خطاهای اندازه‌گیری در تحقیقات علوم پزشکی

معصومه صادقی: دانشجوی کارشناسی ارشد اپیدمیولوژی، مرکز تحقیقات علوم اعصاب، دانشگاه علوم پزشکی کرمان
مریم نصیریان: دانشجوی کارشناسی ارشد اپیدمیولوژی، مرکز تحقیقات بیماری‌های دهان و دندان، دانشگاه علوم پزشکی کرمان
علی‌اکبر حق‌دوست: * دانشیار، گروه اپیدمیولوژی، مرکز تحقیقات فیزیولوژی، دانشگاه علوم پزشکی کرمان

فصلنامه پایش
سال نهم شماره چهارم پاییز ۱۳۸۹ صص ۴۶۱-۴۵۳
تاریخ پذیرش مقاله: ۱۳۸۹/۳/۱۰
[نشر الکترونیک پیش از انتشار-۲۹ تیر ۱۳۸۹]

چکیده

در تحقیقات علوم پزشکی برای پاسخ به سؤالات و یا رد و قبول فرضیات، نیازمند اندازه‌گیری متغیرها هستیم. همیشه اندازه‌گیری دقیق و با حداقل خطا از دغده‌های اصلی محققان محسوب می‌گردد.

در این مقاله سعی شده به صورت مبسوط، اهم خطاهای اندازهگیری در تحقیقات، شرح و دیدگاه مناسبی به محققان در این خصوص داده شود. در ابتدا خطاها به چهار زیر گروه اصلی تقسیم و در هر یک به تفصیل توضیحاتی ارایه خواهد شد. این دسته‌ها به این شرح هستند:

۱- خطای عدم توافق، زمانی بروز می‌کندکه یک اندازه‌گیری توسط چند ابزار و یا چند فرد اندازه‌گیری کننده که با هم دقیقاً هم نظر نباشند، انجام شود

۲- خطای تصادفی، زمانی بروز می‌کندکه مقدار اندازه‌گیری شده با مقدار واقعی تفاوت داشته باشد ولی همه اندازه‌گیری‌ها به صورت منظم به یک سمت متمایل نباشند

۳- خطای منظم یا سوگیری (Bias)، زمانی دیده می‌شودکه اندازه‌گیری‌ها به یک سمت متمایل باشند

۴- خطای مخدوش‌کنندگی خطایی است که در مطالعات تحلیلی ایجاد شده و روابط را به دلیل تأثیر متغیرهای مخدوش‌کننده کمتر یا بیشتر از مقدار واقعی نشان می‌دهد.

کلیدواژه‌ها: خطاهای اندازه‌گیری، خطای منظم، خطای تصادفی، عدم توافق، مخدوش‌کنندگی

* نویسنده پاسخگو: کرمان، چهار راه طهماسب آباد، خیابان ابن سینا، مرکز تحقیقات فیزیولوژی، دانشگاه علوم پزشکی کرمان
تلفن: ۰۳۴۱-۲۲۶۳۷۸۷
E-mail: ahaghdoost@kmu.ac.ir

[Downloaded from payeshjournal.ir on 2026-10-30]

[DOR: 20.1001.1.16807626.1389.9.4.13.7]

مقدمه

در مطالعات و تحقیقات علوم پزشکی برای پاسخ به سؤال تحقیق و یا فرضیه، اندازه‌گیری متغیرها امری اجتناب ناپذیر است. بعنوان مثال اگر هدف، اندازه‌گیری فشار خون مردان و زنان باشد، باید دو متغیر اصلی جنس و فشار خون سنجیده شوند. در اندازه‌گیری هر دو این متغیرها ممکن است اشتباه رخ دهد؛ البته پر واضح است که اندازه‌گیری جنس، بسیار کم خطاتر از اندازه‌گیری فشار خون خواهد بود. برای شروع بحث دو اصطلاح پارامتر و آماره شرح داده می‌شوند. پارامتر، مقدار یک متغیر در کل جامعه است که مقدارآن ثابت است [۱].

از آنجا که در بسیاری از موارد، حجم جامعه مورد مطالعه بسیار بزرگ و حتی نامحدود است و انجام مطالعه بر روی کل جامعه مستلزم بکارگرفتن هزینه سنگین و وقت زیاد است و از طرفی بدلیل ابعاد گسترده کار، کنترل، صحت و دقت به نحو مطلوب امکان‌پذیر نیست، بهتر است به مطالعه نمونه‌ای ازافراد جامعه اقدام گردد [۲، ۳]. در واقع آماره، مقدار یک متغیر در نمونه است که بر اساس آن و با استفاده از روش‌های برآورد آماری می‌توان مقدار پارامتر را تخمین زد [۱].

به عنوان مثال، میانگین فشار خون کل زنان و مردان جامعه پارامتر هستند و می‌توان آنها را از طریق اندازه‌گیری فشار خون یک زیر گروه تصادفی از جامعه که نماینده کل جامعه باشند برآورد کرد. درصورتی برآوردهای معتبر و قابل اعتمادی از پارامترهای جامعه یا بعبارتی اندازه‌گیری دقیق خواهیم داشت که اندازه‌گیری با کمترین خطا انجام شود. هدف اصلی در تحقیقات و مطالعات، جمع‌آوری اطلاعات معتبر جهت برآورد دقیق پارامترهای جامعه است [۲، ۴، ۵]، ولی در هر مطالعه و یا اندازه‌گیری ممکن است خطا رخ دهد.

به منظور اطمینان از نتایج مطالعه این خطاها باید کاهش یابند، در غیر این صورت نتایج همراه با خطا خواهند بود و بنابراین شناختن راه‌های کاهش این خطاها برای هر محققی لازم به نظر می‌رسد. در این مقاله سعی بر آن است به زبانی ساده و با مثال‌های کاربردی، انواع خطاهای اندازه‌گیری شرح و راه‌های مقابله وکاهش آنها در مراحل مختلف تحقیقات مورد بررسی قرارگیرد.

خطاهای اندازه‌گیری (Measurement Errors)

خطاهای اندازه‌گیری از زوایای مختلف تقسیم بندی می‌شوند. در این مقاله تقسیم بندی به گونه‌ای انتخاب شده است که مباحث فنی کمتر دخیل باشند. ساده‌ترین شکل تقسیم بندی خطاها به شرح زیر است:

۱- خطاهای عدم توافق (Disagreement Errors)

۲- خطاهای تصادفی (Random Errors)

۳- خطاهای سیستماتیک (Systematic Errors)

۴- خطاهای مخدوش‌کنندگی (Confounding Errors)

خطاهای عدم توافق (Disagreement Errors)

این نوع خطا زمانی رخ می‌دهدکه یک متغیر در نمونه‌های مورد مطالعه توسط چند ابزار متفاوت و یا افراد متعدد اندازه‌گیری شوند، بطوری که پاسخ‌های آنها حتی در اندازه‌گیری نمونه های مشابه متفاوت باشند [۱]. چنانچه در مثال اندازه‌گیری فشار خون مجبور به استفاده از دستگاه‌های اندازه‌گیری فشارخون با مارک‌ها و مدل‌های مختلف (نوع عقربه‌ای و نوع جیوه‌ای) جهت اندازه‌گیری فشار خون یک نفر باشید و مقادیر بدست آمده تشابه کمی با هم داشته باشند، بیان می‌گردد که عدم توافق وجود دارد. البته ممکن است این خطا به دلیل تفاوت در دقت و یا تجربه افراد اندازه‌گیری‌کننده به وجود آید. راهکارهای پیشنهادی جهت کاهش این خطا عبارتند از:

۱- کاهش تعداد ابزار اندازه‌گیری و یا افراد اندازه‌گیری کننده.

۲- سعی در عینی (Objective) کردن شیوه سنجش؛ به عبارتی کمتر وابسته به نظرات افراد اندازه‌گیری کننده باشند.

۳- کالیبراسیون: یعنی نزدیک کردن پاسخ ابزار به اندازه‌های واقعی (Gold Standard). در مثال فوق، دستگاه اندازه‌گیری فشار خون وقتی کالیبره خواهد بود که در شروع اندازه‌گیری، عقربه دستگاه عدد صفر را نشان دهد، البته اصطلاح کالیبراسیون، مختص وسایل اندازه‌گیری نیست، بلکه در مورد افراد نیز می‌توان آن را انجام داد. در چنین شرایطی تأکید می‌شود با جلسات آموزشی، پاسخ‌های افراد اندازه‌گیری کننده با پاسخ‌های یک فرد خبره و متخصص درآن زمینه مقایسه و با تمرین، این نقطه نظرات و جواب‌ها به هم نزدیک شوند. به عنوان مثال برای افزایش توافق، در یک مصاحبه تمرینی، فردی به عنوان بیمارنما در نظر گرفته شده و از مصاحبه کنندگان خواسته می‌شود تا با وی صحبت کنند. احتمالاً در جلسه اول، نظرات مصاحبه کنندگان تفاوت زیادی باهم خواهند داشت. برای افزایش توافق باید این جلسات تمرین را تا زمانی ادامه داد که جواب و نظرات افرادمصاحبه گر به هم نزدیک شوند و در واقع به یک توافق جمعی و پاسخ تقریباً یکسان دست یابند.

[Downloaded from payeshjournal.ir on 2026-10-30]

[DOR: 20.1001.1.16807626.1389.9.4.13.7]

۴- استاندارد سازی (Standardization): تحت بعضی از شرایط Gold Standard وجود ندارد. در این مواقع امکان انجام کالیبراسیون وجود ندارد. در چنین موقعیت‌هایی می‌توان با استفاده از یک روش ثابت و تعریف شده (استاندارد)، میزان توافق افراد را افزایش داد. در واقع در این روش افراد بر اساس یک روش تعریف شده و مشخص (پروتکل)، عمل اندازه‌گیری را انجام داده و به این ترتیب پاسخ‌های مشترک (توافق) آنها افزایش می‌یابد. اما همیشه این سؤال وجود دارد که آیا می‌توان میزان توافق را اندازه‌گیری کرد؟ پاسخ این سؤال مثبت است. جهت تعیین میزان توافق، در صورت کمی (عددی) بودن متغیر مورد نظر، از ضریب همبستگی پیرسن و در حالتی که متغیر تحت مطالعه کیفی باشد، از ضریب کاپا استفاده می‌شود [۶].

ضریب همبستگی پیرسن عددی بین ۱- تا ۱+ است. زمانی که بین دو متغیر عددی (در اینجا بین دو اندازه‌گیری) ارتباطی وجود نداشته باشد، این ضریب به صفر نزدیک خواهد شد. ضریب همبستگی نزدیک به ۱+ نشان‌دهنده همبستگی مثبت و قوی است. از این رو، در زمینه بررسی توافق انتظار داریم که ضریب مذکور به یک نزدیک باشد. معمولاً ضریب بالای ۰/۷ قابل قبول است و ضرایب کمتر از این عدد نشان‌دهنده عدم توافق کافی در بین دو اندازه‌گیری است.

کاپا (Agreement Beyond the Chance=Kappa) که برای بررسی توافق برای متغیرهای اسمی (گروه بندی شده) بکار می‌رود و همانطور که از اسم آن بر می‌آید، کاپا یعنی توافق فراسوی شانس. برای درک بهترآن به این مثال توجه کنید: اگر از دو نفری که هیچ آشنایی به علوم پزشکی ندارند خواسته شود که از بین ۱۰۰ بیمار بستری در بخش داخلی یک بیمارستان، بیمارانی را که طحال بزرگ دارند، صرفاً براساس نظرات خود و بدون انجام هیچ گونه معاینه‌ای تشخیص دهند، طبیعی است که این دو فرد در تعدادی از موارد، بر اساس شانس و تصادف توافق خواهند داشت (یعنی هر دو تعدادی از بیماران را دارای طحال بزرگ معرفی می‌کنند و یا هر دو تعدادی بیمار را دارای طحال طبیعی می‌شناسند). اما اگر در مثال بالا فرض نمایید که یک استاد رشته پزشکی بخواهد تبحر دانشجویان خود را در تشخیص طحال بزرگ ارزیابی کند، بدین منظور، این ۱۰۰ بیمار، ابتدا توسط خود استاد و سپس توسط دانشجو معاینه می‌شوند. نتایج فرضی را در جدول ۲×۲ (توافقی) مشاهده فرمایید:

تشخیص استاد

تشخیص دانشجو	+	−
+	a (۲۰)	b (۵)
−	c (۱۵)	d (۶۰)

جدول شماره ۱- نتایج معاینات استاد و دانشجو

a: هم استاد و هم دانشجو تشخیص مثبت داده‌اند. (توافق)

b:دانشجو تشخیص مثبت ولی استاد آنها را منفی تشخیص داده است (عدم توافق)

c: استاد تشخیص مثبت داده، ولی دانشجو نتوانسته آنها را تشخیص دهد (عدم توافق)

d: هم استاد و هم دانشجو تشخیص منفی داده‌اند (توافق)

(+): بیمارانی که طحال آنها بزرگ تشخیص داده شده است

(-): بیمارانی که طحال آنها در حد طبیعی تشخیص داده شده است

همانطور که مشاهده می‌فرمایید در ۸۰ (۶۰+۲۰) مورد از ۱۰۰ بیمار جواب استاد و دانشجو یکسان است؛ یعنی هر دو ۲۰ بیمار را دارای طحال بزرگ و ۶۰ نفر را منفی تشخیص داده‌اند. در ۱۵ مورد استاد تشخیص طحال بزرگ (+) داده است ولی دانشجو نتوانسته تشخیص دهد و در ۵ مورد دانشجو ادعا کرده طحال بیمار بزرگ است، ولی استاد این یافته را ردکرده است (تشخیص منفی داده است). طبق این نتایج می‌توان گفت توافق آنها برابر است با ۸۰٪.

$$\text{توافق (Agreement)} = \frac{\text{توافق دوطرفه}}{\text{کل بیماران}} = \frac{80}{100} = 0.8$$

بسیار ساده انگاری خواهد بود اگر اظهار شودکه این توافق صرفاً به دلیل شانس بوده است. بدیهی است توافق آنها صرفاً به دلیل شانس و تصادف نبوده، بلکه توافق مشاهده شده از مجموع توافق بر اساس شانس و توافق فراسوی شانس و تصادف یا بر اساس آگاهی حاصل گردیده است؛ به عبارت بهتر: [(کاپا)توافق فراسوی شانس + توافق بر حسب شانس = توافق ظاهری]

Agreement=agreement by chance + agreement beyond the chance (Kappa)

در این مثال اگر برحسب شانس ۱۰ نفر در خانه a و ۲۰ نفر در خانه d قرارگیرند توافق براساس شانس و تصادف برابر است با: ۰/۳ (۳۰/۱۰۰) یا ۳۰٪.

با توجه به نتایج جدول مقدار توافق فراسوی شانس در مثال فوق عبارت است از: ۰/۵-۰/۳=۰/۸

به بیان بهتر در این مثال از ۸۰ مورد توافق بین استاد و دانشجو ۵۰ مورد به دلیل مهارت دانشجو در تشخیص بوده و ۳۰ مورد

[Downloaded from payeshjournal.ir on 2026-10-30]
[DOR: 20.1001.1.16807626.1389.9.4.13.7]

برحسب شانس. این ساده‌ترین شکل محاسبه توافق بین دو نفر است، ولی فرمول کاپا با همین منطق و کمی پیچیده‌تر است و به عبارتی مقدار واقعی کاپا با عدد ۰/۵ کمی متفاوت خواهد بود.

همانطور که در مثال بالا ملاحظه کردید کاپا همیشه از توافق ظاهری کمتر است. به عنوان یک راهکار کلی، اگر کاپا کمتر از ۰/۲ باشد، درجه توافق مشاهده شده بد در نظر گرفته می‌شود. اگر کاپا بین ۰/۴-۰/۲ باشد، درجه توافق در حد متوسط، اگر کاپا بین ۰/۶-۰/۴ باشد توافق مشاهده شده خوب و اگر کاپا بیشتر از ۰/۶ باشد، درجه توافق مشاهده شده در حد عالی است.

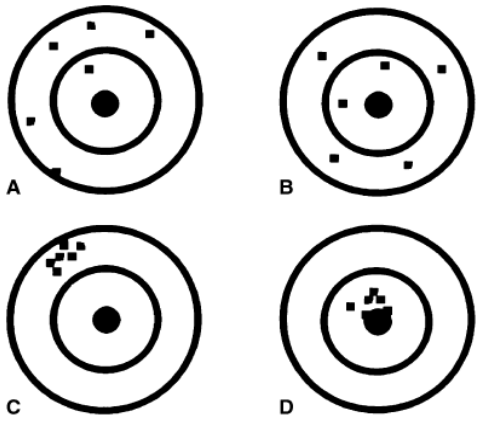
نکته آخر آن که محاسبه ضریب توافق چه برای متغیرهای اسمی و چه برای متغیرهای عددی زمانی که بیش از ۲ مشاهده‌گر و یا ابزار اندازه‌گیری وجود دارد، کمی دشوار و محاسبات مربوطه پیچیده خواهد شد، ولی مفهوم و تفسیر یافته‌های این روش‌های آماری، همانند تفسیری است که از شاخص‌های بیان شده در این مقاله ارایه گردیده است.

خطاهای تصادفی (Random Errors)

خطاهایی که الگوی خاصی ندارند و بر اساس شانس و تصادف ایجاد می‌شوند و مشاهدات را زیادتر یا کمتر از مقدار حقیقی نشان می‌دهند. در واقع خطای تصادفی بوسیله تغییرات ناشناخته وغیرقابل پیش بینی در وسایل اندازه‌گیری و یا تغییر شرایط محیطی اتفاق می‌افتد. به عنوان مثال اندازه‌گیری فشارخون در محیط پر سرو صدا احتمال خطای تصادفی را زیاد می‌کند، چرا که دو حالت تقریباً متضاد ممکن است رخ دهد؛ صداهای مزاحم از یک طرف باعث می‌شوندکه اولین صدای کورتکوف شنیده نشده در نتیجه فشار خون فرد کمتر از واقعیت گزارش گردد و یا این امکان وجود داردکه سر و صدای اضافه همانند صدای کورتکوف شنیده شده و قبل از وقوع واقعی، اولین صدای کورتکوف فرد احساس کند که صدا را شنیده است و بدین ترتیب، فشار خون فرد بیشتر از واقعیت گزارش شود. بعبارتی تحت این شرایط ممکن است فشار خون را یک بار بیشتر و یک بار کمتر از واقعیت گزارش نمود. مثال ساده دیگری از این دست ممکن است اندازه‌گیری وزن توسط یک ترازو باشد؛ اگر در پنج بار اندازه‌گیری وزن یک فرد، به پنج عدد نسبتاً متفاوت برسیم، می‌توانیم فرض کنیم که حداقل خطای تصادفی اندازه‌گیری ما را تحت تأثیر قرار داده است. به بیان دیگر خطای تصادفی زمانی رخ می‌دهد که تفاوت بین مقادیر مشاهده شده و مقادیر واقعی بر اساس شانس و تصادف ایجاد شده و گاهی مثبت و گاهی منفی باشد. از خصوصیات مهم و اصلی خطای تصادفی جهت دار نبودن این تفاوت‌ها است.

ثبات یا پایایی (Reliability)که به آن میزان تکرارپذیری نیزگفته می‌شود عبارت است از احتمال بدست آمدن نتایج تقریباً یکسان و مشابه در صورت تکرار اندازه‌گیری [۵، ۷]. به عنوان یک مثال ساده می‌توان پایایی را شبیه مهارت تیراندازی دانست که در تکرار تیراندازی‌ها، یک نقطه ثابت را هدف قرار می‌دهد و تیرهای شلیک شده وی بسیار نزدیک به هم به سیبل برخورد می‌کنند [۸].

با کاهش خطای تصادفی، میزان پایایی افزایش می‌یابد. با این توضیح برای سنجش میزان خطای تصادفی بایست میزان پایایی اندازه‌گیری را تعیین نمود.



شکل شماره ۱- نتایج افراد در تمرین تیراندازی. در سیبل A و B تیرهای شلیک شده از هم فاصله زیاد دارند و در نتیجه می‌توان بیان نمود که دارای خطای تصادفی زیاد و در نتیجه پایایی کم هستند. در سیبل C پایایی بالا است، چراکه شلیک‌ها به هم نزدیک هستند، ولی چون برآیند شلیک‌ها از مرکز سیبل فاصله دارد، دارای سوگیری و یا خطای منظم است. البته در سیبل A، علاوه بر خطای تصادفی، خطای منظم نیز وجود دارد. در سیبل D شلیک‌ها به هم نزدیک بوده و برآیند آنها نیز نزدیک به مرکز است، پس می‌توان بیان نمود که در این سیبل، خطای منظم و تصادفی کم است و در نتیجه تیراندازی دقیق و معتبر (Precise and Valid) بوده است.

برای کاهش خطای تصادفی و افزایش پایایی، باید ابزار و شیوه اندازه‌گیری دقیق‌تری بکار برد و فرد اندازه‌گیری کننده با روش استاندارد کار نماید. البته با افزایش حجم نمونه و یا تکرار اندازه‌گیری و استفاده از میانگین چند بار اندازه‌گیری می‌توان خطای تصادفی را به میزان زیادی کاهش داد. به عنوان مثال بر اساس استانداردهای بالینی، در صورتی می‌توان فردی را دارای فشار خون بالا معرفی کردکه حداقل دو بار فشار خون فرد، اندازه‌گیری و

[Downloaded from payeshjournal.ir on 2026-10-30]
[DOR: 20.1001.1.16807626.1389.9.4.13.7]

در هر دو بار فشار خون فرد بالا باشد، چرا که اندازه‌گیری یک بار فشار خون، پایایی بالایی ندارد و دلیل آن تغییرات تصادفی قابل ملاحظه فشار خون افراد در طول روز و در روزهای متوالی است.

خطای منظم (Systematic Error)

اسامی متعدد دیگری نیز به این خطا اطلاق شده که از آن جمله می‌توان به تورش، سوگیری یا (Bias) اشاره نمود. خطای منظم برخلاف خطای تصادفی الگو و جهت خاصی دارد. در مثال تیر اندازی (شکل شماره ۱) فرد C مرتکب خطای سیستماتیک شده است، به این دلیل که اگرچه بطور متمرکز (به یک نقطه) شلیک کرده و خطای تصادفی آن کم است، ولی تمام آنها دور از مرکز بوده است. برای روشن‌تر شدن موضوع به چند مثال توجه فرمایید.

مثال ۱: اندازه‌گیری فشار خون با استفاده از دستگاهی که بازوبند (کاف) آن کوچک است؛ طبیعی است که این دستگاه، فشار خون را همیشه بیشتر از واقعیت نشان خواهد داد، چراکه فشار کافی به شریان‌ها برای انسداد جریان وارد نمی‌کند. بنابراین حتی اگر چندین بار با این دستگاه اندازه‌گیری را تکرار کنیم، باز هم میانگین فشار خون‌های مشاهده شده بیشتر از مقدار واقعی خواهد بود.

مثال ۲: اگر از فردی که شنوایی اوکم است، خواسته شودکه فشار خون افراد را اندازه بگیرد، طبیعی است که وی فشار خون‌ها را کمتر گزارش می‌دهد، چراکه ضعف شنوایی وی باعث می‌شود تا اولین صدای کورتکوف را که خفیف است نشنود. اگر این شخص ۵۰ بار هم فشار خون یک فرد را اندازه بگیرد، برآیند نتایج گزارش شده تغییری کمتر از واقعیت خواهد بود. به عبارتی، خطای گزارش این شخص در یک جهت است و با تکراراندازه‌گیری و یا افزایش حجم نمونه نمی‌توان خطای منظم را رفع و یا حتی مقدار آن را کاهش داد. در شکل شماره ۱، در سیبل A هم خطای منظم و هم خطای تصادفی دیده می‌شود و بیانگر امکان وقوع همزمان این دو خطا با هم است. مشاهده می‌کنید که تمام شلیک‌های فرد A دور از مرکز و پراکنده است، به عبارتی علاوه براین که تفنگ این فرد تنظیم نبوده (خطای منظم)، هدف گیری وی نیز خوب نبوده است و بنابراین شلیک‌های وی از هم دور و پراکنده بوده‌اند.

از زوایای مختلف می‌توان انواع زیادی را برای سوگیری‌ها برشمرد [۸]، اما به صورت کلی در این مقاله به دو دسته مهم سوگیری اشاره و زیرشاخه‌های آنها مورد بحث قرار می‌گیرند. این دو دسته مهم سوگیری عبارتند ازسوگیری اطلاعات (Information Bias) و سوگیری انتخاب (Selection Bias).

سوگیری اطلاعات

زمانی اتفاق می‌افتد که اعداد اندازه‌گیری شده اشتباه بوده و با مقدار واقعی تطابق نداشته باشد، به شکلی که برآیند اندازه‌گیری‌ها کمتر یا بیشتر از واقعیت باشند. علل وقوع این سوگیری عبارتند از:

الف- دقیق نبودن ابزار اندازه‌گیری، مثلاً استفاده از دستگاه فشار خونی که کاف آن متناسب با جثه افراد نباشد و یا ترازویی که کالیبره نبوده و وزن‌ها را ۵ کیلوگرم بیش از واقعی گزارش نماید.

ب- عدم دقت از سوی فرد مشاهده کننده یا اندازه‌گیری کننده، مثلاً انجام اندازه‌گیری فشار خون توسط فردی که ضعف شنوایی دارد.

ج- خطا داشتن فرد مورد مشاهده و یا متغیر مورد اندازه‌گیری، مثلاً اگر قبل از اندازه‌گیری فشار خون، افراد استرس داشته و یا به اندازه کافی استراحت نکرده باشند که در این شرایط انتظار داریم فشار خون آنها حتی در صورت اندازه‌گیری دقیق بیش از حد طبیعی باشد. مشابه چنین خطایی زمانی رخ می‌دهد که مثلاً نمونه خون فرد برای اندازه‌گیری قند، مدت طولانی در دمای عادی نگاه داشته شود؛ در این صورت گلبول‌های خون مقداری از قند خون را مصرف می‌کنند و در نتیجه مقدار قند کمتر از واقعی گزارش خواهدشد.

سوگیری انتخاب

زمانی که نمونه انتخاب شده جهت مطالعه نماینده خوبی از کل جامعه نباشد، سوگیری انتخاب را خواهیم داشت. به عنوان مثال جهت اطلاع از غلظت آفلاتوکسین در ده تن پسته نیازی به آزمایش تمام آنها نیست، کافی است غلظت آفلاتوکسین را در یک نمونه یک کیلویی اندازه بگیریم. ولی گاهی اوقات این یک کیلو نماینده کل ده تن پسته نیست و ممکن است میزان آفلاتوکسین درآن زیادتر و یا کمتر از واقعیت باشد. در اینجا اصطلاحاً گفته می‌شود سوگیری انتخاب اتفاق افتاده و این یک کیلو نماینده و نمونه خوبی از ده تن پسته نبوده است (در اصطلاح عام مشت نمونه خروار نبوده) و در نتیجه نمی‌توان نتایج بدست آمده از مطالعه را به کل جامعه تعمیم داد. چنین خطایی در مطالعات بسیار رخ می‌دهد و ممکن است افراد گزینش شده برای ورود به مطالعه، مشابه کل جامعه نباشند. بعبارتی سوگیری درنمونه گیری (Sampling Bias) باعث این خطا شده است. از طرفی ممکن است به دلیل از دست دادن نمونه‌ها در طول مطالعه (Lost to Follow up) و یا نمونه‌های بدون پاسخ (Missing Data or Non-Respondents) نمونه‌های نهایی که

[Downloaded from payeshjournal.ir on 2026-10-30]
[DOR: 20.1001.1.16807626.1389.9.4.13.7]

وارد تجزیه و تحلیل می‌شوند با کل جامعه به صورت جهت داری متفاوت باشند.

مفهوم اعتبار یا روایی

اعتبار(Validity) نشان‌دهنده این مهم است که اندازه‌گیری چقدر به نتایج صحیح و درست می‌انجامد [۸]. زمانی یک اندازه‌گیری معتبر محسوب می‌گردد که هیچکدام از خطاهای تصادفی و یا منظم را نداشته باشد [۹] (مانند قدرت تیر انداز برای زدن به هدف شکل۱-D) [۷]. البته این تعریف مورد قبول همه نیست و عده‌ای اعتبار را معادل نداشتن خطای منظم می‌دانند و صحت (Accuracy) را معادل نداشتن هر دو خطای منظم و تصادفی بیان می‌کنند. ولی به هر شکل با قبول هر تعریف، می‌توان اعتبار را به دو جزء داخلی و خارجی تقسیم کرد.

اعتبار داخلی (Internal Validity) یعنی اعداد بدست آمده از مطالعه در یک گروه چقدر به اعداد واقعی در همان گروه نزدیک است. زمانی که سوگیری در اطلاعات وجود نداشته باشد و اندازه‌گیری‌های انجام شده به صورت منظم از واقعیت به یک سمت منحرف نشده باشند، عملاً می‌توان گفت که اعتبار داخلی کافی در نتایج وجود دارد.

اعتبار خارجی (External Validity) یعنی نتایج بدست آمده درگروه مورد مطالعه چقدر قابل تعمیم به کل جامعه هستند. پس اعتبار خارجی مقدار دقت در تعمیم دادن نتایج به جامعه را نشان می‌دهد و با میزان سوگیری در انتخاب، نسبت معکوس دارد و این یعنی با افزایش سوگیری در انتخاب اعتبار خارجی کاهش می‌یابد.

اشکالات ایجاد شده به دلیل خطاهای منظم یا سوگیری‌ها بسیار خطرناک‌تر از خطاهای تصادفی هستند. افزایش حجم نمونه نمی‌تواند به کاهش خطاهای منظم منجر شود و روش‌های ساده آماری برای برآورد میزان چنین خطاهایی وجود ندارد. البته روش‌های نسبتاً پیچیده‌ای وجود دارند که تا حدودی ممکن است میزان اثر سوگیری را در مطالعه نشان‌دهند، از جمله این روش‌ها آنالیز حساسیتی (Sensitive Analysis) است که به دلیل نیاز به مقدمات زیاد بحث پیرامون آن خارج از هدف این مقاله است.

خطای مخدوش کنندگی (Confounding Error)

آخرین نوع خطایی که در این مقاله شرح داده خواهد شد، مخدوش کنندگی است که باعث می‌شود تا ارتباط ظاهری یک عامل خطر با پیامد کمتر یا بیشتر از واقعیت دیده شود [۱۰]. مخدوش کننده‌ها فقط در مطالعات تحلیلی وجود دارند [۸]، اما متأسفانه در بسیاری از موارد، به اشتباه در مطالعات توصیفی از اصطلاح مخدوش کنندگی استفاده می‌شود؛ به طور مثال در مطالعه‌ای که هدف آن تعیین میانگین فشار خون مردم شهرکرمان است، اگر سن را به عنوان عامل مخدوش کننده بیان کنیم مرتکب اشتباه شده‌ایم، زیرا این مطالعه توصیفی است و در آن مخدوش کنندگی معنی ندارد؛ اگر مثلاً میانگین سن مردم در جامعه به صورت واضحی کمتر از میانگین سنی نمونه‌ها است و در نتیجه میانگین فشار خون بیش از واقعیت بدست آمده، اشکال در نمونه‌گیری(سوگیری در نمونه گیری) است و نه اثر مخدوش کنندگی سن.

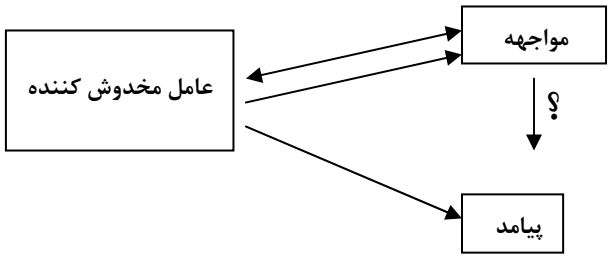
به عنوان یک راهکار کلی، در صورتی می‌توانیم یک عامل را به عنوان عامل مخدوش کننده در نظر بگیریم که ارتباط بین دو متغیر را کمتر یا بیشتر از واقعیت نشان دهد، این در حالی است که در مطالعات توصیفی، هدف، تعیین ارتباط بین متغیرها نیست بلکه در این نوع مطالعات، اطلاعات مربوط به فراوانی رخدادی مشخص، شرح داده شده و توزیع آن در زیر گروه‌ها بر اساس متغیر شخص، زمان و مکان را نشان می‌دهند [۱۱].

بطور کلی، برای اینکه یک عامل را مخدوش کننده بنامیم چند شرط لازم است [۸، ۱۱، ۱۲] که عبارت‌اند از:

۱- عامل مخدوش کننده با پیامد (متغیر وابسته) رابطه علیتی داشته باشد

۲- بین عامل مخدوش کننده و مواجهه (متغیر مستقل) یک رابطه و همراهی وجود داشته باشد

۳- عامل مخدوش کننده یک عامل واسطه‌ای بین مواجهه و پیامد نباشد



شکل شماره ۲- نمای شماتیک ارتباط بین عامل مخدوش کننده، مواجهه و پیامد

[Downloaded from payeshjournal.ir on 2026-10-30]
[DOR: 20.1001.1.16807626.1389.9.4.13.7]

مثال: فرض کنید مطالعه‌ای به دنبال تعیین شدت ارتباط بین مصرف سیگار و سرطان ریه است. برای پاسخ به این سؤال، دو گروه شامل ۱۰۰ نفر سیگاری و ۱۰۰ نفر غیرسیگاری به مدت ۱۰ سال تحت مطالعه قرار گرفته و بعد از ۱۰سال نتایج زیر مشاهده شده است (جدول شماره ۲).

	مبتلا به سرطان	سالم	جمع
سیگاری	۴۰	۶۰	۱۰۰
غیرسیگاری	۱۰	۹۰	۱۰۰
جمع	۵۰	۱۵۰	۲۰۰

جدول شماره ۲- میزان بروز سرطان ریه در دو گروه مختلف از نظر مواجهه با سیگار

همانگونه که مشاهده می‌فرمایید در گروه سیگاری بعد از ۱۰ سال ۴۰ نفر مبتلا به سرطان ریه و در گروه غیرسیگاری بعد از ۱۰ سال،۱۰ نفر به سرطان ریه مبتلا شده‌اند. بعبارتی خطر ابتلا ۱۰ ساله به سرطان در گروه سیگاری ۴۰٪ (۴۰/۱۰۰) و در گروه غیرسیگاری ۱۰٪ (۱۰/۱۰۰) است. با توجه به خطر ابتلا در این دو گروه خطر نسبی (Relative Risk) را می‌توان به این صورت محاسبه کرد:

$$\frac{\text{خطر ابتلادرگروه سیگاری}}{\text{خطر ابتلا گروه غیرسیگاری}} = \frac{0.4}{0.1} = 4$$

این عدد(RR= ۴) نشان می‌دهد که در طول ۱۰ سال افراد سیگاری ۴ برابر بیشتر از افراد غیرسیگاری به سرطان ریه مبتلا می‌شوند. در واقع این عدد (۴) شدت ارتباط بین مصرف سیگار و سرطان ریه را نشان می‌دهد. اما اگر عوامل مخدوش کننده وجود داشته باشند، شدت بدست آمده را نمی‌توان به راحتی به اثر سیگار به تنهایی نسبت داد.

اگر در این مثال مشخص شود که ۲۰ درصد از افراد گروه سیگاری، تریاک مصرف می‌کنند، در حالی که درگروه غیرسیگاری، فقط ۲ درصد، تریاک مصرف می‌کنند، آیا صرفاً به دلیل عدم توازن مصرف تریاک بین دوگروه، می‌توان اظهار کرد تریاک یک عامل مخدوش کننده است؟ ممکن است افراد دو گروه از نظر بسیاری از متغیرهای دیگر مثل وزن، جنس و غیره با یکدیگر تفاوت داشته باشند، آیا همه این متغیرها مخدوش کننده خواهند بود؟ پاسخ بسیار ساده است، با این اطلاعات به تنهایی نمی‌توانیم مصرف تریاک و یا سایر متغیرهایی که در دو گروه متفاوت هستند را مخدوش کننده بنامیم.

حال اطلاعات بیشتری بدست می‌آوریم. فرض کنید در تحقیق دیگری خطر نسبی ابتلا به سرطان ریه ناشی از مصرف تریاک ۶ بدست آمده باشد؛ یعنی افرادی که تریاک مصرف می‌کنند ۶ برابر افرادی که تریاک استفاده نمی‌کنند به سرطان ریه مبتلا می‌شوند. حال با این توضیحات جمع بندی چیست؟

در مثال بالا مشاهده کردیدکه افراد مصرف کننده تریاک در گروه سیگاری‌ها بیشتر هستند (۲۰٪ در مقابل ۲٪)؛ از آنجایی که تریاک خود یک عامل خطر سرطان ریه است (RR=۶)، پس می‌توان نتیجه گرفت که RR=۴ برای سیگار ترکیبی از اثر مصرف سیگار و تریاک بوده است؛ یعنی خطر نسبی ۴ صرفاً به دلیل مصرف سیگار نیست، بلکه نتیجه مصرف توأم سیگار و تریاک بیشتر در این گروه است. بنابراین می‌توان نتیجه گرفت اگر مصرف تریاک در گروه سیگاری‌ها مساوی غیرسیگاری‌ها شود، خطر نسبی حاصل از مصرف سیگار کمتر از ۴ خواهد شد.

با یک نگاه اجمالی به شرط‌های لازم برای این که یک عامل مخدوش کننده باشد، می‌توان دیدکه مصرف تریاک یک عامل مخدوش کننده است؛ تریاک عامل خطر سرطان ریه است (شرط اول)، از طرفی بین مصرف سیگار و تریاک رابطه وجود دارد؛ یعنی سیگاری‌ها بیشتر تریاک مصرف می‌کنند (شرط دوم)، و بالأخره مصرف تریاک یک عامل واسطه‌ای نیست (شرط سوم).

گاهی اوقات برخی از متغیرها به نظر مخدوش کننده می‌رسند ولی در واقع این طور نیستند؛ آیا می‌توانید ادعا کنید که در رابطه بین مصرف سیگار و ابتلا به بیماری‌های قلبی و عروقی چاقی، عامل مخدوش کننده است؟ مطالعات نشان داده که چاقی یک عامل خطر برای ابتلا به بیماری‌های قلبی و عروقی است (شرط اول). ولی در وجود رابطه بین چاقی با سیگار کشیدن تردید وجود دارد (عدم اطمینان از شرط دوم)؛ در چنین حالتی به عامل چاقی مخدوش کننده بالقوه (Potentially Confounder) گفته می‌شود به بیان بهتر عامل مخدوش کننده بالقوه، متغیری است که می‌تواند در یک تحقیق مخدوش کننده باشد و در تحقیق دیگری خیر.

از طرفی گاهی اوقات عواملی وجود دارند که مستقیماً با پیامد رابطه ندارند، ولی با واسطه چندین متغیر بر روی پیامد اثر می‌گذارند و می‌توانند نقش مخدوش کننده داشته باشند. به این متغیرها مخدوش کننده واسطه‌ای یا غیرمستقیم (Proxy Confounder) می‌گویند. معمولاً متغیرهایی مثل سن، جنس و وضعیت اقتصادی ـ اجتماعی از این دسته هستند.

[Downloaded from payeshjournal.ir on 2026-10-30]
[DOR: 20.1001.1.16807626.1389.9.4.13.7]

روش‌های حذف متغیر مخدوش کننده

برای حذف و کنترل اثر مخدوش کنندگی یا باید در مرحله طراحی مطالعه تمهیداتی به خرج داد و یا در زمان تجزیه و تحلیل نتایج. برای از بین بردن عامل مخدوش کننده در مرحله طراحی مطالعه سه شیوه وجود دارد:

۱- تصادفی سازی (Randomization)، فقط در مطالعات مداخله‌ای قابل انجام است و منظور از آن وارد نمودن نمونه‌ها به صورت تصادفی در زیرگروه‌ها است، در این روش برای یکنواخت کردن توزیع متغیرهای مخدوش کننده در گروه‌های مورد مطالعه از شانس و تصادف کمک گرفته می‌شود، چراکه انتظار داریم در صورت توزیع تصادفی نمونه‌ها در گروه‌های مورد مطالعه، توزیع متغیرهای مخدوش کننده به تدریج در گروه‌ها به حالت تعادل برسد.

۲- محدود کردن (Limitation)، یعنی انتخاب نمونه‌ها فقط از یک زیرگروه متغیر مخدوش کننده. در مثال قبلی در مورد رابطه بین مصرف سیگار و سرطان ریه، اگر نمونه‌ها فقط از افراد غیرتریاکی گرفته شوند و نمونه‌های مصرف کننده تریاک از هر دو گروه حذف شوند، طبیعی است که جایی برای تأثیرگذاری تریاک باقی نمی‌ماند.

۳- همسان سازی (Matching)، یعنی انتخاب افراد به گونه‌ای که توزیع عامل مخدوش کننده در زیر گروه‌ها یکسان باشد. در مثال قبل برای این که عامل مخدوش کنندگی مصرف تریاک از بین برود می‌توان افراد را به گونه‌ای انتخاب کرد که مصرف تریاک در دو گروه سیگاری و غیرسیگاری یکسان باشد.

برای از بین بردن اثر عوامل مخدوش کننده در زمان تجزیه و تحلیل باید از روش‌های آماری خاص مانند مثل تجزیه و تحلیل طبقه بندی شده (Stratified) و یا مدلسازی استفاده کرد. البته باید توجه داشت به شرطی می‌توان در زمان تجزیه و تحلیل اثرات متغیرهای مخدوش کننده را از بین برد که در حین مطالعه عوامل مخدوش کننده اندازه‌گیری شده باشد؛ به طور مثال اگر جنس، سن و یا سطح تحصیلات یک عامل مخدوش کننده هستند، در حین جمع‌آوری اطلاعات باید این متغیرها سنجیده و ثبت شده باشند.

نتیجه گیری

همانگونه که بیان شد در تمام تحقیقات، اندازه‌گیری‌ها بدلیل محدودیت‌های موجود در وسایل اندازه‌گیری، محیط اندازه‌گیری، فرایند اندازه‌گیری و همچنین خطای انسانی تقریبی بوده و احتمالاً مقادیر اندازه‌گیری شده با مقادیر واقعی کاملاً مطابقت ندارند. اما شناخت این خطاها، بررسی اثر احتمالی و تلاش برای به حداقل رساندن آنها جزو وظایف یک محقق است. بی تردید توجه به این مهم در تحقیقات علوم پزشکی که با جوامع انسانی در ارتباط بوده و هدف از انجام طرح‌های پژوهشی در این حیطه تعمیم نتایج حاصل ازآنها به جوامع بشری است، از اهمیت خاصی برخوردار است؛ چراکه ارایه نتایج اشتباه باعث بوجود آمدن پیامدهای گاه جبران ناپذیری خواهد شد.

در این مقاله سعی بر این بود تا به زبانی ساده و روان و به دور از پیچیدگی‌های تخصصی، اهم خطاهای اندازه‌گیری شرح داده شوند. دانستن و شناخت این خطاها و راه‌های کاهش اثرات آنها اولین قدم برای انجام پژوهش‌های درست و دقیق است. طبیعی است که مصادیق و مفاهیم این خطاها در انواع تحقیق و زمینه‌های مختلف متفاوت است.

سهم نویسندگان

معصومه صادقی: بررسی متون و تدوین متن اولیه مقاله و تحلیل داده‌ها

مریم نصیریان: بررسی متون و تدوین متن اولیه مقاله و تحلیل داده‌ها

علی‌اکبر حق‌دوست: تدوین ابعاد موضوع، ویرایش مطالب جمع‌آوری شده و نهایی نمودن و پی‌گیری مراحل انتشار مقاله

منابع

۳- محمد کاظم، ملک افضلی حسین، نهاپتیان وارتکس. *روشهای آماری و شاخصهای بهداشتی*. چاپ دوازدهم، سلمان، تهران ۱۳۸۳

1- Knapp RG, Miller Clinton MD. *Clinical epidemiology and biostatistics*. 3rd Edition, Williams & Wilkins: United States of America, 2006

2- Levy PS, Lemeshow S. *Sampling of population: Methods and Applications*. 3rd Edition, Wiley-Interscience: United States of America, 2002

4- Sacket DL. Bias in analytical research. *Journal of Chronic Disease* 1979; 32: 51-63

[Downloaded from payeshjournal.ir on 2026-10-30]

[DOR: 20.1001.1.16807626.1389.9.4.13.7]

5- Fosgate GT. Non-differential measurement error does not always bias diagnostic likelihood ratios towards the null. *Emerging Themes in Epidemiology* 2006; 3: 7-10

۶- سرافراز علی اکبر، غفارزادگان کامران. *آمار پزشکی (پایه/بالینی)*. چاپ سوم، دانشگاه علوم پزشکی مشهد، تهران۱۳۷۶

7- Streiner DL, Norman GR. Precision and accuracy: two terms that is neither. *Journal of Clinical Epidemiology* 2006; 59: 327-30

8- Zklo M, Nieto J. E*pidemiology Beyond the Basics*. 2nd Edition, Aspen Publisher: United States of America, 2003

9- Rothman K J, Greenland S, Lash TL. *Modern Epidemiology*. 3 rd Edition, Lippincott Williams & Wilkins: Philadelphia, 2008

۱۰- شجاعی تهرانی حسین. *فرهنگ اپیدمیولوژی*. چاپ دوم، سماط، تهران ۱۳۸۲

۱۱- ملک افضلی حسین، مجدزاده رضا، فتوحی اکبر، توکلی سامان. *روش شناسی پژوهشهای کاربردی در علوم پزشکی*، چاپ اول، دانشگاه علوم پزشکی وخدمات بهداشتی ـ درمانی تهران، تهران ۱۳۸۳

12- Gordis L. *Epidemiology*. 3rd Edition, Saunders, an Imprint of Elsevier Inc: Philadelphia, 2006

[Downloaded from payeshjournal.ir on 2026-10-30]

[DOR: 20.1001.1.16807626.1389.9.4.13.7]