فصلنامه علمی- پژوهشی بیمه سلامت ایران| دوره ۹، شماره ۱، بهار ۱۴۰۵| صفحات ۶۹ تا ۷۰

ملاحظات روش‌شناختی در استفاده از داده‌های حجیم بیمه سلامت در آزمون‌های فرض آماری

زهرا ذوالقدر[۱*] PhD، مریم رحمتی[۲] PhD، هومن شمسینی‌غیاثوند[۳] BA

[۱] مرکز ملی تحقیقات بیمه سلامت، تهران، ایران
[۲] واحد خدمات کارآزمایی بالینی، گروه جمعیت‌شناسی سلامت نافیلد، دانشگاه آکسفورد، آکسفورد، بریتانیا
[۳] گروه آمار، دانشکده علوم پایه، دانشگاه بین‌المللی امام خمینی، قزوین، ایران

[*] **نویسنده مسئول:** زهرا ذوالقدر، پست الکترونیک:zahra.zolghadr@gmail.com

Citation: Zolghadr Z, Maryam R, Shamseini Ghiyasvand H. Methodological considerations in using health insurance big data for statistical hypothesis testing. Iran J Health Insur. 2026;9(1):69-70.

تاریخ دریافت: ۱۴۰۵/۰۱/۳۰ **تاریخ تصحیح:** ۱۴۰۵/۰۲/۲۹ **تاریخ پذیرش:** ۱۴۰۵/۰۳/۰۵ **تاریخ انتشار آنلاین:** ۱۴۰۵/۰۳/۱۰

سردبیر محترم؛

گسترش سامانه‌های اطلاعاتی در نظام بیمه سلامت موجب ایجاد پایگاه‌های داده بزرگ و غنی از اطلاعات تاریخی شده است که فرصت‌های ارزشمندی را برای انجام مطالعات پژوهشی فراهم می‌آورد. پژوهشگران در بسیاری از مطالعات تحلیلی مرتبط با خدمات سلامت و بیمه‌ای از این داده‌ها برای آزمون فرض‌های آماری و بررسی روابط بین متغیرها استفاده می‌کنند. با این حال، استفاده از داده‌های بسیار حجیم در چارچوب آزمون فرض‌های آماری مستلزم توجه به برخی ملاحظات روش‌شناختی مهم است.

در مطالعات مبتنی بر آزمون فرض، حجم نمونه یکی از عوامل تعیین‌کننده در توان آزمون و تفسیر نتایج آماری محسوب می‌شود. همان‌گونه که حجم نمونه کمتر از مقدار مورد نیاز می‌تواند موجب کاهش توان آزمون و عدم شناسایی اثرهای واقعی گردد، حجم نمونه بسیار بزرگ نیز ممکن است باعث معناداری آماری تفاوت‌ها یا روابطی شود که از نظر عملی یا بالینی ناچیز هستند. این پدیده در ادبیات آمار با عناوینی مانند تورم معناداری یا تمایز میان معناداری آماری و معناداری عملی مطرح شده است [۱، ۲].

از منظر نظری، در بسیاری از آزمون‌های آماری پارامتریک، آماره آزمون تابعی از حجم نمونه است. افزایش حجم نمونه موجب کاهش خطای استاندارد برآوردها شده و در نتیجه مقدار آماره آزمون افزایش می‌یابد و احتمال رد فرض صفر بیشتر می‌شود [۳]. برای مثال، در آزمون برابری میانگین‌ها آماره آزمون به صورت زیر تعریف می‌شود [۴]:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

که در آن افزایش مقادیر n_1 و n_2 باعث کوچک‌تر شدن مخرج کسر و در نتیجه بزرگ‌تر شدن مقدار t می‌شود.

در آزمون برابری نسبت‌ها نیز آماره آزمون به شکل زیر است [۴]:

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

که در آن نیز افزایش حجم نمونه سبب کاهش خطای استاندارد و افزایش مقدار آماره آزمون خواهد شد.

به طور مشابه، در آزمون معناداری ضریب همبستگی پیرسون داریم [۴]:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

که نشان می‌دهد با افزایش n مقدار آماره آزمون افزایش یافته و حتی ضرایب همبستگی کوچک نیز ممکن است از نظر آماری معنادار گزارش شوند.

Copyright © 2026 The Author(s); Published by National Center for Health Insurance Research. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[DOI: 10.66224/ihio.9.1.5]

بنابراین اگر امکان تحلیل کل داده‌های موجود در یک پایگاه داده، در یک بازه زمانی و مکانی مشخص برای پژوهشگر میسر باشد، می‌توان این داده‌ها را معادل جامعه مورد مطالعه تلقی کرد و تحلیل را در چارچوب روش‌های توصیفی یا رویکردهای داده‌کاوی انجام داد. در تحلیل‌های داده‌کاوی که مبتنی بر کشف الگوها و روابط پنهان در داده‌ها هستند، افزایش حجم داده‌ها نه تنها محدودیت محسوب نمی‌شود بلکه می‌تواند به بهبود دقت مدل‌ها و کشف الگوهایی پایدارتر کمک کند. در مطالعات تحلیلی مبتنی بر آزمون فرض نیز، توصیه می‌شود پژوهشگران پیش از انجام تحلیل آماری، حجم نمونه مورد نیاز را متناسب با نوع آزمون آماری و با استفاده از فرمول‌ها یا روش‌های تعیین حجم نمونه محاسبه کرده و سپس از طریق نمونه‌گیری مناسب به انتخاب داده‌ها اقدام کنند. این رویکرد می‌تواند از بروز معناداری آماری ناشی از بزرگی بیش از حد حجم داده‌ها جلوگیری کرده و تفسیر نتایج را از نظر علمی معتبرتر سازد.

نتیجه‌گیری

در انتخاب رویکرد مناسب در تحلیل داده‌های حجیم می‌توان این ملاحظات را در نظر گرفت: در صورت دسترسی به کل داده‌های یک جامعه، روش‌های توصیفی و داده‌کاوی با بهره‌گیری از حجم بالای داده، کارایی بالایی در کشف الگوهای پنهان دارند. اما در مطالعات مبتنی بر آزمون فرضیه‌ها، تعیین دقیق حجم نمونه با روش‌های علمی پیش از تحلیل، ضروری است تا از نتایج کاذب و تفسیرهای گمراه کننده جلوگیری شود. رعایت این ملاحظات می‌تواند به ارتقای دقت علمی و تفسیرپذیری نتایج پژوهش‌های مبتنی بر داده‌های ثبتی و اداری در نظام سلامت کشور کمک نماید.

تأییدیه اخلاقی: مطالعه حاضر مورد تأیید مرکز ملی تحقیقات بیمه سلامت قرار گرفت.
تضاد منافع: بین نویسندگان هیچگونه تضاد منافعی وجود ندارد.
سهم نویسندگان: هر سه نویسنده در ایده‌پردازی، نگارش پیش‌نویس و بازبینی انتقادی محتوا مشارکت داشتند. نویسنده اول و دوم تحلیل‌های نظری و مباحث روش‌شناختی را تدوین کردند و نویسنده سوم در سامان‌دهی متن، یکپارچه‌سازی استدلال‌ها و آماده‌سازی نسخه نهایی همکاری نمود. همه نویسندگان نسخه نهایی را خوانده و تأیید کرده‌اند.
منابع مالی: مقاله حاضر مورد حمایت مالی قرار نگرفته است.

References

1. Jacob C. The earth is round (P<. 05). Am Psychol. 1994;49(12):997-1003. doi: 10.1037/0003-066X.49.12.997.
2. Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. J Grad Med Educ. 2012;4(3):279-82. doi: 10.4300/JGME-D-12-00156.1.
3. Lenth RV. Some practical guidelines for effective samplesizedetermination.AmStat.2001;55(3):187-93. doi: 10.1198/000313001317098149.
4. Parsian A. Fundamentals of Probability and Statistics for Science and Engineering Students. Second Edition. Isfahan: Isfahan University of Technology Publishing Center; 2022. [Persian]

[DOI: 10.66224/ihio.9.1.5]