیگران بهترین راه این است که کار را با برخورد مناسبی شروع کنند: «مطالعهٔ مقاله، مستلزم صرف وقت است.» من برای مطالعهٔ یک مقالهٔ ساده حداقل ۴ ساعت، و برای مطالعهٔ مقاله‌های پیچیده‌تر مدت زمان بسیار بیشتری را در نظر می‌گیرم. لزومی ندارد این مدت زمان یکجا به این کار اختصاص داده شود، اما تصوّری از مدت زمان کل که صرف این کار خواهد شد، می‌دهد. داشتن محیط مناسب برای کار مفید خواهد بود. یک صندلی راحت، نور مناسب، کاغذ و قلم و در صورت امکان رایانه از لوازم سودمندِ کارند.

۲.۲ توجه خود را به این مطلب متمرکز کنید که چرا مقاله را می‌خوانید.

این ضرب‌المثل که «به خاطر توجه به درختان از دیدن جنگل عاجزید» اغلب در موقع خواندن مقاله‌ای پیچیده مصداق پیدا می‌کند. قبل از اینکه مطاله را شروع کنید، باید مشخص کنید که چرا مقاله را می‌خوانید. توجه خود را به این نکتهٔ اصلی متمرکز کنید.

یک آماردان، سه دلیل اساسی برای خواندن یک مقاله دارد: علاقهٔ کلی، مرتبط بودن با یک کاربرد خاص، یا کسب دانش بیشتر دربارهٔ یک روش خاص آماری. پیشنهادهای گلسر (۱۹۸۶) برای وجود هدف چهارمی که همان داوری مقاله است، با توصیه‌های این مقاله سازگاری دارد. این دلایل از هم جدا نیستند. با این حال، مفید خواهد بود که در مطالعهٔ مقاله‌ای خاص، توجه خود را تنها به یکی از این دلایل معطوف کنیم. اگر مقاله، جنبهٔ عمومی داشته باشد، باید توجه خود را بر مقدمه و زمینهٔ آن متمرکز کنید. اگر در مقاله، اطلاعاتی دربارهٔ یک کاربرد داده شده باشد؛ باید توجه خود را بر بخش نتایج و داده‌ها متمرکز کنید. مطالعهٔ مقاله‌ای برای ارتقای روش‌شناسی آماری، مستلزم بیشترین دقت نظر در مقاله است.

صرف‌نظر از اینکه دلیل شما برای مطالعهٔ مقاله، چه باشد؛ سودمند خواهد بود که کار را با توجه به نخستین نکتهٔ هاف آغاز کنید «این مطلب را چه کسی و در کجا گفته است؟» آیا مؤلفان شهرت کافی دربارهٔ این روش‌شناسی دارند؟ آیا مقاله در مجله‌ای چاپ شده است که این روشها محتملاً مورد موشکافی دقیق هیأت تحریریه قرار می‌گیرند؟

۳.۲ مسأله را به زبان خودتان بیان کنید.

چکیده، مقدمه، و بحث مقاله را آن قدر بخوانید که بتوانید آن را در یک یا دو جمله بیان کنید. مسأله را در قالب نمادهایی که با آنها آشنایید، بیان کنید. یک (یا اگر به فکرتان می‌رسد، چند روش) راه حل ممکن را در قالب دانش فعلی خودتان، طرح کنید. شاید سودمند باشد که مقاله را به طور اجمالی مرور کنید، و در هر پاراگراف جملات اصلی را بخوانید تا مطمئن شوید که کلیهٔ ابعاد مسأله را می‌فهمید. اگر پس از انجام این کار باز هم نتوانستید مسأله را به وضوح در قالب جملات خودتان بیان کنید، به مراجع مقاله نگاه کنید. ببینید که آیا پیشتر تلاشی برای حل مسأله به عمل آمده است یا خیر؟ این مقالهٔ مقدم‌تر ممکن است مقاله را در قالب آشناتری بیان کرده باشد یا ممکن است به وسیلهٔ کسی نوشته شده باشد که می‌دانید محقق یا نویسندهٔ خوبی است. یک کتاب درسی عمومی که مسألهٔ مورد بحث در مقاله را مطرح کرده باشد، ممکن است مطالبی مرتبط با مسأله را برای حل مسأله در اختیار شما قرار دهد. ممکن است لازم باشد کار را چند بار تکرار کنید تا مسأله برای شما جا بیفتد.

زمانی که بتوانید مسأله را در قالب جملات خودتان بیان کنید، مقدمه را همراه با بخش روشها مجدداً بخوانید. توجه خود را به فرضهای انجام شده و محدودیتهایی معطوف کنید که این فرضها در راه حل ارائه شده به وجود آورده‌اند. روشهای ارائه شده در مقاله را با راه‌حل پیشنهادی که خود طرح کرده‌اید، مقایسه کنید. این دو راه‌حل تا چه حدّ متفاوت هستند؟ شما چه

1) Statistics in Medicine 2) JASA

[Downloaded from andisheyeamari.irstat.ir on 2024-05-08]

نکاتی را نادیده گرفته بودید که این روش آنها را در نظر گرفته است؟ چه فرضهایی در مقایسه با فرضهای انجام شده در این مقاله کرده بودید؟ اگر مقاله را از بُعد عام مطالعه می‌کنید، شاید تا همین جا برایتان کافی باشد. توصیه می‌کنم طرحی کلی از آنچه تاکنون انجام داده‌اید، برای مراجعات آتی تهیه کنید (به بخش ۸.۲ مراجعه کنید).

۴.۲ مسأله مشابهی پیدا کنید که با آن آشنایی دارید و جزئیات تکنیکی مسأله جدید را با مرتبط کردن آن با مسأله‌ای که با آن آشنایید، انجام دهید.

فرض کنید که می‌خواهید از نتایج این مقاله یا در یک مسألهٔ کاربردی استفاده کنید و یا اینکه می‌خواهید نظریه‌ای جدید را بفهمید و آن را بسط دهید. صرفاً خواندن یک مقاله درک کاملی از محتوای آن ایجاد نمی‌کند. استفاده از مقاله یا تدریس روشهای موجود در مقاله است که درک کاملتری را به وجود می‌آورد. اگر فرصتی برای تدریس مقاله در یک جمع یا کلاس درس داشته باشید، باید این کار را انجام دهید. حتی اگر نمی‌توانید آن را به دیگران بیاموزید، روشهای مقاله را مورد استفاده قرار دهید.

کار را با مرتبط کردن این مسأله با مسأله‌ای که از قبل با آن آشنایی دارید، شروع کنید. استدلالها و اعمال مسأله‌ای را که با آنها آشنا هستید دنبال کنید و آنها را به گونه‌ای بسط دهید که مسألهٔ جدید را شامل شود. برای مثال برای درک مقاله‌ای دربارهٔ برآورد پارامترها در مدلهای خطی با پارامترهای تصادفی، می‌توانید این مسأله را با مسأله‌ای دربارهٔ برآورد پارامترهای ثابت در مدلهای خطی ارتباط دهید، و تفاوتهای موجود در ساختارهای ماتریسی، تأثیر آنها بر جوابهای گاوس ـ مارکوف، تغییرپذیری برآوردها و غیره را امتحان کنید. شروع کار با مسأله‌ای آشنا، پایهٔ مستحکمی برای ورود به قلمروی ناشناخته در اختیارتان می‌گذارد. راه حلی برای مسأله جدیدتان، با پایهٔ دانش در حال رشد شما همخوانی خواهد داشت، و می‌توانید از آن در مسأله‌های دیگر استفاده کنید.

۵.۲ مسأله را در مورد داده‌ها به کار برید.

این نکته شبیه به نکتهٔ ۴.۲ است، اما تأکید بر آن دارد که از دانش جدید خود به طور ملموس استفاده کنید. با تکنیکهای جدید با استفاده از یک مجموعهٔ داده‌ها کار کنید. دربارهٔ این مسأله فکر کنید که اگر این داده‌ها دارای توزیعی با ساختاری متفاوت می‌بودند، چه اتفاقی می‌افتاد. اگر فرضها نقض شوند، چه اتفاقی می‌افتد؟ بسیاری از مقاله‌های کاربردی داده‌هایی را در مقاله می‌آورند که طرز استفاده از تکنیکهای مورد بحث را نشان می‌دهند.

مثالهای عددی ساده نیز ممکن است مفید باشند. سعی کنید که اعمال ماتریسی جدید را در مورد یک ماتریس ۲ × ۲ به کار ببرید و ببینید چه می‌شود. اگر مناسب باشد تکنیکها را به صورت برنامه‌ای درآورید و آماره‌هایی را که تولید می‌شوند، بررسی کنید.

۶.۲ نظریهٔ آماری را از جزئیات تکنیکی جدا کنید.

برای درک و استفادهٔ سهل از یک تکنیک جدید، لازم است هم نظریه و هم روش اجرای آن را بفهمید. با این حال، وقتی شروع به خواندن می‌کنید، فکر خوبی خواهد بود که نظریه را از اجرای آن جدا کنید. برآورد حاصل از نظریهٔ مدل آمیخته ممکن است مستلزم به کارگیری آلگوریتم EM برای محاسبات باشد. درهنگام مطالعهٔ مقاله؛ به وضوح بدانید که کدام قسمت مربوط به الگوریتم EM و کدام قسمت مربوط به نظریه می‌شود. در یک مقالهٔ بسیار پیچیده، می‌توانید یک بار نکات این مقاله را برای مباحث نظری و یک بار برای اجرای نظریه مرور کنید.

۷.۲ مقاله را حداقل سه بار بخوانید و تأکید شما هر بار بر بخشهای مختلف مقاله باشد.

تا به حال مقاله را یک بار خوانده‌اید، همه بخشها را با مقداری تفصیل بررسی کرده‌اید، و درکی کلی از مقاله تحصیل کرده‌اید. در بار دوم که مقاله را مطالعه می‌کنید، سازگاری درونی استدلالها را بررسی کنید، و توجه خود را به بخشهای مربوط به روشها و نتایج معطوف کنید. آیا فرضها لازم و کافی هستند؟ آیا منطق و نمادگذاری مقاله واضح و قابل فهم هستند؟ آیا قسمتهای مربوط به آمار، نظریه احتمال، و ریاضیات مقاله را درک می‌کنید؟ آیا می‌توانید این تکنیکها را به آماردانانی در سطح درک و تجربهٔ خود توضیح دهید و از آن دفاع کنید؟

ضمن اینکه این سؤالها را مطرح می‌کنید، به این موضوع هم فکر کنید که مؤلفان چه می‌توانستند بکنند تا کار را برای شما آسانتر کنند؟ هر چیزی که در اینجا به آن فکر می‌کنید، باید در مقاله‌های خودتان منظور کنید. اینک زمان مناسبی است که مجدداً مراجع مقاله را بررسی کنید، و احتمالاً باید مقاله‌های هم مضمونی را که دیگر پرسشهای شما را پاسخ می‌دهند، مورد مداقّه قرار دهید. می‌توانید دربارهٔ مقاله با همکارانتان صحبت کنید، مسائل حل نشده‌ای را در نظر بگیرید که منتظر ارائهٔ پاسخی از طرف شما هستند. شاید در این مرحله بخواهید برخی نظرات خود را برای گروهی مطرح کنید و عقاید آنها را دریافت کنید. باید از انجام این کار احساس اطمینان کنید زیرا اکنون بر بخشهایی از مسأله تسلط دارید و می‌توانید آنچه را که هنوز دربارهٔ آن احساس سردرگمی می‌کنید، توصیف کنید. بوئن (۱۹۸۲) پیشنهادهای خوبی دربارهٔ ارائهٔ مقالات در حضور جمع و پاسخ دادن به آنها دارد.

سرانجام مقاله را از نظر سازگاری برونی یا قابل تعمیم بودن آن مطالعه کنید. مقدمه، نتایج، مثالها و بخشهای بحث را به دقت بخوانید. ببینید چگونه می‌توان این تکنیک را به کار برد، این تکنیک برای چه نوع داده‌هایی مفیدند، کجا این تکنیک در محدودهٔ پاسخهایی برای مسائلی از این نوع می‌گنجد و اینکه آیا برنامه‌های کامپیوتری امتحان شده، پایدار، و قابل پشتیبانی موجود است یا خیر. دیگر کاربردهای این تکنیک مشابه را که در مراجع موجودند، مورد توجه قرار دهید.

[Downloaded from andisheyeamari.irstat.ir on 2024-05-08]

۸.۲ ایجاد یک پایگاه اطلاعاتی از یادداشتهای توضیحی را مورد توجه قرار دهید.

یک فهرست توضیحی از مراجع، مرور بر یک تکنیک را آسانتر می‌کند. شما در این مرحله مقاله را خلاصه نمی‌کنید، تنها نیاز به این دارید که آن قدر یادداشت بردارید که خواندن مجدد مقاله را آسانتر کند. این پایگاه اطلاعاتی می‌تواند در حد چند کارت یادداشت، ساده باشد، یا می‌تواند در حد یک برنامه کامپیوتری مراجع، پیچیده باشد. جدول ۱ نمونه‌ای از پایگاه اطلاعاتی است که من از آن استفاده می‌کنم، اما شما از خود ابتکار به خرج دهید. شما نمی‌خواهید با عدم استفاده از این شیوه برای مدتی معین، همه تلاشهایی که در خواندن مقاله برای اولین بار به عمل آورده‌اید، به هدر رود. همچنین نام نویسندگان، نظریه‌پردازان، و کارورزان را ضمن اینکه به آنها برخورد می‌کنید، یادداشت کنید. همه آماردانان در کلیهٔ زمینه‌ها خوب نیستند، اما می‌توانید در هر زمینه، بهترین را انتخاب کنید که از او تقلید کنید. همچنین می‌توانید یادداشتهایی دربارهٔ بهترین سخنرانان در همایشها و عوامل خوب بودن آنها را به مجموعه خود اضافه کنید.

۳ آخرین تذکرات

پیشرفت خود را بر اساس آنچه انجام داده‌اید بسنجید و نه بر اساس آنچه باید انجام دهید. تعداد مقاله‌ها به صورت نمایی رو به ازدیاد است، و شما هرگز قادر نخواهید بود که همهٔ آنها را بخوانید. با این حال، شاید بتوانید که همهٔ مقاله‌های خوب دربارهٔ یک موضوع خاص را مطالعه کنید. اگر به کمک یک پایگاه اطلاعاتی، سوابق مقاله‌ها را نگاهداری کنید، از تعداد مقاله‌هایی که می‌خوانید، متعجب خواهید شد.

جدول ۱ نمونه‌ای از یک مدخل برای یک پایگاه اطلاعاتی کتابشناختی.

مرجع

Andrews, D. F. (1971), "Sequentially Designed Experiments for Screening out Bad Models with F-tests", Biometrika, 58(3), 427.

نظریه‌های آماری به کار رفته: مدلهای خطی، طرحهای دنباله‌ای، آزمونهای F دنباله‌ای.

فنون محاسباتی به کار رفته: تولید انحرافهای نرمال تصادفی، ایجاد یک فضای طرح و انتخاب نقاط نمونه‌ای مبتنی بر تجمع داده‌ها، شبیه‌سازی داده‌ها.

فرضهای توزیعی: توزیع نرمال، نیازمند اندازه‌های تکراری است.

دیگر فرضهای مربوط: مدلهای مورد بحث در این مقاله، در سلسله مراتبی از مدلهای چندجمله‌ای موجودند، و می‌خواهیم بهترین مرتبه را برای چندجمله‌ای انتخاب کنیم.

مجموعه داده به کار رفته: داده‌های شبیه‌سازی شده.

مراجع باز ارجاعی مربوط: هیچ مورد از این مراجع که به طرح فضایی برای آزمایش بپردازد، موجود نیست.

نکات: امکان استفاده در امتحانهای تک مرحله‌ای یا انتخاب مدلهایی برای زوال نرخها در اندازه‌های تکراری. پیشنهاد اساسی این است که فضاهای طرح را انتخاب کنیم که اجازه دهند تعیین کنیم که کدام یک از اعضای مجموعه‌ای از مدلهای ممکن، نامعتبر است، و سپس آزمایش خود را بر معتبرترین مدل اجرا کنیم.

مراجع

Boen. J. R., and Zahn, D. A. (1982). *The Human Side of Statistical Consulting.* Belmont. CA: Wadsworth.

Colton. T. (1979). Statistics in Medicine, Boston: Little, Brown.

Cowen. A. C. (1991). "Teaching and Testing Mathematical Reading. *American Mathematical Monthly*, 98, 50-53.

Gleser. I., J. (1986). "Some Notes on Refereeing." *The American Statistician*, 40, 310-315.

Huff. D. (1954), *How to Lie with Statistics*, New York: W. W.

Norton. Parke, N. (1958), *Guide to the Literature of Mathematics and Physics* (2n. ed.). New York: Dover.

[Downloaded from andisheyeamari.irstat.ir on 2024-05-08]

Pemberton, J. (1969). *How to Find Out in Mathematics: A Guide to Sources of Information*, Oxford: Pergamon Press.

Phanstiel. O. (1990), "How to Read Chemistry," *Journal of Chemical Education*. 67, 57-59.

Polya. G. (1945), *How to Solve It: A new Aspect of Mathematical Methods*. Princetion: Princetion University Press.

Sackett. D. L., Haynes, R., B., Guyatt, G. H., and Tugwell. P. (1991) *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Boston Little, Brown.

Schechtman. K. B., and Spitznagel. E. L., (1987), "Teaching Biostatistics with an Emphasis on Reading the Medical Literature,"in *ASA Proceedings of Statistical Education*, pp. 111-115.

اصل این مقاله با عنوان
How to Read the Statistical Methods Litrature: A Guide for Students
در The American Statistician, May 1997, Vol. 51, No. 2. به چاپ رسیده است.

[Downloaded from andisheyeamari.irstat.ir on 2024-05-08]