MSSH
Methodology of Social Sciences and Humanities

Vol.29, No. 114, Spring 2023
Quarterly
ISSN: 1608-7070
Method@rihu.ac.ir

The Role of Sample Size on Interpretation of the Result in Applied Research

A Study on the Analysis of Regression Models

Alireza Pakgohar (Assistant Professor, Payam Noor University, A_pakgohar@pnu.ac.ir)

ARTICLE INFO

Article History

Received: 2022/08/04

Accepted: 2023/01/05

Key Words:

Hypothesis Test,
Effect Size,
Sample Size,
Significant Level,
Confidence Interval,
Big Data,
P-Value

ABSTRACT

Many applied types of research have been carried out using statistical hypothesis tests and many researchers reject or accept the research hypothesis only by looking at the p-value. On the other hand, due to the relationship between the p-value and the sample size, in large samples, despite the small size of the work, the hypothesis is accepted with very high confidence, and the researcher relies solely on the p-value to support practical insignificance results. Thus, many research studies can be classified as research that has many tests of meaningful assumption but lacks practicality and scientific significance. In this paper, first, considering the issue of big data, the issue of data volume and data diversity in big data from a statistical perspective, then the p-value, effect size, and confidence interval as three decision criteria in hypothetical tests on different samples in the range (173-19361) The impact of big data on these three indicators has been examined and in particular. The results showed: A large sample is not an advantage to increasing reliability in hypothetical tests. It can make meaningful claims that are not of practical importance and in small samples fall into the category of random effects and sampling error. Also, the effect size is not affected by the sample size and converges with increasing constant sample size. Finally, the data showed that the confidence interval is visually better than other indicators.

فصلنامه علمی روش‌شناسی علوم انسانی
بهار ۱۴۰۲، سال ۲۹، شماره ۱۱۴ (صفحات ۱۹-۳۴)
ISSN: 1608-7070
مقاله پژوهشی
Method@rihu.ac.ir

نقش اندازه نمونه در تفسیر نتیجه پژوهش‌های کاربردی: مطالعه‌ای بر مدل‌های رگرسیونی

علیرضا پاک‌گوهر (استادیار، دانشگاه پیام نور؛ A_pakgohar@pnu.ac.ir)

اطلاعات مقاله

تاریخ دریافت: ۱۴۰۱/۰۵/۱۳

تاریخ پذیرش: ۱۴۰۱/۱۰/۱۵

واژگان کلیدی:
آزمون فرض آماری،
اندازه اثر،
اندازه نمونه،
سطح معناداری،
فاصله اطمینان،
کلان‌داده،
مقدار-احتمال

چکیده

امروزه بیشتر پژوهشگران تحقیقات کاربردی صرفاً با نگاه به مقدار-احتمال، فرضیه پژوهش را رد یا قبول می‌کنند؛ همچنین با توجه به رابطه‌ای که بین مقدار-احتمال و اندازه نمونه وجود دارد، معمولاً در نمونه‌های بزرگ، با وجود کوچک بودن اندازه اثر با یک اطمینان بسیار بالا فرضیه ادعا پذیرفته می‌شود و محقق با تکیه صرف به مقدار-احتمال به‌سمت حمایت از نتایج بی‌اهمیت عملی سوق داده می‌شود؛ بدین ترتیب بسیاری از مطالعات پژوهش می‌توانند در زمره تحقیقاتی قرار گیرند که دارای تعدادی آزمون فرض معنادار شده، اما فاقد وجاهت کاربردی و اهمیت علمی است.

نخست این مقاله با توجه به موضوع کلان‌داده، مسئله حجم داده و تنوع داده‌ها در کلان‌داده از نظر آمار پرداخته، سپس مقدار-احتمال، اندازه اثر و فاصله اطمینان به‌عنوان سه معیار تصمیم در آزمون‌های فرض روی نمونه‌های مختلف در بازه (۱۹۳۶۱-۱۷۳) بررسی و به‌طور خاص به تأثیر کلان‌داده روی این سه شاخص توجه کرده است.

نتایج نشان داد کلان‌داده به‌عنوان یک نمونه بزرگ، نه‌تنها مزیتی برای افزایش اطمینان در آزمون‌های فرض ندارند، بلکه می‌توانند موجب معنادار شدن ادعاهایی شوند که از نظر عملی اهمیت چندانی ندارند و در نمونه‌های نه‌چندان بزرگ در زمره اثرهای تصادفی و خطای نمونه‌گیری قرار می‌گیرند؛ همچنین اندازه اثر تحت‌تأثیر اندازه نمونه قرار نگرفته، با افزایش اندازه نمونه به‌سمت یک مقدار ثابت همگرایی دارد. درنهایت داده‌ها نشان دادند که فاصله اطمینان از نظر بصری بهتر از شاخص‌های دیگر عمل می‌کند.

۱. مقدمه

امروزه داده‌کاوی در اثر پیشرفت‌های فناوری توانایی جمع‌آوری، انتقال و ذخیره مجموعه‌های بزرگ داده یکی از علوم مطرح در تحلیل داده به شمار می‌رود و مطالعات آماری بسیاری روی کلان‌داده‌ها انجام شده است؛ برای مثال اوربای و جاپ[1] (۲۰۰۹) در یک مطالعه روی بیش از ۱۰۸ هزار وسیله نقلیه استفاده شده در بازار عمده فروشی خودرو استفاده کرده‌اند. گاس و یائو[2] (۲۰۱۱) از ۳/۷ میلیون رکورد شامل معاملات برای خدمات تأمین فدرال دولت فدرال ایالات متحده استفاده کرده‌اند. پاک گوهر و همکاران با مطالعه روی تقریباً ۲۷۷ هزار پرونده تصادفات ترافیکی ایران شامل حدود سه میلیون داده با نمونه‌های بزرگ سروکار داشته و از تکنیک‌های آماری و داده‌کاوی در تحلیل شدت تصادفات بهره برده‌اند؛ اما با توجه به رابطه معکوسی که بین اندازه نمونه و مقدار-احتمال وجود دارد، در مطالعات روی کلان‌داده‌ها با توان بسیار بالایی فرضیه معنادار بودن یک اثر جزئی پذیرفته شده، و نتایجی که اهمیت چندانی ندارند از نظر آماری مهم و معنادار دانسته می‌شوند. این دوگانگی در اهمیت موضوع و اهمیت آماری از سوی محققان مختلفی مورد توجه قرار گرفته است، مثلاً چتفیلد[3] (۱۹۹۵) می‌گوید: مسئله این نیست که آیا تفاوت‌ها (منظور اختلاف بین آماره و مقدار پارامتر تحت فرض صفر است) معنادار هستند (چون تقریباً همیشه در نمونه‌های بزرگ هستند)، اما آیا این تفاوت قابل توجه است؟ سطح معناداری را فراموش کنید، اهمیت عملی نتیجه چیست؟

افزایش توان آزمون در نمونه‌های بزرگ بدان معناست که محققان می‌توانند اثرهای جزئی را با تکیه به اعتبار نمونه بزرگی که در اختیار دارند تشخیص دهند، اما تفسیر نتایج با تکیه صرف به مقدار-احتمال فقط می‌تواند به پشتیبانی از فرض ادعایی بینجامد که اهمیت عملی کمی در مطالعات کاربردی دارد.

در این مقاله به رابطه اندازه نمونه و مقدار-احتمال متمرکز و به اندازه اثر و فاصله‌های اطمینان توجه شده است. نخست در بخش ۲ به تعریف و ماهیت کلان‌داده و چالش‌هایی پرداخته است که با روش‌های آماری کلاسیک دارد. در بخش ۳ درباره اینکه چگونه یک مقدار-احتمال به سرعت با افزایش اندازه نمونه به صفر می‌رسد؛ بحث و بررسی شده، سپس به مسئله مقدار-احتمال در مدل رگرسیونی پرداخته است. آنگاه اندازه اثر، فاصله اطمینان و مزایای ضمنی آنها را بررسی می‌شود. در بخش ۴ تأثیر اندازه نمونه بر مقدار-احتمال در آزمون فرض ضریب رگرسیونی را بررسی و با توجه به اهمیت بررسی اندازه اثر، پس از بررسی اهمیت آن، روش‌های محاسبه اندازه اثر در مدل‌های گوناگون رگرسیونی معرفی شده است. بخش ۵ رویکردهای مختلف مقدار-احتمال، اندازه اثر و فاصله اطمینان برای مقایسه آماره و پارامتر ضریب رگرسیونی استانداردشده ارائه می‌دهد. در این رویکرد تلاش می‌شود با استفاده از نمودارها تصویر روشن‌تری برای درک بیشتر خواننده از میزان تأثیر اندازه نمونه بر شاخص‌های مورد اشاره ارائه گردد و سرانجام در بخش ۶ با استفاده از روش‌های مختلف نمونه‌گیری از یک کلان‌داده واقعی مربوط به اطلاعات تصادفات ترافیکی در ایران فرایند تأثیرپذیری مقدار-احتمال و اندازه اثر به‌ازای اندازه‌های مختلف نمونه بررسی می‌شود.

۲. مفهوم کلان‌داده

امروزه نقش و اهمیت داده‌ها به‌عنوان یک منبع ارزشمند برای بهره‌برداری، تحلیل و استخراج اطلاعات بر کسی پوشیده نیست؛ همچنین توسعه روزافزون مجموعه داده‌ها سبب پیچیدگی داده‌ها می‌شود و پایگاه‌های داده‌های سنتی قادر به پوشش و پشتیبانی آنها نیستند و بدیهی است که نتوانیم از ابزارهای سنتی همچون صفحه گسترده برای تجزیه و تحلیل آنها استفاده کنیم (عارفی اصل، ۱۳۹۷)؛ ازاین‌رو کلان‌داده[4] به مجموعه‌ای از داده‌ها گویند که دارای حجم زیاد، ساختار گوناگون و پیچیده می‌باشند. به‌طوری‌که این حجم گسترده داده‌ها موجب شده است که ذخیره‌سازی، آنالیز، مجازی‌سازی، استفاده مؤثر و عملی از داده‌ها و... جهت پردازش با مشکلات زیادی روبه‌رو شوند (Zikopoulos et al., 2012; Kaisler, 2013). بنابراین

1. Overby & Jap
2. Ghose & Yao
3. Chatfield
4. Big Data

کلان‌داده محصولی از یک محیط تکنولوژیکی است که در آن تقریباً هر چیزی به‌طور دیجیتالی دریافت، اندازه‌گیری و ثبت شده، درنهایت به داده تبدیل می‌شود (عارفی‌اصل، ۱۳۹۷).

۲-۱. کلان‌داده از نظر آمار

به‌طورکلی محدوده کلان‌داده‌ها شامل سه بخش اصلی نوع، سرعت و حجم می‌شود (Chen et al., 2014; Kaisler, 2013).

حجم داده: معمولاً در کلان‌داده‌ها اندازه داده‌های تولید شده در مبناهایی بزرگ‌تر از ترابایت یا حتی اگزابایت است، اما در آمار با توجه به تأثیر اندازه نمونه بر شاخص‌های بسیاری مانند مقدار-احتمال نیازی به حجم بسیار بزرگ نیست؛ بلکه تعداد داده بیشتر از ۱۰ هزار رکورد نیز می‌تواند به‌عنوان یک کلان‌داده تعریف شود. این مسئله مشابه مقیاس اندازه‌گیری است که از نظر ریاضی تفاوت‌هایی برای مقیاس فاصله‌ای و مقیاس نسبتی در نظر گرفته می‌شود، اما در آمار این دو مقیاس تفاوتی با هم ندارند. پس گرچه حجم بسیار بالا به‌عنوان یک ویژگی اصلی برای کلان‌داده‌ها به شمار می‌رود، ولی در آمار یک انبار داده با حجم تقریباً بزرگ نیز به‌عنوان یک کلان‌داده به شمار می‌رود. شاید بتوان این موضوع را یک ویژگی مناسب و مفید برای رویکرد آمار در نظر گرفت، اما باید گفت حجم کلان‌داده در مقیاس ترابایت تا زتابایت می‌تواند فلسفه بسیاری از آزمون‌های آماری را بی‌معنا کند؛ به‌گونه‌ای‌که در این مقاله نیز به آن اشاره می‌شود یک اندازه نمونه به اندازه کافی بزرگ به محاسبه مقدار-احتمال همگرا به صفر می‌انجامد. بدین سبب از دیدگاه بیشتر صاحب‌نظران، روش‌های آماری برای تحلیل کلان‌داده‌ها مناسب نیستند و روش‌های داده‌کاوی به‌عنوان روش‌هایی هستند که بیشتر پیش‌فرض‌های آزمون‌ها و مدل‌های پارامتری را دربرنمی‌گیرد و برخی از تکنیک‌های آن با حجم بزرگ داده‌ها سازگار است؛ مانند تکنیک‌های درخت تصمیم، شبکه عصبی، انواع خوشه‌بندی و... (Kaisler, 2013).

تنوع داده: کلان‌داده‌ها تنها تولیدکننده داده‌های عددی نیستند بلکه بسیاری از انواع داده‌های غیرعددی همانند متن، تصویر، ویدئو، صوت و... را تولید می‌کنند که به‌عنوان بزرگ‌ترین مانع برای استفاده از آزمون‌های مرسوم و کلاسیک آماری به شمار می‌روند. به‌طورکلی انواع مختلف محتوای کلان‌داده را می‌توان در سه دسته داده‌های ساختاریافته، نیمه‌ساختاریافته و غیرساختاریافته تقسیم کرد (سنمی، ۱۳۹۵).

داده‌های ساخت یافته، داده‌هایی را توصیف می‌کند که درون یک طرح رابطه‌ای گروه‌بندی شده باشند؛ مانند سطرها و ستون‌ها درون یک پایگاه داده استاندارد. پیکربندی و سازگاری داده‌ها این اجازه را می‌دهد تا به پرس‌وجوهای ساده برای رسیدن به اطلاعات مفید براساس پارامترهای سازمان‌دهی شده، و به نیازهای عملیاتی پاسخ دهد. داده‌های نیمه‌ساختاریافته از یک طرح صریح و ثابت پیروی نمی‌کند. این داده‌ها به‌طور ذاتی خود توصیف هستند و شامل تگ یا نشانه‌های دیگر برای به اجرا درآوردن سلسله مراتبی از سوابق و زمینه‌های درون داده می‌باشند، مانند وبلاگ‌ها و رسانه‌های اجتماعی. داده‌های غیرساختاریافته شامل فرمت‌هایی است که نمی‌تواند به راحتی در جداول رابطه‌ای برای تجزیه و تحلیل و یا پرس‌وجو نشان داده شود؛ همچون تصاویر، فایل‌های صوتی و ویدیوئی (همان).

بدیهی است آزمون‌های آماری صرفاً روی داده‌های ساختاریافته انجام می‌شوند و برای داده‌های نیمه‌ساختاریافته روش‌های تحلیل کیفی قابل انجام است که با روش‌های مرسوم آماری و شیوه تحلیل بسیار متفاوت هستند.

۳. معیارهای تصمیم در آزمون‌های فرض

در این بخش به سه معیار تصمیم‌گیری در آزمون فرض برای زمانی‌که با کلان‌داده‌ها سروکار داریم، پرداخته می‌شود که عبارتند از: مقدار-احتمال، اندازه اثر و فاصله اطمینان. رویکرد بحث مرور دیدگاه صاحب‌نظران آمار در زمینهٔ ضعف مقدار-احتمال در تأثیرپذیری از اندازه نمونه است.

معمولاً در آزمون‌های فرض آماری با تکیه بر فرض صفر و تحت این فرض برای پارامتر موردنظر خود یک مقدار مشخص، مثلاً صفر در نظر گرفته می‌شود که بیانگر اندازه ناچیز و تصادفی است؛ برای مثال در آزمون فرض ضریب رگرسیون خطی ساده معمولاً فرض صفر مقدار پارامتر شیب خط را صفر در نظر می‌گیرد که بیانگر «بدون تأثیر» بودن آن است.

همان‌طورکه فرض صفر، علت اختلاف آماره مشاهده شده و

پارامتر را به‌عنوان عامل‌های خطای تصادفی در نظر می‌گیرد، فرض ادعا این اختلاف را معنادار و قابل ملاحظه می‌داند و ازآنجاکه مطالعات کلان‌داده‌ها خطای استاندارد برآوردگر، مقداری بسیار جزئی می‌شود و همین امر به تخمین بسیار دقیق و به‌طور معادل پذیرش فرض ادعا خواهد انجامید.

مثال: برای آزمون میانگین با فرض $H_0: \mu = 0$ در مقابل $H_1: \mu \neq 0$ به ازای $s = 0.00102$ و $\bar{X} = 0.00002$ از یک نمونه به‌اندازه n=10000 فرض ادعا با اطمینان ۹۵ درصد پذیرفته می‌شود.

پرسش اینجاست که آیا این اختلاف بین میانگین نمونه و پارامتر جامعه اهمیت علمی برای محققان دارد؟ برای درک بهتر باید یک بار دیگر به‌مقادیر میانگین نمونه و پارامتر میانگین و اختلاف این دو توجه کرد. در کدام مطالعه اختلاف ۰/۰۰۰۰۲ بین آماره و پارامتر مورد توجه محقق بوده است و آیا از نظر آماری مقدار قابل توجهی می‌باشد؟

مقدار-احتمال‌ها با استفاده از برآورد پارامتر مورد نظر، فاصله بین داده‌های مشاهده شده و مقدار پارامتر تحت فرض صفر را اندازه‌گیری می‌کنند. معمولاً فاصله در واحدهای انحراف استاندارد از این برآورد اندازه‌گیری می‌شود؛ برای مثال آزمون برای ضریب رگرسیون برابر یا یک، براساس فاصله یک واحدی خطای استاندارد نسبت به صفر است (بازرگان‌لاری، ۱۳۸۴).

ازآنجاکه خطای استاندارد برآوردگر با افزایش اندازه نمونه کوچک می‌شود، در یک کلان‌داده با خطای استاندارد بسیار ناچیز روبه‌رو می‌شویم؛ به‌گونه‌ای‌که حتی اختلاف بسیار کوچک بین برآوردگر و مقدار پارامتر از نظر آماری معنادار خواهد شد. کوهن[1] (۱۹۹۲) می‌گوید: کمی اندیشه واقعیتی قابل درک را در بین آمارشناسان نشان می‌دهد: فرض صفر همیشه در دنیای واقعی نادرست است، زیرا هر اختلافی بین آماره و پارامتر باید به این صورت باشد که به‌ازای یک نمونه به‌اندازه کافی بزرگ نتیجه معناداری به‌دست آورد و به رد آن بینجامد؛ بنابراین اگر فرض صفر همیشه نادرست است، چه دلیل مهمی در ردّ آن وجود دارد؟

مقدار-احتمال

موضوع مقدار-احتمال به تولید مقالات فراوانی انجامیده است (همچون: برومیده و شاهقیان، ۱۳۸۳؛ Tanha, 2017). در برخی از مقالات کوشیده‌اند که از معیار مقدار-احتمال اصلاح شده استفاده شود (همچون: اسماعیلی و همکاران، ۱۳۹۰؛ Tanha, 2017).

مقدار-احتمال که معمولاً به‌عنوان خروجی برنامه‌های آماری در نظر گرفته می‌شود؛ در واقع یک متغیر تصادفی است که از توزیع آماره آزمون سرچشمه گرفته می‌شود. توزیع مقدار-احتمال تحت فرض مقابل، تابعی از اندازه نمونه و مقدار درست پارامتر آزمون است (شریفیان و خزاعی، ۱۳۹۱)؛ بنابراین با توجه به تأثیری که اندازه نمونه در تعیین مقدار-احتمال دارد، نمی‌توان آزمون فرض را مورد خدشه قرار داد؛ ضمن آنکه همیشه با یک نمونه به‌اندازه کافی بزرگ سروکار نداریم. پس شواهد پیشین در مورد فرض صفر قصد دارد تا تمرکز ما را از تکیه صرف به معناداری آزمون (تکیه به مقدار-احتمال) به‌اندازه اثر سوق دهد. به عبارت دیگر به‌ازای نمونه‌های بسیار بزرگ یک تصمیم آماری باید فراتر از ردّ فرض صفر براساس مقدار-احتمال باشد. درعوض محققان باید محتاطانه ارزیابی کنند که آیا یک مقدار-احتمال کوچک به‌صورت مصنوعی و به‌ازای اندازه بزرگ یک نمونه به‌دست آمده است یا خیر و باید میزان و حساسیت اثر را نیز به‌دقت اندازه‌گیری کنند؛ به‌عبارتی، نتیجه‌گیری براساس سطح معناداری با ادعای عدم پذیرش فرض صفر بی‌معناست، مگر اینکه با توجه به‌بزرگی واقعی اندازه اثر تفسیر شود.

به‌طورکلی مطالعات آماری در زمینه مقدار-احتمال، صرف‌نظر از بحث اندازه نمونه به دو جریان تقسیم می‌شود. یک جریان، استفاده از آزمون فرض را به‌طورکلی رد می‌کند و از حرکت به سمت تخمین پارامتر دفاع می‌کند، به‌طوری‌که در گزارش‌های خود به برآوردهای نقطه‌ای و بازه‌های اطمینان بسنده می‌کنند (Hubbard & Armstrong, 2006). جریان دوم مخالف آزمون فرض نیست، بلکه بر لزوم تمرکز بر اهمیت عملی و نه نگاه صرف به‌سطح معناداری آماری تأکید دارد.

از دیدگاه دیزدیر (Disdier et al, 2008) و همکاران در مطالعات تجربی بهتر است از فاصله‌های اطمینان استفاده شود.

1. Cohen

بازه اطمینان می‌تواند برای شاخص‌های مختلف به‌دست آید؛ مانند بازه اطمینان آماره، بازه اطمینان برای اندازه اثر و...؛ درحالی‌که مقدار-احتمال تنها احتمال اینکه فرض صفر با توجه به یک اثر واقعی رد شود، توصیف می‌کند، بازه اطمینان دامنه‌ای برای نشان دادن بزرگی واقعی پارامتر موردنظر ارائه می‌دهد. با افزایش اندازه نمونه، بازه اطمینان معمولی باریک‌تر می‌شود؛ بنابراین اطلاعات موجود در بازه‌های اطمینان سبب می‌شود تخمین دقیق‌تری به دست آید. این خاصیت موجب می‌گردد که حتی اگر محقق مطمئن نیست که نمونه برای استفاده از مقدار-احتمال‌ها بسیار بزرگ است یا خیر، تکیه بر بازه اطمینان همیشه ایمن باشد. حتی برخی از محققان پیشنهادی متفاوت در مورد مسئله کاهش مقدار-احتمال دارند. آنها پیشنهاد می‌کنند به‌ازای افزایش نمونه، آستانه مقدار-احتمال نیز کمتر شود. به‌عبارتی حداکثر مقدار خطای نوع اول (α) کوچک‌تر شود (Greene, 2003). براساس این استدلال به‌جای ادعای معناداری (با مقایسه خطای نوع اول)، مقدار-احتمال برای نمونه‌های بسیار بزرگ با مقادیر یک‌درصد، یک‌دهم درصد یا حتی کوچک‌تر مقایسه شود؛ با وجودی که این رویکرد نیز نمی‌تواند قابل دفاع باشد، زیرا اندازه مقدار-احتمال به ما میزان اطمینان از پذیرش فرض مقابل را می‌دهد و از سوی دیگر هیچ قاعده مشخصی برای تعیین آستانه مقدار-احتمال تعریف نشده‌است؛ بنابراین هیچ قرارداد علمی مشخصی نمی‌تواند به ما بگوید فرضیه ادعای یک آزمون فرض را می‌توان با اطمینان ۹۵درصد پذیرفت، اما برای یک مطالعه مشابه با نمونه بیشتر اجازه پذیرش فرضیه ادعا را نداریم.

اندازه اثر

امروزه اندازه اثر به‌اندازه مقدار-احتمال مورد توجه شیوه‌نامه‌های نگارش مقالات تحقیقات کاربردی قرار گرفته است؛ به‌گونه‌ای که در بعضی از آنها به پژوهشگران توصیه می‌شود که در کنار مقدار-احتمال، اندازه اثر نیز گزارش گردد (Wasserstein, 2016)

تا تفاوت واقعی بین گروه‌ها یا قدرت رابطه بین دو متغیر را در مقیاس عددی اندازه گیری کند (Sullivan &Feinn, 2012). کلاین(Kline, 2004) در کتابش اهمیت اندازه اثر را در بزرگی اندازه آن می‌داند و معتقد است این بزرگی مقدار اندازه اثر می‌باشد که اهمیت علمی یافته پژوهش را نشان می‌دهد و نمی‌توان مانند آزمون فرض صرفاً به این اندازه بسنده کنیم که میزان اطمینان ما از معناداری فرضیه ادعا برابر $(1-\alpha)\%$ است. او می‌گوید: «شما باید نتایج را براساس بزرگی اندازه اثر توصیف کنید؛ نه فقط به اینکه آیا یک درمان بر افراد تأثیر می‌گذارد یا خیر، بلکه چقدر آنها را تحت‌تأثیر قرار می‌دهد. کوهن (۱۹۹۲) حتی اندازه اثر را معیار آزمون فرض دانسته است: «نتیجه اولیه یک مطالعه را اندازه اثر می‌داند، نه مقدار-احتمال». گزارش مقدار-احتمال بدون ذکر اندازه اثر موجب می‌شود در مورد نقش متغیر مستقل بر متغیر وابسته تصویر درستی ارائه نشود (Sugathan & Jacob, 2021).

فاصله اطمینان

به محققانی که با کلان‌داده‌ها کار می‌کنند، توصیه می‌شود که از بازه‌های اطمینان استفاده کنند و اندازه‌های اثر را گزارش دهند (Cannon et al., 2006; Disdier et al., 2008 ; Goolsbee & Guryan, 2006). بعضی ازمزایای عمدهٔ گزارش فاصله‌های اطمینان نسبت به مقدار-احتمال عبارتند از: الف) با گزارش فواصل اطمینان در کلان‌داده‌ها مسئله تمایل محققان را به ارائه مقدار-احتمال کوچک جهت توجیه توان آزمون برطرف می‌کند؛ مثلاً فرض کنید محقق با ارائه فاصله اطمینان برآوردگر پارامتر، دیگر نمی‌تواند معناداری آزمون فرض و پشتیبانی از فرض ادعا را توجیه نماید؛ زیرا در یک کلان داده، فاصله اطمینان دربرگیرنده یک بازه بسیار کوچک حول برآوردگر پارامتر است و طبعاً اگر اندازه اثر ناچیز باشد، تقارب و نزدیکی برآوردگر و پارامتر سبب کاهش تأثیر کاذبی می‌شود که کوچکی عدد مقدار-احتمال به‌خواننده القا می‌نماید. ب) پژوهشگران با گزارش فاصله‌های اطمینان زمینهٔ انجام مطالعات مروری مانند فراتحلیل، ترکیب مطالعات پیشین و کمک به پیشرفت دانش علمی در زمینه ارائه اطلاعات آماری را فراهم می‌کنند. این امر به‌ویژه زمانی صادق است که بازه اطمینان از قابلیت ارتجاعی یا کشش[1] متغیر Y

1. Elasticity

(درصد تغییر در Y به‌ازای هریک درصد تغییر در X) برخوردار باشد.

با توجه به آنکه یک نمونه بزرگ به یک فاصله اطمینان[1] (CI) با بازه کوچک می‌انجامد، آستانه‌های فاصله‌های اطمینان مخصوصاً برای تعیین اندازه پارامتر (ناشناخته) و دامنه آن بسیار مفید خواهند بود؛ به‌همین سبب می‌توان از نمودارهای فاصله‌های اطمینان برای کلان‌داده‌ها استفاده زیادی کرد. گولسبی[2] و همکاران (۲۰۰۶) بیان می‌دارند با نمودارهایی که از فاصله اطمینان ترسیم می‌شوند، می‌توان از این استدلال که «نتایج از نظر کیفی مشابه هستند» فراتر رفته و دامنه تخمین‌ها را از نظر کمّی با یکدیگر مقایسه کرد (Goolsbee et al., 2006). آخرین نکته آنکه یک محقق مطالعات تجربی برای بررسی مدل خود تمایل دارد مدلش را با مدل‌های دیگر مقایسه نماید؛ بدین منظور می‌توان از شاخص‌هایی مانند شاخص حساسیت، مشخصه‌سازی و... استفاده کرد که در اینجا از بحث درباره این شاخص‌ها اجتناب می‌کنیم (ر.ک: پاک‌گوهر، ۱۳۹۵).

۴. رگرسیون خطی و کلان‌داده‌ها

معمولاً در آزمون‌های فرض آماری، یک فرضیه پژوهش مطرح شده و درباره پذیرفته شدن یا ردّ این فرضیه تصمیم‌گیری می‌شود و در مقابل، فرضیه صفر به‌عنوان یک پیش فرض درست در نقطه مقابل فرض پژوهش قرار می‌گیرد. نکته اصلی آنکه اساس آزمون آماری همین فرض صفر است که مفهوم «بی‌تأثیر» بودنی را دربردارد که در مطالعات آزمایشی مفهوم بی‌تأثیری دربرگیرندهٔ آن است که هیچ تفاوتی بین گروه‌ها وجود ندارد و درمطالعات توصیفی به‌معنای آن می‌باشد که هیچ ارتباطی بین متغیرها وجود ندارد.

در یک مدل رگرسیون خطی نیز عمدتاً دو فرضیه اصلی بررسی می‌شود؛ تأثیر داشتن مقدار ثابت یا عرض از مبدأ و مقدار شیب خط رگرسیون که معمولاً آن را با عبارت شیب خط (در مدل رگرسیون خطی ساده) و ضریب رگرسیونی تلخیص داده، و معمولاً در آزمون فرض مدل رگرسیون خطی، فرضیه صفر برای ضریب رگرسیونی مقدار عددی صفر پیشنهاد می‌شود و در مقابل، فرضیه ادعا برخلاف آن یک مقدار مثبت یا منفی برای ضریب رگرسیونی در نظر می‌گیرد. بدیهی است اندازه نمونه مورد مطالعه تأثیر مستقیمی بر آماره محاسبه شده دارد و نیز با افزایش آن با انحرافاتی بسیار جزئی فرضیه ادعا پذیرفته می‌شود.

قضیه (۱) نشان می‌دهد که توزیع حدی برآوردگر $\hat{\beta}$ تمام جرم خود را در پارامتر جمعیت دارد.

قضیه ۱: در مدل رگرسیون خطی، مقدار احتمال آزمون فرض $H_0: \beta = 0$ در مقابل $H_1: \beta \neq 0$ به‌ازای افزایش نمونه به‌اندازه کافی بزرگ به سمت صفر میل می‌کند.

اثبات: می‌دانیم مقدار-احتمال به‌صورت زیر محاسبه می‌شود:

$$P-Value = 2(1-\phi(df, |T|)), \qquad (۱)$$

که در آن ϕ توزیع تجمعی توزیع تی استودنت با درجه آزادی باقیمانده (df)و منظور از $|T|$ مقدار قدرمطلق آماره آزمون T مشاهده شده است که از رابطه زیر به‌دست می‌آید:

$$|T| = \frac{\hat{\beta}-0}{\sigma_{\hat{\beta}}} \qquad (۲)$$

بدیهی است آماره T یک تابع صعودی نسبت به اندازه نمونه (n) است، زیرا در مدل رگرسیون خطی ساده خطای استاندارد برابر است با:

$$\sigma_{\hat{\beta}} = \sqrt{\frac{MSE}{S_x^2(n-1)}} \qquad (۳)$$

و به ازای یک نمونه بسیار بزرگ، مقدار آماره به سمت بی‌نهایت میل می‌کند و به همین دلیل مقدار-احتمال به سمت صفر نزدیک می‌شود؛ به عبارت دیگر در رابطه (۱) خواهیم داشت:

1. Confidence Interval
2. Goolsbee

$$\lim_{n\to\infty} P-Value = \lim_{n\to\infty} 2(1-\phi(df,|T|)) = 2(1-\lim_{n\to\infty}\phi(df,|T|)) = 2(1-\phi(df,\lim_{n\to\infty}\frac{\sqrt{n-1}|\hat{\beta}|S_X}{\sqrt{MSE}})),$$

بنابراین:

$$\lim_{n\to\infty} P-Value = \lim_{n\to\infty} P(|\hat{\beta}-\beta|<\varepsilon) = \begin{cases} 0, & \beta\neq 0 \\ 1 & \beta = 0 \end{cases}$$

زیرا:

$$\begin{cases} 2(1-\phi(df,0)) = 1, & \beta = 0 \\ 2(1-\phi(df,\infty)) = 0, & \beta \neq 0 \end{cases}$$

در یک مطالعه رگرسیون فرض پوچ برای ضریب رگرسیونی مقدار صفر را پیشنهاد می‌دهد؛ درحالی‌که فرض مقابل برخلاف آن مقدار مثبت یا منفی در نظر می‌گیرد. بدیهی است در آزمون فرض مدنظر مقدار نمونه مورد مطالعه تأثیر مستقیمی بر آماره محاسبه شده دارد و با افزایش آن حتی انحرافاتی بسیار جزئی فرض ادعا پذیرفته می‌شود. به‌طور رسمی‌تر مقدار-احتمال‌ها که براساس برآوردگرهای ثابت بنا شده‌اند دارای رفتار محدودکننده تحت فرض تهی هستند؛ در این صورت در یک آزمون فرض مبتنی بر $H_0:\beta=0$ داریم:

$$\lim_{n\to\infty} P-Value = \lim_{n\to\infty} P(|\hat{\beta}-\beta|<\varepsilon) = \begin{cases} 0, & \beta\neq 0 \\ 1 & \beta = 0 \end{cases}$$

که در آن $\hat{\beta}$ برآوردگر پارامتر شیب خط رگرسیونی براساس داده‌های مورد مشاهده است؛ به‌عبارت دیگر توزیع حدی برآوردگر $\hat{\beta}$، تمام جرم خود را در پارامتر جمعیت دارد.

۴-۱. اندازه اثر در مدل رگرسیون

در مدل‌های رگرسیونی با توجه به وابستگی مقدار-احتمال به اندازه نمونه پیشنهاد می‌شود؛ افزون بر گزارش آن در آزمون‌های آماری حساسیت متغیر وابسته نسبت به تغییرات در متغیر مستقل نیز گزارش شود. یکی از راهکارهای این اقدام، ارائه مقدار اندازه اثر است.

اندازه اثر در مدل رگرسیونی، میزان تعامل یک متغیر پیش‌بینی‌کننده با یک متغیر وابسته را مشخص می‌کند؛ در مثالی ساده، اگر بخواهیم اندازه اثر درمان یک داروی فشار خون را بررسی کنیم، کافی است کاهش میانگین فشار خون را به صورت درصد در نظر گیریم. معمولاً اندازه اثر مدنظر محقق می‌تواند براساس مرور ادبیات پژوهش (مطالعه پژوهش‌های دیگر) یا از نتایج یک مطالعه پایلوت یا براساس دیدگاه خبرگان و متخصصان در آن زمینه به‌دست آید. با توجه به تنوع توابع متغیر وابسته Y و گستره روش‌های رگرسیونی، محاسبه اندازه اثر در زمره علاقه‌مندی‌های پژوهش‌های آماری است.

برای نمونه ویتینگاف[1] و همکاران (۲۰۱۱)، چگونگی محاسبه اندازه اثر برای تابع‌های گوناگون مدل‌های رگرسیونی را انجام داده‌اند که خلاصه نتایج آن در جدول ۱ ارائه می‌شود.

1 Vittinghoff

جدول ۱. مقدار اندازه اثر در مدل‌های رگرسیونی

فرم تابع	اندازه اثر (زمانی که با ضریب β کار میکنیم)
رگرسیون متغیر پیوسته	
$y = f(x)$	یک واحد تغییر در x با میانگین تغییر واحد β در y همراه است
$\ln(y) = f(x)$	به ازای یک واحد افزایش در x، مقدار y به طور متوسط با $100(e^{\beta} - 1)$درصد افزایش می‌یابد. یا به‌طور معادل تقریباً برابر $100\,\beta$ هنگامی‌که $\lvert\beta\rvert < 0.1$
$y = f(\ln(x))$	به ازای یک درصد افزایش در x، مقدار y به طور متوسط با $\ln(1.01)\beta \cong \beta/100$ افزایش می‌یابد
$\ln(y) = f(\ln(x))$	به ازای یک درصد افزایش در x، مقدار y به طور متوسط $\left(e^{\beta \ln(1.01)} - 1\right)$ افزایش می‌یابد یا به‌طور معادل تقریبا برابر β هنگامی‌که $\lvert\beta\rvert < 0.1$
رگرسیون لجستیک	
اگر X عددی باشد	تغییر واحد x با یک تغییر میانگین در شانس Y = 1 با ضریب β همراه است
اگر X دو رشته‌ای باشد	شانس Y = 1 به ازای x = 1 با ضریب β بیشتر از x = 0 است.

در زمینهٔ مقدار-احتمال و اندازه اثر در مدل‌های رگرسیونی مطالعاتی انجام شده است (cf: Coenders & Pawlowsky-Glahn, 2020).

با وجود آن، بحث مقاله پیش‌رو در مورد اندازه‌های اثر فراتر از موارد خاصی است که به‌آنها اشاره می‌شود؛ برای مثال در روش حداقل مربعات،[1] اثر حاشیه‌ای برای هر مقدار متغیر X یکسان است؛ درحالی‌که در مدل‌هایی مانند پروبیت[2] باید مشخص گردد که آیا اندازه اثر میانگین X مانند میانه یا سایر پارامترهای مرکزی دیگر محاسبه خواهد شد؟

مثلاً فرض کنید که یک مدل رگرسیونی پروبیت برازش داده و بخواهیم ضریب متناظر با متغیر X_1 را تفسیر کنیم. محقق می‌تواند سایر متغیرهای مستقل (X) را در یک مقدار مشخص مانند میانه خود، ثابت نگه دارد و سپس با تغییر یک واحدی متغیر X_1، تغییرات متغیر وابسته Y را به‌عنوان تابعی از آن اندازه‌گیری کند. این روش تحلیل حاشیه‌ای به‌ویژه برای مقایسه و ارزیابی میزان تأثیر متغیرها بسیار مفید و انعطاف‌پذیر است.

در مدل‌های غیرخطی که کاملاً متداول هستند، تحلیل حاشیه‌ای یک روش قوی‌تر (و گاهی تنها راه) برای تفسیر اندازه اثر در مقایسه با نگاه کردن به عدد مقدار-احتمال یا بزرگی ضریب رگرسیونی است. برای توضیح بیشتر، اگر X_1 و X_2 متغیرهای توضیحی Y باشند؛ تفسیر مستقیم اثر حاشیه‌ای X_1 بر مبنای ضریب آن نادرست است، زیرا نمی‌توان هم‌زمان مقدار عرض از مبدأ و X_1 را به یک واحد افزایش داد.

الزاماً گزارش اندازه اثر نباید ارائه شدت تغییرات Y نسبت به X برحسب مقیاس واحد/درصد باشد؛ درحقیقت، اگر محقق بتواند اثرها را با چیزی معادل‌سازی کند که درک آن آسان باشد، بسیار مفیدتر خواهد بود. مثال ۲ به‌مفهوم اندازه اثر و میزان درک خواننده از نتیجه پژوهش می‌پردازد.

مثال ۲. فرض کنید یک پژوهشگر درمی‌یابد که خوردن سیب در روز، شانس بیماری را از ۳ درصد به ۲ درصد کاهش می‌دهد. می‌توان گفت «خوردن روزانه سیب، احتمال مراجعه به‌پزشک را به‌طور متوسط ۳۳ درصد کاهش می‌دهد» و یا «یک سیب در رژیم غذایی روزانه، احتمال بیماری را از ۳ درصد به ۲ درصد کاهش می‌دهد». اینگونه گزارش کردن می‌تواند اطلاعات بیشتری ارائه کند، زیرا:

۱. قابلیت مقایسه نقطه‌ای دارد؛ برای مثال (X=0، بدون سیب) در رگرسیون پروبیت در مقابل (X=1، با سیب).

۲. به‌مفهوم سنتی، اندازه اثر را همراه دارد؛ برای مثال $\frac{3\% - 2\%}{3\%} = 33\%$.

۳. بزرگی نسبی اندازه اثر را گزارش می‌دهد؛ برای مثال کاهش بیماری از ۳٪ به ۲٪.

این مثال آزمون فرض، همچنین معناداری عملی یک یافته تحقیق را به حدود و برآورد نقطه‌ای مدنظر خواننده ارجاع می‌دهد و صرفاً به معناداری آزمون فرض نمی‌پردازد.

1. Ordinary Least Squares
2. Probit

همان‌طور که ملاحظه شد، مثال ۲ صرفاً به‌معناداری آزمون فرض نمی‌پردازد بلکه یافته پژوهش را به‌شاخصی قابل فهم برای خواننده ارجاع می‌دهد. تغییر در احتمال بیماری از ۳ درصد به ۲ درصد برای سیاست‌گزاران حوزه سلامت بسیار قابل‌توجه است و حتی می‌تواند برای آن کسی که مراقب سلامتی خود است، بسیار اهمیت داشته باشد. این شیوهٔ تفسیر نتایج آماری نه‌تنها به‌سادگی از نتایج آماری قابل استخراج است، بلکه انتقال نتایج تحقیقات را نیز از سوی محققان به مخاطبان عمومی تسهیل می‌کند.

۵. روش‌شناسی پژوهش

دراین بخش به‌تغییرات مقدار–احتمال، فاصله اطمینان، اندازه اثر و ضریب استاندارد رگرسیونی به‌ازای افزایش اندازه نمونه توجه می‌شود؛ به‌منظور درک بصری بیشتر، روند تغییرات شاخص‌های اشاره شده در قالب رسم نمودار ارائه شده است. به این منظور نخست یک نمونه کوچک به‌اندازه تقریباً یک درصد از کلان‌داده در اختیار انتخاب کرده، سپس بنابر مراحل گام به گامی که در جدول۲ با عنوان الگوریتم تولید نمودار بر پایه اندازه نمونه معرفی شده، نمونه‌های بعدی استخراج شده‌اند. این نمونه‌ها با ضریب ثابتی افزایش می‌یابند که در مثال کاربردی بخش پنج ضریب موردنظر ۵ درصد تعیین شده است.

جدول ۲. الگوریتم تولید نمودار بر پایه اندازه نمونه

برای یک مجموعه دادگان N تایی یک نمودار بر پایه اندازه نمونه‌های افزایشی K تایی:
۱. یک اندازه n_0 برای نمونه اولیه مدل رگرسیون تعیین شود.
۲. یک نمونه تصادفی n_0 تایی از مجموعه داده‌های اصلی انتخاب شود.
۳. مدل رگرسیونی مورد نظر برازش و ضرایب برآورد شده مدل، خطای استاندارد و مقداراحتمال‌ها به‌دست آید
۴. به اندازه مقدار گرد شده از ضریب N/K به‌اندازه نمونه پیشین افزوده شود.
۵. مراحل ۳ و ۴ تکرار شود تا زمانی که از کل داده‌های اصلی استفاده شود.
۶. درنهایت یک نمودار خطی از ضرایب و مقدار–احتمال‌ها در مقابل اندازه نمونه‌های مختلف طراحی شود.

نمودار مقدار–احتمال

دراین نمودار، اندازهٔ مقدار–احتمال متناظر با آزمون فرض تأثیر متغیر مستقل بر وابسته به‌صورت تابعی از اندازه نمونه نشان می‌دهد که از کمترین مقدار نمونه تا بیشترین نمونه مطالعه‌شده به‌دست می‌آید. این نمودار شدت کوچک شدن مقدار–احتمال را به‌ازای افزایش نمونه نشان می‌دهد.

نمودار فاصله اطمینان ضریب رگرسیون

این نمودار، بازه اطمینان ضریب رگرسیونی متغیر مستقل را به‌صورت تابعی از اندازه نمونه ارائه و شدت کوچک شدن خطای استاندارد ضریب رگرسیونی به‌ازای افزایش نمونه را نشان می‌دهد.

نمودار مقدار اندازه اثر

این نمودار، مقدار اندازه اثر متغیر مستقل بر وابسته را به‌صورت تابعی از اندازه نمونه نشان می‌دهد که از کوچک‌ترین تا بزرگ‌ترین اندازه نمونه تغییرپذیر است. این نمودار نیز شدت نوسان مقدار اندازه اثر را به‌ازای افزایش نمونه نشان می‌دهد.

۶. ارائه یک مثال کاربردی

از مسائلی که بسیار مورد توجه محققان در ایران بوده، تصادفات جاده‌ای است. شدت تصادف شاخصی می‌باشد که به‌روش‌های مختلفی محاسبه می‌شود. در این مقاله از فرمول ارائه شده در مرجع Chandler (2013) استفاده شده است که شدت تصادف را براساس رابطه (۴) نشان می‌دهد.

$$TS = 542N_{FAT} + 11N_{INJ} + N_{PDO} \qquad (۴)$$

در آن منظور از TS شاخص شدت تصادف، N_{FAT} تعداد کشته شدگان صحنه تصادف و N_{INJ} تعداد مجروحین صحنه تصادف و NPDO تعداد خودرو حادثه دیده در صحنه تصادف است.

برای بررسی مسئله مقدار–احتمال که در کلان‌داده‌ها وجود دارد و راه‌حل‌های پیشنهادی مختلف از اطلاعات تصادفات جاده‌ای شش‌ماه نخست سال ۱۳۹۸ استفاده شده که این داده‌ها شامل ۱۹۳۵۱ پرونده تصادفات جاده‌ای ثبت شده است. خلاصه آمار متغیرهای اصلی در جدول ۳ ارائه می‌شود.

جدول ۳. توصیف متغیرهای مدل رگرسیون تصادفات جاده‌ای ایران

نام متغیر	توضیح متغیر	میانگین	انحراف معیار
Fatal	تعداد کشته شدگان در صحنه تصادف	۰/۱۹	۰/۵۹۲
Injury	تعداد مصدومان در صحنه تصادف	۱/۴۱	۱/۴۷۵
vehicle	تعداد خودروهای درگیر تصادف	۱/۳۹	۰/۵۷۸
Severity	شاخص شدت تصادف	۱۱۷/۳۲	۳۲۱/۸۱۹

با توجه به هدف این مقاله، مدل رگرسیونی (۵) به کار رفته است:

(۵)

$$LN(Severity) = B_0 + B_1\ Fatal + B_2\ Injury + B_3\ Vehicle$$

نخست از نمونه کامل در اختیار، جهت تخمین معادله رگرسیون استفاده شده و نتایج در جدول ۴ ارائه شده است.

براساس نتایج جدول ۴ در مدل رگرسیون خطی که در رابطه ۴ آمده است، تمام متغیرهای مستقل به‌طور معناداری مؤثر هستند. ضریب تعیین و ضریب تعیین تعدیل یافته مدل برابر ۰/۶۲۳ به‌دست آمدند.

جدول ۴. خروجی مدل رگرسیون خطی برای متغیر تابع لگاریتم شدت تصادف $Y = LN(Severity)$

متغیر (Var)	ضریب رگرسیون (B)	خطای استاندارد (S.E)	ضریب استاندارد رگرسیون (β)	مقدار_احتمال (P_Value)
Constant	2.163	.019	–	<0.001
Fatal	1.947	.012	0.731	<0.001
Injury	1.951	.005	.271	<0.001
vehicle	.299	.012	.026	<0.001

علت برابری تقریبی این دو ضریب به‌علت نمونه بزرگی است که در اختیار داریم و باید این نکته مدنظر محققان قرار گیرد که در کلان داده‌ها صرفاً مقدار ضریب تعیین تعدیل شده گزارش می‌شود.

براساس گزارش جدول ۱ به ازای یک واحد افزایش در تعداد کشته‌شدگان صحنه تصادف، میزان شدت تصادف به‌طور تقریبی ۱۰۸ درصد افزایش می‌یابد؛ همچنین به‌ازای یک واحد افزایش در تعداد مصدومان صحنه تصادف، میزان شدت تصادف تقریباً ۳۱ درصد زیاد می‌شود. به‌همین ترتیب به‌ازای یک واحد افزایش در تعداد خودروهای درگیر تصادف، میزان شدت تصادف به‌طور متوسط ۳ درصد افزایش می‌یابد که به‌عنوان اندازه اثر متغیرهای مستقل مدل رگرسیونی به‌دست آمده‌اند.

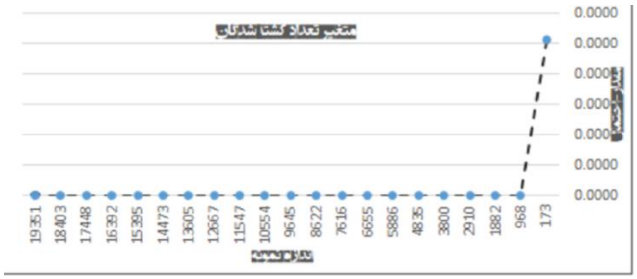
شکل‌های ۱، ۴ و ۷ میزان تأثیر اندازه نمونه بر مقدار–احتمال را نشان می‌دهند؛ چنان‌که نمودار مقدار–احتمال در شکل‌های ۱و۴نشان می‌دهد، افزایش نمونه تأثیری در مورد تصمیم‌گیری برای پذیرش یا ردّ فرضیه ادعا یا افزایش اطمینان محقق نداشته؛ به‌گونه‌ای‌که در بازه نمونه (۱۹۵۳۱– ۱۷۳) میزان اطمینان همواره بیش از ۹۹/۹ درصد بوده است (P-Value<0/0001). در شکل ۹ مقدار–احتمال به‌جز در یک نمونه (n=968)، در سایر موارد میزان اطمینان پذیرش فرضیه ادعا بیش از ۹۵ درصد بوده و برای نمونه‌های بیش از ۷ هزار این اطمینان به بیش از ۹۹/۹ درصد رسیده است.

شکل‌های ۲، ۵ و ۸ نمودار فاصله اطمینان ضریب رگرسیون را به‌ترتیب برای متغیرهای تعداد کشته، مصدوم و خودرو درگیر تصادف به‌ازای اندازه‌های مختلف نمونه گزارش داده‌اند. گفتنی است فاصله‌های اطمینان با سطح معناداری ۵ درصد و به‌صورت ضریب (± 1.96) در خطای استاندارد محاسبه شده‌اند. این نمودارها درک بصری بهتری نسبت به نمودارهای مقدار–احتمال ارائه می‌دهند و افزایش نمونه همچنان برای محقق جهت بررسی وضعیت اختلاف بین برآوردگر و پارامتر قابل استفاده است؛ همچنین نتایج این دو شکل به‌صورت بصری نشان می‌دهد تأثیر اندازه نمونه بر مقدار–احتمال بیشتر از تأثیری است که بر بازه اطمینان می‌گذارد.

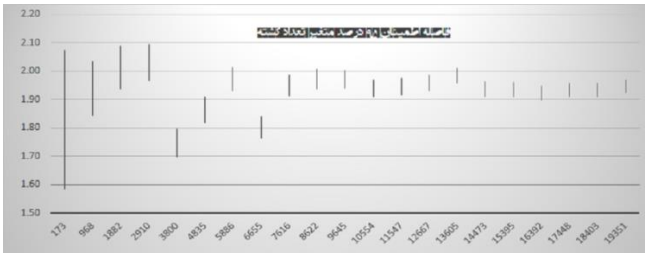
شکل‌های ۳، ۶ و ۹ نمودارهای محاسبه اندازه اثر هستند که به‌ترتیب برای متغیرهای تعداد کشته‌ها، مجروحان و خودروهای درگیر تصادف به‌ازای اندازه‌های مختلف نمونه به‌دست آمده‌اند.

همان‌طورکه ملاحظه می‌شود؛ نمونه‌های بزرگ، تأثیر معناداری بر کاهش یا افزایش مقدار ضرایب رگرسیونی استاندارد شده و اندازه اثر ندارند، بلکه با افزایش نمونه میزان نوسان در تخمین پارامتر نیز کاهش می‌یابد. به‌طورکلی تخمین پارامترهای ضریب استاندارد رگرسیون و اندازه اثر در نمونه‌هایی به‌اندازه ۵ درصد از نمونه اصلی نیز هم‌اندازه تخمین این دو پارامتر در نمونه اصلی است. در اینجا تکیه ما صرفاً به‌نمونه‌هایی هم‌اندازه با ضریب ۵ درصد نمونه اصلی نیست، بلکه در نمونه‌های تقریباً هزارتایی نیز می‌توانند نتیجه‌ای مشابه نمونه‌های بزرگتر ارائه دهند. در این مطالعه با نمونه‌ای هزارتایی می‌توان به یک اشباع نظری در کفایت داده برای دستیابی به شاخص مورد نظر رسید و دیگر

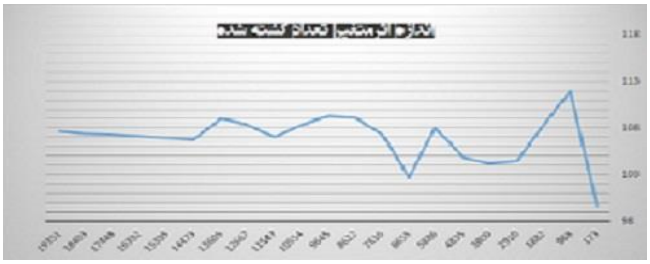
لزومی به مطالعه یک نمونه حدود بیست برابری نیست. گفتنی است که در این مطالعه نوسانات مقدار اندازه اثر و ضریب رگرسیونی یکسان است، زیرا با توجه به اطلاعات جدول ۱ یک رابطه خطی بین این دو ضریب برقرار می‌باشد، اما این رابطه در مدل‌های دیگر رگرسیونی به‌ازای تابع پیش‌بینی مدل می‌تواند یک رابطه غیرخطی باشد؛ با وجود آن، نوسانات هر دو مشابه هستند و در رسم گرافیکی تفاوت محسوسی با یکدیگر ندارند.



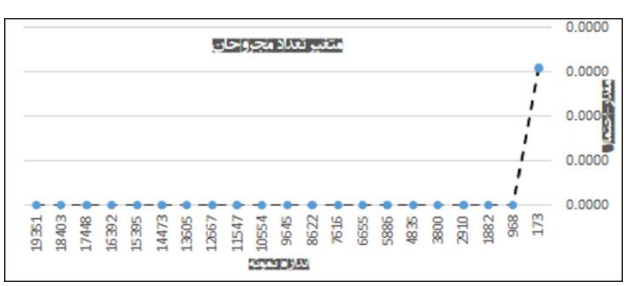
شکل ۲. نمودار مقدار–احتمال تأثیر متغیر تعداد فوتی بر شدت تصادف



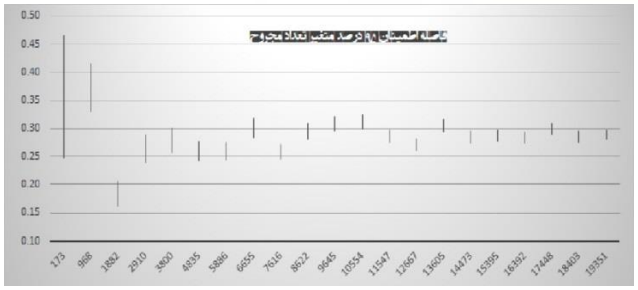
شکل ۳. نمودار فاصله اطمینان متغیر تعداد فوتی بر شدت تصادف



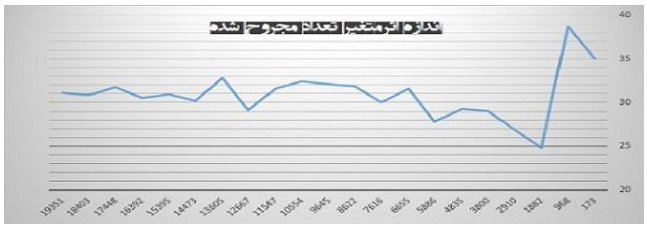
شکل ۴. نمودار اندازه اثر متغیر تعداد فوتی بر شدت تصادف



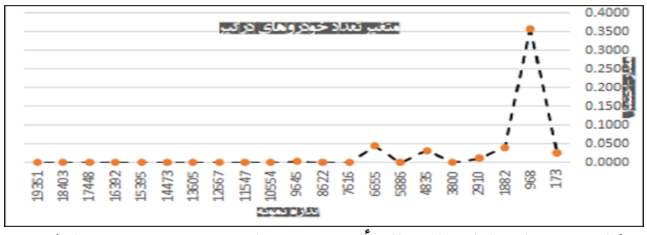
شکل ۵. نمودار مقدار–احتمال تأثیر متغیر تعداد مجروح بر شدت تصادف



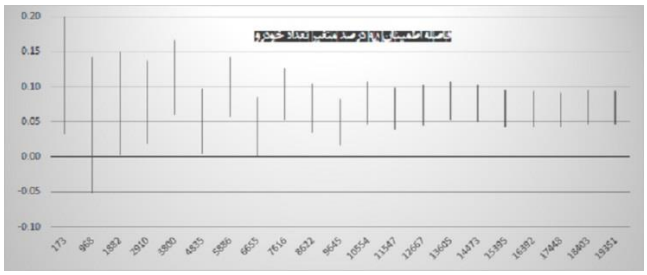
شکل ۶. نمودار فاصله اطمینان متغیر تعداد مجروح بر شدت تصادف



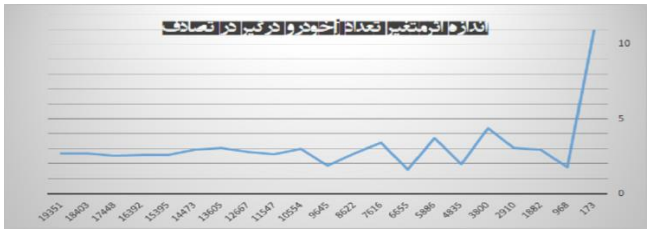
شکل ۷. نمودار اندازه اثر متغیر تعداد مجروح بر شدت تصادف



شکل ۸. نمودار مقدار–احتمال تأثیر متغیر تعداد خودرو بر شدت تصادف



شکل ۹. نمودار فاصله اطمینان متغیر تعداد خودرو بر شدت تصادف



شکل ۱۰. نمودار اندازه اثر متغیر تعداد مجروح بر شدت تصادف

۷. بحث و نتیجه‌گیری

هدف از این پژوهش برجسته کردن یک چالش مهم در نحوه تفسیر آزمون‌های فرض در پژوهش‌های کاربردی است. از نظر آمار کلان‌داده را می‌توان یک نمونه در حدود مگابایت دانست و نمی‌توان از همه انواع داده‌ها در یک انبار از کلان داده استفاده کرد و تنها داده‌های ساختاریافته به کار می‌رود و سایر انواع داده‌ها اعم از نیمه ساختاریافته و غیرساختاریافته به روش‌های تبدیل داده قابلیت بهره‌برداری دارند.

کلان‌داده‌ها صرفاً فرصت‌های بزرگ برای محققان تجربی نیستند، بلکه مشکلات بالقوه‌ای را در تفسیر اهمیت آماری ایجاد می‌کنند. کلان‌داده‌ها توانسته‌اند بستر مناسبی برای داده کاوی نمونه‌ها فراهم کنند و نیز در آزمون‌های فرض، فرضیات ادعا را قوی‌تر استنتاج کنند؛ اما همین بزرگی اندازه نمونه می‌تواند موجب شود تا برخی از تأثیرات جزئی نیز معنادار شوند. این‌گونه تأثیرات، تنها با یک نمونه به‌اندازه کافی بزرگ معنادار می‌شود و در نمونه‌های کوچک‌تر، تصادفی به شمار می‌رود؛ درحالی‌که مقدار اندازه اثر جدا از اینکه تا چه‌اندازه معنادار است، باید مورد توجه محققان قرار گیرد. در واقع افزایش نمونه در راستای افزایش اطمینان در جهت پذیرش فرضیه ادعا تا حد مشخصی مفید است و کوچک‌تر شدن مقدار-احتمال بیش از سه رقم اعشار قابل درک نیست و در تحقیقات کاربردی استفاده نمی‌شود؛ به‌طوری‌که در پژوهش‌های عملی، مقدار-احتمال در بازه (۰/۰۵ و ۰/۰۰۱) استفاده می‌شود و نمونه‌ای که مقدار-احتمال کمتر از ۰/۰۰۱ تولید نماید، عملاً بی‌فایده است. بنابراین می‌توان نتیجه گرفت کلان‌داده، نه‌تنها مزیتی برای افزایش اطمینان در آزمون‌های فرض ندارند بلکه می‌توانند سبب معنادار شدن ادعاهایی شوند که از نظر عملی اهمیت چندانی ندارند و در نمونه‌های نه‌چندان بزرگ در زمره اثرهای تصادفی و خطای نمونه‌گیری قرار می‌گیرند. به این منظور پیشنهاد می‌شود: در مطالعه بر کلان‌داده‌ها، پیش از توجه به اندازه مقدار-احتمال به اندازه اثر به‌عنوان معیاری جهت تشخیص اهمیت عملی آزمون فرض توجه شود و به‌ازای مدل‌های مختلف یک مقدار مرزی برای آن در نظر گرفته شود، به‌منظور تعیین آنکه آیا این مقدار مورد توجه است یا خیر. مقدار مرزی اندازه اثر صرفاً به این منظور استفاده می‌شود که اهمیت عملی تفاوت بین آماره و پارامتر پیش از انجام آزمون فرض تعریف شده باشد؛ زیرا همان‌طور که اشاره شد، بیشتر آزمون‌های فرض روی تفاوتی (رابطه‌ای) انجام می‌شود که در عالم واقع اندازه اثر آن برای پژوهشگران و بهره‌برداران نتایج پژوهش ناچیز برداشت می‌شود.

در این مطالعه چندین رویکرد به‌عنوان گزارش‌های تکمیلی مقدار-احتمال پیشنهاد شد: گزارش اندازه اثر و تعیین مقادیر مرزی جهت تعیین اهمیت اندازه اثر. گزارش بصری فاصله اطمینان برای مقدار-احتمال، اندازه اثر، اندازه ضریب استاندارد رگرسیونی به‌ازای نمونه‌های مختلف. ما به‌عنوان یک راهکار اساسی پیشنهاد می‌کنیم که هنگام کار با نمونه بزرگ به‌جای استفاده یک‌باره از همه داده‌ها، نمونه را به‌نمونه‌های کوچک‌تر (۵ درصد از نمونه کلی یا نمونه‌های هزارتایی) تقسیم کنیم و براساس نتایج نمونه‌ها مقدار-احتمال و اندازه اثر متوسط به‌عنوان ملاک در نظر گرفته شود. در این رویکرد، افزون بر آنکه سبب می‌گردد مقدار-احتمال تحت‌تأثیر اندازه نمونه به‌سمت صفر میل نکند، می‌توان برای ضرایب اشاره شده، یک فاصله و جهت درک بصری یافته‌ها یک نمودار نیز ترسیم کرد. البته به دلیل آنکه این مقاله کوشید که تأثیر نمونه را برای شاخص‌های اشاره شده بررسی کند، از بازنمونه‌گیری نمونه‌های ۵ درصدی و محاسبه اندازه اثر و مقدار-احتمال اجتناب شد. این روش پیشنهادی در مطالعات دیگر نیز به صورت مشابه ارائه می‌شود (Kafadar, 2021).

در مطالعات دیگران به ضعف‌های تفسیر مبتنی بر مقدار-احتمال اشاره شده و به‌عنوان چالشی در تفسیر نتایج آماری با تکیه صرف به مقدار-احتمال ایرادهایی مطرح شده است (Andrade, 2019) که ما به‌طور خلاصه آنها را در بحث اندازه اثر و مقدار-احتمال طرح کردیم. بعضی از محققان به استفاده همزمان از فاصله اطمینان و مقدار-احتمال توصیه کرده‌اند (Betensky, 2019)، ولی این توصیه نمی‌تواند برای نمونه‌های بسیار بزرگ راه‌حل مناسبی باشد؛ زیرا اندازه نمونه به‌تنهایی می‌تواند مقدار-احتمال یک آزمون فرض را به سمت صفر میل

دهد. در این صورت هر فرضیه پژوهشی با اطمینان ۹۹ درصد قابل تأیید خواهد بود.

بنجامین و همکاران (۲۰۱۸) خواستار کاهش آستانه مقدار-احتمال از ۰/۰۵ به ۰/۰۰۵ به‌عنوان راه‌حلی برای نبود تکرارپذیری در علم شده‌اند (Benjamin, 2018)؛ البته این روش نیز به دلایلی که توضیح داده شد، مشکل اساسی نمونه‌های بزرگ را حل نمی‌کند، ضمن آنکه برای نمونه‌های کوچک نیز امکان پذیرش فرضیه پژوهش با سطح اطمینان بالای ۹۹ درصد وجود ندارد. افزون بر آن مثال‌های نقضی برای روش پیشنهادی ایشان در ارائه شده است (Betensky, 2019).

منابع

۱. اسماعیلی، حمید؛ مینا توحیدی؛ سید روح‌اله روزگار و مهدی امیری (۱۳۸۹)، «P-value اصلاح شده، معیاری بهتر از P-value معمولی در فضای پارامتری محدودشده»، دهمین کنفرانس آمار ایران.

۲. اسماعیلی، حمید؛ مینا توحیدی؛ سید روح‌اله روزگار و مهدی امیری (۱۳۹۰)، «P- مقدار معمولی و اصلاح شده، چگونه بهتر قضاوت کنیم؟». مجله علوم آماری، س۵، ش۱، ص۱-۲۲.

۳. بازرگان لاری، عبدالرضا (۱۳۸۴)، رگرسیون خطی کاربردی، شیراز: مرکز نشر دانشگاه شیراز.

۴. برومیده، علی‌اکبر و حسن شاهقلیان (۱۳۸۳)، «با برخی از اشتباهات رایج در تحلیل‌های آماری آشنا شویم»، مجله اندیشه آماری، س۹، ش۱، ص۲۳-۳۱.

۵. پاک‌گوهر، علیرضا (۱۳۹۵)، «مقایسه کارایی روش‌های رده‌بندی کننده رگرسیون لجستیک و رگرسیون درختی برای متغیر وابسته باینری»، نشریه گستره علوم آماری، س۱، ش۲، ص۷-۱۴.

۶. سنمی علمداری، یعقوب (۱۳۹۵)، «مروری بر کلان داده‌ها BIG DATA». اولین همایش ملی نگرشی نوین در مهندسی برق و کامپیوتر.

۷. عارفی‌اصل، سولماز (۱۳۹۷)، «کلان داده، چالش و فرصتی بزرگ پیش روی حرفه حسابداری و حسابرسی». شانزدهمین همایش ملی حسابداری ایران.

۸. شریفیان، نسترن و امید خزاعی (۱۳۹۱)، «توزیع p-مقدار تحت درست بودن فرض مقابل»، مقاله ارائه شده در چهل و سومین کنفرانس ریاضی کشور. دانشگاه تبریز.

9. Andrade, C. (2019), The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives, *Indian journal of psychological medicine*, 41(3), 210–215.

10. Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R. & Johnson, V. E. (2018), Redefine statistical significance. *Nature human behaviour*, 2(1), 6-10.

11. Betensky, R. A. (2019), The p-value requires context, not a threshold, T*he American Statistician*, 73(sup1), 115-117.

12. Cannon, Edmund Stuart & Cipriani, Giam Pietro (2006), Euro-illusion: A natural experiment, *Journal of Money, Credit, and Banking*, 38(5), 1391-1403.

13. Chandler, Brian E; Myers, Matthew; Atkinson, Jennifer E; Bryer, Tom; Retting, Richard; Smithline, Jeff. Venglar, Steven P. (2013), *Signalized Intersections Informational Guide*. United States. Federal Highway Administration, Office of Safety.

14. Chatfield, C. (1995), *Problem Solving: A Statistician's Guide*, Chapman & Hall/CRC.

15. Coenders, Germa & Pawlowsky-Glahn, Vera (2020), On interpretations of tests and effect sizes in regression models with a compositional predictor, *SORT-Statistics and Operations Research Transactions*, 201-220.

16. Cohen, Jacob (1992), Things I have learned (so far), *Presented at the Annual Convention of the American Psychological Association*, 98th, Aug, 1990, Boston, MA, US; Presented at the aforementioned conference, American Psychological Association.

17. Disdier, Anne–Célia; & Head, Keith (2008), The puzzling persistence of the distance effect on bilateral trade, *The Review of Economics and statistics*, 90(1), 37–48.

18. Ghose, Anindya & Yao, Yuliang (2011), Using transaction prices to re-examine price dispersion in electronic markets, *Information Systems Research*, 22(2), 269-288.

19. Goolsbee, Austan & Guryan, Jonathan (2006), The impact of Internet subsidies in public schools, *The Review of Economics and Statistics*, 88(2), 336-347.
20. Greene, WH. (2003), *Econometric analysis*, 4th edn Prentice-Hall, Upper Saddle River, NJ.
21. Hubbard, Raymond & Armstrong, J. Scott. (2006), Why we don't really know what statistical significance means: Implications for educators, *Journal of Marketing Education*, 28(2), 114–120.
22. Kafadar, K. (2021), Statistical significance, p–values and replicability, *The Annals of Applied Statistics*, 15(3), 1081–1083.
23. Kaisler, S. (2013), "Big data: Issues and challenges moving forward," *46th Hawaii International Conference on System Sciences* (HICSS), IEEE.
24. Kiarash Tanha; Neda Mohammadi & Leila Janani (2017), P–value: What is and what is not, Medical Journal Of the Islamic Republic of Iran, (1).
25. Overby, Eric & Jap, Sandy. (2009), Electronic and physical market channels: A multiyear investigation in a market for products of uncertain quality, *Management Science*, 55(6), 940-957.
26. Pakgohar, Alireza; Tabrizi, Reza Sigari; Khalili, Mohadeseh & Esmaeili, Alireza (2011), The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach, *Procedia Computer Science*, 3, 764-769.
27. Rory, Icompliment (2020), Effect Size Is Just as Important as P-Value. Emergency *Medicine News*, 9.
28. Tukey, John W. (1991), The philosophy of multiple comparisons, *Statistical science*, 100-116.
29. Vittinghoff, Eric; Glidden, David V; Shiboski, Stephen, C. & McCulloch, Charles, E. (2011), *Regression methods in biostatistics: linear, logistic, survival and repeated measures models*, Springer Science & Business Media.
30. Windmeijer, F., Liang, X., Hartwig, F. P. & Bowden, J. (2021), The confidence interval method for selecting valid instrumental variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(4), 752-776.
31. Zikopoulos, C. Eaton, D. Deroos, T. Deutsch and G. Lapis (2012), *Undrestanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, United State of America: Mc Graw Hill Companies.