

## دستورالعمل‌های آماری برای مقالات علوم پزشکی

نویسندگان: دکتر مسعود رودباری<sup>۱\*</sup> و دکتر سیدحسن صانعی<sup>۲</sup>

۱. استادیار آمار زیستی دانشکده بهداشت دانشگاه علوم پزشکی زاهدان

۲. دکتری آمار حیاتی

Email: mroudbari@yahoo.co.uk

\* نویسنده مسئول:

### چکیده

سابقه: رعایت دستورالعمل‌های نگارش مقالات پزشکی خصوصاً دستورالعمل‌های مربوط به قوانین آماری، از جمله موضوعاتی است که محققین علوم پزشکی و هیأت تحریریه مجلات پزشکی غالباً با آن سروکار دارند. بسیاری از مقالات به دلیل عدم رعایت این اصول، امکان چاپ پیدا نکرده، یا به دلیل لزوم بازنگری عودت داده می‌شوند و با تأخیر زمانی طولانی چاپ می‌شوند. از سوی دیگر، دلایلی چون دغدغه‌های متعدد محققین علوم پزشکی، ماهیت دشوار علم آمار، عدم تجربه، و گذشت زمان طولانی از دوره آموزش آنان، موجب می‌شوند که اشتباهات فاحش آماری در اغلب مقالات علمی به چشم بخورد.

هدف: این مقاله ضمن بررسی نیازهای آماری که قسمت‌های مختلف مقالات پزشکی به آن‌ها احتیاج دارد، به ارائه دستورالعمل‌های آماری پرداخته است تا با رعایت آن‌ها از سوی نگارندگان، امکان چاپ مقالاتشان افزایش یابد.

روش بررسی: این دستورالعمل‌ها می‌تواند به نیازهای آماری بسیاری از محققین پاسخ واضح و روشن داده، آن‌ها را از مراجعه مستقیم به کتب مرجع آماری بی‌نیاز سازد.

یافته‌ها: دستورالعمل‌های ارائه شده در این مقاله شامل نکاتی است که در قسمت‌های روش‌کار، تجزیه و تحلیل و تفسیر داده‌ها کاربرد دارد، به طوری که محققین را تا حدودی از مشاوری آماری بی‌نیاز می‌سازد. غالب این دستورالعمل‌ها به قسمت تجزیه و تحلیل داده‌ها مربوط بوده، شامل روش‌های تجزیه و تحلیل با توجه به انواع مشاهدات و روش‌های مختلف مطالعه و همچنین اشتباهات رایجی است که محققین در تشخیص روش‌های مختلف آماری با آن مواجه هستند. نتیجه‌گیری و بحث: امید است ارائه این دستورالعمل‌ها بتواند گامی کوچک، اما مؤثر در بالا بردن کیفیت مقالات پزشکی در مجلات علمی برداشته، محققین را در ارائه هر چه بهتر و علمی‌تر تحقیقاتشان توانمند سازد.

واژه‌های کلیدی: دستورالعمل، آمار، مقالات، پزشکی

دوماهنامه علمی - پژوهشی

دانشگاه شاهد

سال سیزدهم - شماره ۶۳

تیر ۱۳۸۵

تاریخ وصول: ۸۳/۱۰/۲۰

تاریخ پذیرش: ۸۴/۴/۲۸

### مقدمه

با پیشرفت علم و تحقیق، همه روزه شاهد انتشار مقالات علمی، خصوصاً مقالات علمی پزشکی در این گونه مجلات هستیم که نتیجه و ماحصل تحقیقات محققین این رشته از علوم است. اکثر مجلات علمی برای چاپ مقالات از الگو و ساختار خاصی پیروی می‌کنند که این الگو به تصویب شورای تحریریه و

سردبیری مجله مورد نظر رسیده، با چاپ آن در ابتدا یا انتهای مجله از صاحبان مقالات خواسته می‌شود که در نوشتن مقالات خویش این الگو را مدنظر داشته، رعایت کنند. بنابراین از چاپ مقالاتی که این الگو را رعایت نکرده باشند معذورند و ممکن است آن را برای تصحیح به صاحبان اثر عودت دهند.

بعضی از دستورالعمل‌ها برای برخی از محققین که در آمار تجربه کافی ندارند، قاعداً ناآشنا به نظر می‌رسند و فقط برای پوشش دادن به مواردی که برخی از محققین به آن‌ها نیاز دارند، آورده می‌شوند تا این‌گونه محققین بتوانند با استفاده از کتب آماری [۲، ۳، ۴، ۵، ۶] و یا بهره‌گیری از مساعدت یک آمارشناس مجرب استفاده کنند. در مورد برخی دیگر از عناوین و سرفصل‌های آماری، دستورالعملی ارائه نمی‌شود و در این‌گونه موارد مراجعه به آمارشناسان خبره توصیه می‌شود.

هدف این دستورالعمل‌ها، یادآوری محققین به مواردی است که از نظر علم آمار مهم بوده، معمولاً محققین در چگونگی به‌کارگیری آن‌ها در مقالاتشان با مشکل مواجهند. البته موارد ارائه شده در این مقاله، کلیه مواردی را که محققین باید از علم آمار بدانند شامل نمی‌شود، بلکه تنها مشتمل بر مواردی است که می‌توان در یک مقاله به آن‌ها اشاره کرد. شیوه تعیین مناسبت روش‌های آماری و نیز چگونگی انتخاب آن‌ها موضوعی فراتر از یک مقاله است و حتماً باید با مشورت یک آمارشناس صورت‌پذیرد. در پاره‌ای از اوقات، افراد - آگاهانه یا ناآگاهانه - از روش‌های آماری در جهت اهداف پروژه خود سوء استفاده می‌کنند که باید به این موارد نیز توجه خاص معطوف کرد.

دستورالعمل‌های ارائه شده در این مقاله به ترتیب شامل مواردی است که در بخش‌های روش‌ها، تجزیه و تحلیل، و بحث مقالات علمی به‌کار می‌روند. برخی از دستورالعمل‌ها به علت کاربرد گسترده در بخش‌های مختلف مقالات علمی ارائه می‌شوند که در این صورت باید علت آن نیز ذکر گردد.

از آن‌جا که بعد از انجام یک تحقیق، امکان تصحیح خطا در طرح آن پروژه وجود ندارد، یادآوری این موضوع بسیار حائز اهمیت است که قبل از شروع تحقیق و در زمان نوشتن پروپوزال و برنامه‌ریزی اجرایی آن (نه در زمان تجزیه و تحلیل اطلاعات)، لازم است با یک آمارشناس با تجربه مشاوره انجام شود.

از جمله بخش‌های بسیار مهم و کلیدی در هر مقاله علمی که در اکثر مقالات پزشکی وجود دارد، بخش‌های «روش کار»، و «تجزیه و تحلیل داده‌ها» است که به دلیل تخصصی بودن و این‌که اکثر تجزیه و تحلیل‌های این مقالات به کمک علم آمار انجام می‌شود، راهنمایی و مشورت یک آمارشناس را می‌طلبد و همین موضوع، انگیزه نگارش این مقاله شده است [۱]. این در حالی است که بسیاری از محققین و ارسال‌کنندگان این مقالات، علی‌رغم آشنایی با آمار مقدماتی، اصول مهم آماری را نمی‌دانند [۲]. لذا بسیاری از مطالبی که در بخش‌های تجزیه و تحلیل مقالات علمی ارائه می‌شود ضعیف و یا نادرست هستند و به علت عدم وجود راهنمایی‌های جامع و کامل آماری از سوی مجلات پزشکی، مقالات نادرست برای اصلاحات به محققین عودت می‌شوند که این امر خود علاوه بر طولانی شدن زمان بررسی و یا عدم چاپ این مقالات، موجب می‌گردد مقالات ارسالی به‌روز نباشند و مشمول مرور زمان گردند.

طبیعی است مقالات دارای اشتباهات واضح آماری، قابل استفاده و چاپ نبوده، نتایج و تفاسیر حاصل از آن‌ها غلط و غیرقابل استفاده است. هرچند در پاره‌ای از موارد، تفسیرهایی که از تجزیه و تحلیل‌های درست و صحیح آماری صورت می‌گیرد نیز نادرست بوده، غیرقابل استفاده‌اند که این‌گونه اشتباهات در پاره‌ای از مقالات به چشم می‌خورد.

کاربرد توأم آمار با پزشکی در مجلات پزشکی نیز از جمله مواردی است که هنر خاصی را می‌طلبد و فقط شمار اندکی از مقالات قادر به ارائه صحیح و توأم این دو بوده‌اند؛ زیرا تلفیق صحیح این دو مقوله نیازمند فنون و دانش خاص است، به‌طوری که خوانندگان مقالات مزبور، تلفیق صحیح این دو علم را به خوبی درک خواهند کرد.

در این مقاله سعی شده دستورالعمل‌های متداول و لازم آماری برای نوشتن مقالات پزشکی به‌صورتی آسان و قابل فهم ارائه گردند و در اختیار محققین محترم قرار گیرند تا بتوانند آن‌ها را در مقالات خویش به‌کار گیرند.

### کارآزمایی‌های بالینی

در کارآزمایی‌های بالینی باید رژیم‌های درمانی، از جمله مراقبت ویژه از بیمار، ضوابط تغییر یا توقف درمان و همچنین اختصاص درمان به بیماران به تفصیل بیان شود. روش انتخاب تصادفی بیماران و چگونگی انجام این کار نیز نیاز به تشریح و توجیه کافی دارد. از آن جا که انتخاب غیر تصادفی افراد در مطالعه به عنوان یک نقص در مطالعه تلقی می‌گردد باید به همراه دلایل آن ذکر شود. مطالعاتی که در آن از روش‌های انتخاب غیر تصادفی استفاده می‌شود (مثل استفاده از نمونه‌های بیمارستانی) امکان چاپ در مجلات معتبر علمی را ندارند.

علاوه بر این، روش‌های کورسازی مورد استفاده و دقت‌هایی که برای اطمینان از صحیح بودن پاسخ بیماران به کار می‌روند، باید تشریح و معیارهای مقایسه روش‌های درمانی موجود در پروپوزال طرح فهرست شوند [۸].

### روش‌های آماری

از جمله مواردی که باید از توضیح کافی برخوردار باشد، روش‌های آماری به کار برده شده است که باید در قسمت روش کار به آن پرداخت. اگر روش‌های آماری متفاوتی مورد استفاده قرار گرفته ضروری است به وضوح توضیح داده شود که کدام روش، در کجا مورد استفاده قرار گرفته است. برخی روش‌های شناخته شده متداول آماری نظیر آزمون «تی» و آزمون کای دو، آزمون‌های ویلکاکسون و من‌ویتنی، ضریب همبستگی پیرسون و رگرسیون خطی نیازی به توضیح ندارند، اما روش‌های آماری که بیش از یک شکل دارند، مثل آزمون «تی» (زوجی و غیرزوجی)، تجزیه و تحلیل واریانس و همبستگی رتبه‌ای [۹] باید دقیقاً توضیح داده شوند. همچنین روش‌های پیچیده آماری نیاز به توضیح دارند و چنانچه روش‌های متداولی نیستند که همگان با آن مأنوس باشند باید همراه با یک مرجع و مأخذ دقیق و ترجیحاً از یک کتاب درسی با ذکر شماره صفحه آن معرفی گردند. توضیح مختصر درباره علت به کارگیری یک روش خاص، وقتی که از روش مشهور و متداول‌تر اجتناب شده است، سودمند

### دستورالعمل‌های آماری مربوط به بخش روش‌ها

طراحی مطالعه (تجربی، کارآزمایی بالینی، مشاهده‌ای و...) و نحوه جمع‌آوری داده‌ها از جمله مواردی است که ذکر آن در بخش روش‌ها ضروری است. باید اطلاعات کافی و جامعی درباره طراحی مطالعه و امکان تکرار آن توسط سایر محققین در این قسمت ارائه گردد. همچنین باید روش‌های آماری استفاده شده با جزئیات کامل در این قسمت توضیح داده شوند؛ زیرا همه مخاطبین و خوانندگان مقاله ارائه شده آگاهی و دانش آماری لازم را ندارند. بنابراین، چنین موضوعاتی باید به گونه‌ای شیوا، رسا، جامع و بدون ابهام بیان شوند. سایر مواردی که باید در این قسمت ذکر گردند عبارتند از:

- الف) اهداف مطالعه و فرضیات عمده تحقیق،
  - ب) نوع نمونه و بیان ضوابط ورود و خروج از مطالعه،
  - ج) جامعه آماری و نحوه انتخاب اعضای نمونه،
  - د) تعداد نمونه و شیوه محاسبه حجم نمونه انتخاب شده.
  - ه) نوع متغیرها و روش اندازه‌گیری آن‌ها.
- توجه به این نکته ضروری است که مطالعات مختلف ممکن است نیازمند اطلاعات خاص‌تر باشند که در زیر به بیان پاره‌ای از آن‌ها خواهیم پرداخت.

### مطالعات مشاهده‌ای

در این نوع مطالعات، طراحی مطالعه باید به وضوح تشریح شود. برای مثال اگر نوع مطالعه، مورد-شاهدی است نحوه انتخاب گروه مورد و فرایند همسان‌سازی اعضای آن باید به تفصیل توضیح داده شود. همچنین دلیل استفاده از مطالعه مورد-شاهدی، همگروهی و یا مقطعی باید به وضوح بیان گردد. روش انتخاب افراد نمونه و دستیابی به میزان بالای مشارکت در مطالعه از سوی افراد نمونه، موضوع بسیار مهم دیگری است که باید مد نظر قرار گیرد. همچنین کلیه تصمیمات و اقدامات مفید دیگری که برای تشویق و ترغیب شرکت‌کنندگان به منظور مشارکت بیش‌تر آنان در تحقیق صورت گرفته باید بیان گردد [۷].

خواهد بود. بیان نام نرم‌افزار آماری و برنامه استفاده شده به همراه روش آماری مورد استفاده الزامی است.

### دستورالعمل‌های آماری مربوط به بخش نتایج (تجزیه و تحلیل آماری)

#### مطالعه توصیفی

قبل از شروع تجزیه و تحلیل، ارائه توضیحاتی درباره متغیرهای مورد بررسی و ماهیت داده‌ها ضروری به نظر می‌رسد. متغیرهای عمومی (زمینه‌ای) که برای اعتبار و تفسیر تجزیه و تحلیل در طرح وارد می‌شوند باید به تفصیل شرح داده شوند. بدین منظور، استفاده از روش‌های نموداری همانند دیاگرام پراکنش یا هیستوگرام و نیز شاخص‌های آماری [۶] به منظور نشان دادن همبستگی نمونه مورد بررسی با جامعه کل (اعتباربخشی) مفید خواهد بود. در متغیرهای پیوسته، مثل وزن و فشار خون می‌توان با استفاده از میانگین، انحراف معیار و یا میانه و دامنه صدکی مثل فاصله بین صدک‌های ۲۵ و ۷۵ به خلاصه کردن داده‌ها پرداخت. در مواقعی که توزیع متغیر پیوسته نامتقارن باشد، روش استفاده از میانه و دامنه صدکی توصیه می‌شود. در چنین مواردی برای تشریح تغییرپذیری، استفاده از خطای معیار نامناسب است. برای اطلاعات کیفی رتبه‌ای، مثل مراحل بیماری، استفاده از میانگین و انحراف معیار صحیح نبوده باید از نسبت‌ها استفاده شود [۷].

در کارآزمایی‌های بالینی، ذکر تعداد حذف‌شدگان، دلایل حذف آن‌ها و همچنین نحوه انتساب افراد به گروه‌های درمانی باید مشخص شود. میزان پاسخ و خصوصیات افرادی که به درمان پاسخ نداده‌اند به همراه مقایسه با افرادی که در مطالعه شرکت کرده‌اند نیز برای مطالعات مشاهده‌ای ضروری است. همچنین لازم است نمونه از جامعه‌ای که باید نتایج به آن تعمیم داده شود، آمده باشد [۸].

مقایسه توزیع ویژگی‌های اصلی در گروه‌های مختلف مورد بررسی در کارآزمایی‌های تصادفی شده، مثل گروه درمان، مفید خواهد بود. تفاوت‌های موجود در بین گروه‌ها، حتی اگر از نظر آماری معنادار نباشد،

واقعی است و باید به صورت صحیح و مناسب در تجزیه و تحلیل‌ها لحاظ گردد. به همین علت، آزمون‌هایی که بدین منظور صورت می‌گیرند، فقط صحت تصادفی بودن را بررسی می‌کنند و قادر به نشان دادن قابل مقایسه بودن گروه‌ها نیستند.

#### فرضیات اصلی

بسیاری از تجزیه و تحلیل‌های آماری، نظیر آزمون‌های «تی» (یک نمونه‌ای، دو نمونه‌ای، و زوجی)، همبستگی و تجزیه و تحلیل واریانس که به روش‌های پارامتری موسومند، همگی به مفروضات معینی درباره توزیع متغیر(های) مورد بررسی وابسته‌اند. غالباً این مفروضات، شامل نرمال بودن داده‌ها و در صورت وجود دو و یا چند گروه، مساوی بودن واریانس (یکسان بودن تغییرات) در گروه‌ها هستند [۱۰].

در داده‌های دارای توزیع بسیار نامتقارن (چوله) و یا با تفاوت زیاد تغییرات در بین گروه‌ها می‌توان قبل از شروع تجزیه و تحلیل به انجام یک تبدیل (به بخش تبدیل اطلاعات مراجعه شود) مبادرت کرد تا بتوان با استفاده از سایر روش‌های آماری (توزیع آزاد یا تجزیه و تحلیل ناپارامتری) که به توزیع متغیرها وابسته نیست، کمک گرفت. برای مثال آزمون من یو ویتنی توزیعی ناپارامتری، معادل آزمون «تی» دو نمونه‌ای است. روش‌های ناپارامتری می‌توانند برای مجموعه‌های کوچک داده‌ها که فرضیات لازم برای آن‌ها برقرار نیست، به کار برده شوند [۹].

همان‌گونه که بیان گردید، در بسیاری از موارد نرمال بودن توزیع متغیر مورد بررسی از اهمیت ویژه‌ای برخوردار است. برای مثال، هنگامی که اطلاعات به صورت دو انحراف معیار در طرفین میانگین و با عنوان یک دامنه نرمال ۹۵ درصد و یا فاصله مرجع قرار گرفته‌اند، این اهمیت مشهود و محسوس است. در چنین مواردی، توضیح لازم و کافی برای پذیرفتن یا در نظر گرفتن این مفروضات توزیعی باید ارائه گردد.

## آزمون فرضیه

هدف اصلی آزمون‌های فرضیه که غالباً از آن‌ها به‌عنوان آزمون‌های معناداری نیز یاد می‌شود، آزمودن تعدادی فرضیه است که از پیش در نظر گرفته شده‌اند. باید به خاطر سپرد آزمون‌هایی که بعد از بررسی مقدماتی داده‌ها پیشنهاد می‌گردند، معمولاً نتایج کاذبی را به ارمغان خواهند آورد؛ زیرا در چنین مواقعی مقدار  $p$  محاسبه شده، بسیار کوچک است. برای مثال، آزمون تفاضل بین بزرگ‌ترین و کوچک‌ترین مقدار از مجموعه چند میانگین (یا نسبت)، بدون ارائه دلیل خاص، معتبر نیست. روش‌های خاصی، نظیر مقایسه چندگانه، برای مقایسه دوتایی بین چند گروه وجود دارد. چنانچه در مواردی سه گروه یا بیش‌تر که دارای ترتیب طبیعی هستند، مورد مقایسه قرار گیرند (مثل گروه‌های سنی، مراحل مختلف سرطان‌ها و...) باید با روش‌های خاص تجزیه و تحلیل شوند تا بتوان روند بین گروه‌ها را مورد ارزیابی قرار داد [۱۱].

در آزمون‌ها، معمولاً استفاده از آزمون دو دامنه مورد تأکید و توصیه است و چنانچه قرار است آزمون یک دامنه‌ای انجام‌پذیرد، محقق باید توجیه کافی، قوی و مستدلی برای به‌کارگیری چنین آزمونی داشته باشد [۶]. ارائه نتایج آزمون فرضیه به همراه فاصله اطمینان، مورد توصیه اکید نگارندگان است [۶].

## فواصل اطمینان

بسیاری از مطالعات با برآورد کمیت‌هایی چون میانگین‌ها، تفاضل‌ها یا خطر نسبی سروکار دارند. محاسبه فاصله اطمینان برای چنین برآوردهایی مطلوب پژوهشگران و محققین است. فاصله اطمینان، دامنه‌ای از مقادیر حول برآورد متغیر مورد نظر است که با ۹۵ درصد احتمال، مقدار واقعی متغیر را شامل می‌شود. به عبارت دیگر، چنانچه فاصله اطمینان برای همان متغیر را به وسیله ۱۰۰ بار نمونه‌گیری (۱۰۰ مطالعه تحقیقاتی) محاسبه کنیم، در ۹۵ بار، فاصله اطمینان مقدار واقعی متغیر را شامل می‌شود. رابطه بسیار تنگاتنگی بین نتایج یک آزمون فرضیه و فاصله اطمینان مربوط وجود دارد؛

به‌طوری که اگر تفاوت بین گروه‌ها در سطح معناداری ۰/۰۵ معنادار باشد، آنگاه فاصله اطمینان ۹۵ درصد مربوط، مقدار صفر را دربر نمی‌گیرد. برعکس، اگر تفاوت بین گروه‌ها در سطح ۰/۰۵ معنادار نباشد، مقدار صفر در فاصله اطمینان ۹۵ درصد مورد نظر قرار خواهد گرفت. فاصله اطمینان، اطلاعات زیادی را در اختیار قرار می‌دهد؛ زیرا بر دامنه‌ای از مقادیر واقعی دلالت دارد که با مشاهدات نمونه‌ای سازگار است [۲]. صنایع در فصل مربوط به آزمون‌های آماری، مثال‌های متعدد و گوناگونی را دال بر ارتباط تنگاتنگ این دو روش ارائه داده است [۶].

بدین ترتیب می‌توان گفت که فواصل اطمینان، دقت برآورد را مشخص می‌کنند. بنابراین، فواصل اطمینان پهن و عریض به معنای فقدان اطلاعات کافی هستند، خواه این تفاوت از نظر آماری معنادار باشد یا نباشد. این مطلب برای کسانی که تفسیری پر دامنه از نتایج یک مطالعه کوچک ارائه می‌دهند، هشدار محسوب می‌گردد. در مطالعات مقایسه‌ای که به بررسی تفاضل بین گروه‌ها می‌پردازند، به جای این که فواصل اطمینان جداگانه‌ای برای نتایج هر گروه گزارش شود، باید برای تفاضل دو گروه مورد مقایسه، یک فاصله اطمینان واحد ارائه گردد.

## مشاهدات جفتی

تشخیص حالت مشاهدات غیر جفتی از مشاهدات جفتی ضروری است. برای مثال، مقایسه بین اندازه‌گیری‌های دو گروه مختلف در نمونه‌هایی که درمان‌های متفاوتی را دریافت می‌کنند (غیر جفتی)، از مقایسه بین دو بار اندازه‌گیری در زمان‌های مختلف بر روی افراد یکسان، مثل قبل و بعد از درمان (جفتی)، متفاوت است؛ زیرا هم آزمون مورد استفاده و هم روش محاسبه فاصله اطمینان برای این دو نوع مشاهدات تفاوت دارد. برای مثال، داده‌ها غیر جفتی، آزمون «تی» دو نمونه‌ای را می‌طلبد، در حالی که برای داده‌های جفتی از آزمون «تی» جفتی استفاده می‌شود [۸]. بنابراین باید همواره به طریقی مشخص شود که در کدام مطالعه کدام نوع آزمون (جفتی یا غیر جفتی) استفاده شده است.

مختلف است. اغلب به وسیله محاسبه میانگین‌ها و انحراف معیار چنین داده‌هایی در زمان اندازه‌گیری تجزیه و تحلیل می‌شوند و به وسیله نموداری خطی که این میانگین‌ها را به هم وصل می‌کند، نشان داده می‌شود. شکل این منحنی میانگینی ممکن است ایده خوبی درباره شکل منحنی‌های افراد ارائه نکند، مگر آن‌که اندازه‌های افراد خیلی شبیه به هم باشد. تجزیه و تحلیل برخی از ویژگی‌های افراد، نظیر زمان دستیابی به پیک (بالاترین مقدار) یا مدت زمان لازم برای گذشتن از یک سطح مشخص (مدت زمان رسیدن یک فاکتور در بدن به بالای مقداری خاص)، با ارزش‌تر خواهد بود. این مطلب همچنین کمک می‌کند تا از مشکلات مربوط به آزمون فرضیه‌های چندگانه اجتناب شود (به قسمت مربوط به کثرت آزمون فرضیه مراجعه شود).

اندازه‌گیری‌های تکراری متغیر مشابه بر روی یک فرد تحت همان شرایط آزمایشگاهی که از آن به‌عنوان «خواندن‌های تکراری» یاد می‌شود، در هنگام مقایسه گروه‌هایی از افراد نباید به‌عنوان مشاهدات مستقل در نظر گرفته شود. در جایی که تعداد دفعات تکرار برای همه افراد مشابه است، تجزیه و تحلیل مشکل نیست، به ویژه تجزیه و تحلیل واریانس در جایی که آزمون‌های «تی» بر داده‌های غیرتکراری کاربرد دارد، به کار می‌رود. چنانچه تعداد دفعات تکرار در بین افراد متفاوت باشد، تجزیه و تحلیل کامل ممکن است خیلی پیچیده باشد. اگر تعداد مشاهدات بین افراد به‌طور چشمگیر متفاوت باشد، استفاده از بزرگ‌ترین و کوچک‌ترین عضو یک سری از اندازه‌گیری‌ها (مثل فشارخون ماکزیمم در طی حاملگی) می‌تواند گمراه‌کننده باشد [۲].

### تبدیل داده‌ها

بسیاری از متغیرهای پزشکی دارای توزیع‌های غیرمتمارن و مخصوصاً چوله مثبت، همراه با برخی از مقادیر بسیار بزرگ هستند و احتمالاً به‌منظور ایجاد داده‌های مناسب برای تجزیه و تحلیل نیاز به تبدیل ریاضی دارند. در چنین مواقعی، استفاده از تبدیل‌ها نظیر تبدیل لگاریتمی، تبدیل جذر، و تبدیل معکوس مفید است که غالباً تبدیل لگاریتمی ممکن است مناسب‌تر باشند [۱۰].

همان تمایز باید زمانی که سه یا چند مجموعه از داده‌ها مورد مقایسه قرار می‌گیرند، وجود داشته باشد. تمام روش‌های آماری بیان شده در این بخش می‌توانند به بیش از دو گروه نیز تعمیم یابند. البته به‌ویژه آزمون «تی» دونه‌ای و جفتی به شکل‌های مختلف تجزیه و تحلیل واریانس قابل تعمیم هستند.

### واحدهای تجزیه و تحلیل

واحد تجزیه و تحلیل، بستگی به هدف، شرایط، فرضیات و موقعیت مطالعه تحقیقاتی دارد. برای مثال، بسیاری اوقات ممکن است چندین اندازه‌گیری روی یک بیمار خاص انجام شود، اما بیمار همچنان به‌عنوان کانون توجه تحقیق باقی بماند. ساده‌ترین مورد آن است که محقق می‌خواهد بخشی از آناتومی انسانی را که به‌صورت جفتی آفریده شده مورد مطالعه قرار دهد. برای مثال چشم، گوش، دست و پا از مواردی هستند که برخی اوقات می‌توان اندازه‌گیری‌های زیادی از آن گرفت. در چنین مواردی، شمارش متعدد بیماران خاص (یعنی هر اندازه را برای یک بیمار حساب کردن) می‌تواند به‌طور جدی به نتایج نامعقول منتهی شود. به ویژه این مسأله باعث بزرگ شدن حجم نمونه شده، ممکن است به دلیل افزایش توان آزمون، منجر به معنادار شدن آماری غیرواقعی و مصنوعی گردد. بنابراین، بیمار، واحد بررسی است و بنابراین باید واحد تجزیه و تحلیل نیز همان بیمار باشد.

به عکس، گاه گروه‌ها، کانون توجه هستند. مثلاً گروه‌هایی از کارآزمایی‌های تصادفی (خوشه‌ای)، نظیر بخش‌های بیمارستان یا مطب‌های عمومی ممکن است به‌صورت تصادفی به زیرگروه‌هایی تقسیم شده، بر روی هر زیرگروه مداخله‌های مختلف صورت گیرد. در این‌گونه مطالعات، تجزیه و تحلیل داده‌ها برای بیماران خاص به‌صورت مشاهدات مستقل اشتباه است، بلکه واحد صحیح تجزیه و تحلیل باید «خوشه» باشد.

### اندازه‌گیری‌های تکراری

یک طرح تحقیقاتی مستلزم ثبت اندازه‌گیری‌های متوالی یک یا چند متغیر در افراد مشابه در فواصل زمانی

### همبستگی

استفاده از نمودار پراکنش داده‌ها برای کلیه ضرایب همبستگی ارائه شده در یک گزارش توصیه می‌گردد؛ هرچند در زمانی که تعداد متغیرهای مورد بررسی زیاد باشد، این کار امکان‌پذیر نیست و مستلزم صرف وقت و فضای بیش‌تر در گزارش است. وقتی که چندین متغیر مورد بررسی قرار می‌گیرند، نشان دادن ضرایب همبستگی بین کلیه متغیرها در یک جدول به‌صورت ماتریس همبستگی مفید خواهد بود و از نشان دادن فقط بزرگ‌ترین مقدار یا مقادیر معنادار آن سودمندتر است.

برای داده‌هایی که نامرتب توزیع شده‌اند، می‌توان ضریب همبستگی رتبه‌ای را به جای پیرسون ضریب همبستگی پیرسون (r) محاسبه کرد. همچنین همبستگی رتبه‌ای می‌تواند برای متغیرهایی که با مقادیر خاص بالا و یا پایین محدود شده‌اند (مثلاً وزن هنگام تولد زیر ۲۵۰۰ گرم) یا برای متغیرهای رتبه‌بندی شده طبقه‌ای، استفاده شود. همچنین در مواقعی که ارتباط بین متغیرها خطی نباشد و یا در شرایطی که مقادیر یک متغیر به جای آن‌که نامحدود باشند، به‌وسیله محقق انتخاب شده‌اند، همبستگی رتبه‌ای اولویت دارد [۶].

ضریب همبستگی، شاخص مفیدی از درجه ارتباط خطی بین دو متغیر کمی است؛ اما در عین حال یکی از روش‌های آماری است که غالباً از آن استفاده نابجا می‌شود. برای مثال، در بسیاری از شرایط نباید از ضریب همبستگی استفاده کرد. محاسبه ضریب همبستگی ساده برای داده‌هایی که شامل بیش از یک مشاهده در همه یا تعدادی از نمونه‌ها می‌شوند، اشتباه است، زیرا چنین مشاهداتی مستقل نیستند. همچنین محاسبه ضریب همبستگی برای مقایسه روش‌های مختلف اندازه‌گیری بر روی متغیر مشابه مناسب نیست، زیرا همبستگی فقط ارتباط خطی - نه توافق بین دو روش - را ارزیابی می‌کند [۲]. استفاده از ضریب همبستگی برای ارتباط دادن تغییرات بر روی زمان به مقدار اولیه، ممکن است نتایج فوق‌العاده گمراه‌کننده‌ای به همراه داشته باشد.

محاسبه ضریب همبستگی برای داده‌هایی که شامل زیرگروه‌های مشخص با سطوح مختلف میانگین در یک

بعد از مرحله تجزیه و تحلیل داده‌ها که باید نتایج به‌صورت گزارش درآیند مطلوب است نتایج به مقیاس اصلی تبدیل و در همان مقیاس برای خوانندگان ارائه شوند. در حالت کلی تبدیل لگاریتمی باید از آنتی‌لگاریتم میانگین لگاریتم داده‌ها - که به «میانگین هندسی» معروف است - استفاده شود. به هیچ وجه نباید از انحراف معیار یا خطای معیار، آنتی‌لگاریتم گرفت، اما به جای آن می‌توان از فاصله اطمینان در مقیاس لگاریتمی، آنتی‌لگاریتم گرفت تا فاصله اطمینان را در مقیاس اصلی نشان دهد. برای سایر تبدیل‌ها، زمانی که یک نمونه تکی موجود است، روش‌های مشابهی اتخاذ خواهد شد، اما برگشت تبدیل برای حدود فاصله اطمینان تفاضل بین میانگین نمونه‌ها، تنها برای تبدیل لگاریتمی معنا خواهد داشت [۲].

در صورت استفاده از یک تبدیل، کنترل این که آیا تغییرات انجام شده، متغیر تبدیل شده را به شرایط مورد نظر، مثل یک توزیع تقریباً نرمال، نزدیک کرده است، حائز اهمیت است. بنابراین همواره نباید چنین فرض کرد که تبدیل لگاریتمی الزاماً راه حل مناسبی برای تبدیل کلیه متغیرهای چوله مثبت است.

### نقاط دور افتاده (پرت)

مشاهداتی که با بدنه اصلی داده‌ها ناسازگاری فاحش دارند نباید از تجزیه و تحلیل حذف شوند، مگر آن‌که دلایل موجه دیگری وجود داشته باشند که به اعتبار آن‌ها خدشه وارد کنند. بنابراین، چنانچه هرکدام از این مشاهدات که «داده‌های پرت و دورافتاده» نامیده می‌شوند، حذف شوند باید در گزارش تحقیق ذکر شوند. از آن جا که نقاط دور افتاده به دلیل وجود اختلاف چشمگیر با بدنه اصلی داده‌ها، می‌توانند اثر قطعی و مشخصی بر تجزیه و تحلیل آماری داشته باشند، لذا به منظور ارزیابی درست مقدار بستگی نتایج با این مقادیر، مطلوب است داده‌های حاوی مشاهدات دورافتاده، یک بار با این مشاهدات و یک بار بدون آن‌ها تجزیه و تحلیل شوند.

### داده‌های بقا

مطلوب است گزارش مطالعات مربوط به داده‌های بقا، شامل نمایش نموداری یا جداولی برای ارائه محاسبات ضروری طول عمر، به همراه جزئیاتی درباره تعداد بیمارانی که در مطالعه تحت پیگیری بوده‌اند، باشد. جدول طول عمر یا نمودار بقا در مورد زمان‌های بقای سانسور (گم) شده به‌طور مؤثر بحث می‌کند که این زمان‌ها ناشی از مواقعی هستند که بیماران گم شده، یا هنوز زنده‌اند، ولی معلوم است که زمان بقای آن‌ها فقط چند روز است. به‌منظور پرهیز از تفسیر نادرست در عدم قطعیت، در قسمت آخر منحنی، بریدن منحنی بقا در زمانی که فقط چند نفر از افراد هنوز در معرض خطر هستند، ممکن است مفید باشد. در صورت وجود داده‌های سانسور شده، محاسبه میانگین زمان بقا معقول نیست، زیرا توزیع زمان‌های بقا معمولاً دارای چولگی مثبت است و لذا محاسبه میانه زمان بقا ارجحیت دارد و توصیه می‌شود [۱۳].

مقایسه نسبت افراد زنده بین گروه‌های درمانی در زمان‌های ثابت اختیاری، ممکن است گمراه‌کننده باشد و معمولاً کارایی کم‌تری در مقایسه با جداول طول عمر با روشی نظیر آزمون لگاریتم رتبه‌ها (Log rank) دارد. روش‌های محاسبه برآوردهای زمان بقا و فواصل اطمینان در [۱] آورده شده‌اند.

زمانی که تعداد کافی از موارد مرگ موجود است، با رسم نسبت کسانی که در شروع هر فاصله زنده‌اند، ولی در طی آن فاصله مرده‌اند (برای فواصل زمانی مساوی) می‌توان نشان داد که چگونه خطر مرگ با زمان تغییر می‌کند. تعدیل عواملی که ممکن است بر پیش‌آگهی بیماری تأثیر بگذارد با استفاده از مدل‌های رگرسیونی مناسب با داده‌های بقا امکان‌پذیر است.

مقایسه بقا بین گروه افرادی که به درمان پاسخ داده‌اند و آن‌هایی که به آن پاسخ نداده‌اند گمراه‌کننده است و هرگز نباید صورت گیرد.

یا هر دو متغیر هستند، مثلاً ترکیب داده‌ها برای مردان و زنان موقعی که یکی از متغیرها «قد» است، می‌تواند گمراه‌کننده باشد.

رگرسیون و همبستگی تکنیک‌های جداگانه و با اهداف متفاوتی هستند و نیازی به همراهی با یکدیگر ندارند. تفسیر ضرایب همبستگی در ذیل بحث می‌شود (ارتباط و علیت).

### رگرسیون

به‌منظور نمایش میزان برازندگی خط رگرسیون به اطلاعات، ارائه یک خط رگرسیون مناسب به همراه یک نمودار پراکنش بر داده‌های خام بسیار مطلوب است. بدون مشاهده داده‌ها، ترسیم خط برازنده شده بر آن‌ها، اطلاعات تقریباً بیش‌تری از خود معادله رگرسیون در اختیار می‌گذارد. ارائه مقدار شیب خط همراه با خطای معیار آن، عرض از مبدأ خط رگرسیون، و شاخص پراکنندگی نقاط حول خط برازنده شده (انحراف معیار خطا) اطلاعات مفیدی را در اختیار خوانندگان می‌گذارد. بدیهی است همانند سایر برآوردهای آماری می‌توان برای خط رگرسیون نیز فاصله اطمینان و فواصل پیشگویی شده را تشکیل داد [۱۲]. واضح است خطوطی که این مقادیر را به هم وصل می‌کنند با خط رگرسیون موازی نیستند؛ زیرا انحراف معیار برآورد شده در نقاط مختلف متفاوت است، اما منحنی‌های مذکور نسبت به مقادیر روی محور طول‌ها و بدون توجه به توده مشاهدات، عدم قطعیت (uncertainty) بیش‌تری را در مورد پیشگویی نشان می‌دهند [۱].

رگرسیون بر روی داده‌هایی که شامل زیرگروه‌های متمایز هستند، ممکن است نتایج گمراه‌کننده‌ای به همراه داشته باشد، خصوصاً اگر گروه‌ها در سطح میانگین‌های متغیر وابسته ( $y$ ) متفاوت باشند. با استفاده از تجزیه و تحلیل کواریانس ممکن است نتایج قابل اطمینان‌تری به‌دست آورد [۱].

تفسیر تجزیه و تحلیل رگرسیون در قسمت‌های بعد (پیشگویی و آزمون‌های تشخیصی) بحث شده است.

## تجزیه و تحلیل‌های پیچیده

در بسیاری از مطالعات، مشاهدات مورد نظر اولیه ممکن است تحت تأثیر متغیرهای دیگر قرار گیرند. این متغیرها می‌توانند «هر صفت متفاوت» بین افراد باشند و همچنین می‌توانند بر پیامد مورد مشاهده تأثیر بگذارند. برای مثال، در کارآزمایی‌های بالینی، این متغیرها ممکن است شامل خصوصیات بیمار یا علائم و نشانه‌ها باشند. با روش‌های رگرسیون چندگانه مناسب می‌توان تمام و یا برخی از متغیرهای هم‌نوسان (covariate) را به‌منظور تشریح یا پیشگویی یک متغیر پیامد ترکیب کرد. این متغیرها ممکن است متغیر پیوسته (فشار خون)، کیفی (ترموموز بعد از عمل جراحی) یا زمان بقا باشند که به ترتیب با استفاده از رگرسیون خطی چندگانه، رگرسیون لجستیک چندگانه یا تجزیه و تحلیل رگرسیونی خطرهای نسبی کاکس می‌توان به این منظور نایل آمد. حتی در کارآزمایی‌های بالینی تصادفی، محققین ممکن است نیاز به اطمینان داشته باشند که اثر درمان هنوز بعد از تعدیل هم‌زمان برای چندین عامل خطر موجود است. زمانی که مدل‌ها برای به‌دست آوردن برآوردهای تعدیل شده برای سایر متغیرها به کار می‌روند، باید به روشنی مشخص شود که کدام متغیرها تعدیل شده‌اند. بر چه اساسی انتخاب گردیده‌اند و در صورت تناسب متغیرهای پیوسته در تجزیه و تحلیل به چه صورت رفتار کرده‌اند.

روش‌های آماری چند متغیره که با بیش از یک متغیر پیامد هم‌زمان سر و کار دارد، در واقع نیازمند کمک افراد ماهر بوده، فراتر از قلمرو این مقاله‌اند. هر روش پیچیده آماری باید به‌صورتی ارائه شود که برای خواننده قابل درک باشد. گنجاندن موارد تکنیکی در یک ضمیمه می‌تواند مفید باشد.

## دستورالعمل‌های آماری بخش نتایج: ارائه نتایج

## ارائه آماره‌های خلاصه

مقادیر میانگین را نباید بدون همراهی برخی از اندازه‌های پراکندگی، نظیر انحراف معیار و خطای معیار ارائه کرد. انحراف معیار (SD) برای نمایش تغییرات در

بین افراد و خطای معیار میانگین (SE) برای نمایش دقت میانگین نمونه‌ای به کار می‌رود. لذا در هر مقاله خاص باید به روشنی مشخص شود که کدام یک از دو مورد فوق، همراه میانگین ارائه شده است. استفاده از نماد  $\pm$  برای الحاق انحراف معیار و خطای معیار به میانگین (مثل  $1/9 \pm 1/4$ ) باعث سر در گمی می‌شود و باید از آن اجتناب کرد. به همین دلیل، در حال حاضر برخی از مجلات پزشکی اجازه استفاده از چنین نماد تعریف نشده‌ای را نمی‌دهند. ارائه میانگین‌ها برای مثال به‌صورت  $1/4$  (با SE مساوی  $1/9$ ) و یا  $4/2$  (با SD مساوی  $7/4$ ) ارجحیت دارد. فواصل اطمینان، روش خوبی برای ارائه بیان منطقی از عدم قطعیت میانگین‌های نمونه، نسبت‌ها و سایر آماره‌ها هستند. برای مثال، فاصله اطمینان ۹۵ درصد برای میانگین واقعی جامعه، تقریباً در فاصله دو خطای معیار کم‌تر از میانگین نمونه تا دو خطای معیار بیش‌تر از آن است. توصیه می‌شود برای نشان دادن فواصل اطمینان، به جای استفاده از نماد  $\pm$  از نمادی واضح‌تر (از  $10/4$  تا  $18/0$ ) استفاده شود.

هنگامی که مقایسات جفتی صورت می‌گیرد، مثل وقتی که از آزمون «تی» جفتی استفاده می‌شود، نشان دادن میانگین و انحراف معیار تفاضل بین مشاهدات و یا خطای معیار تفاضل میانگین (هرکدام مناسب‌تر هستند)، مهم است [۲].

برای داده‌هایی که با روش‌های توزیع آزاد (ناپارامتری) مورد تجزیه و تحلیل قرار می‌گیرند، نشان دادن میانه و دامنه میانی پوشش (برای مثال ۹۵ درصد مشاهدات) به جای استفاده از میانگین و انحراف معیار مناسب‌تر است [۲]. همچنین چنانچه تجزیه و تحلیل بر روی داده‌های تبدیل شده انجام شده است، ارائه میانگین و انحراف معیار داده‌های خام احتمالاً نمی‌تواند شاخص خوبی برای نشان دادن گرایش‌های مرکزی و پراکندگی داده‌ها باشد.

یکی از نکات مهم که باید مورد توجه نویسندگان مقالات قرار گیرد، مربوط به زمانی است که درصدها مورد استفاده قرار می‌گیرند. در این‌جا همواره باید

مخرج کسر را که این درصدها براساس آن محاسبه و ارائه شده‌اند را به روشنی مشخص کرد، لکن باید توجه داشت در نمونه‌هایی که از نظر آماری کوچک هستند، استفاده از درصدها نمی‌تواند خیلی مفید باشد. زمانی که بین درصدهای مربوط به یک موضوع مغایرتی وجود دارد، تعیین تفاضل مطلق و تفاضل نسبی آن‌ها حائز اهمیت است. برای مثال، یک کاهش از مقدار ۲۵ درصد به ۲۰ درصد را می‌توان به صورت ۵ درصد (تفاضل مطلق) و یا ۲۰ درصد (تفاضل نسبی) بیان کرد [۲].

### نتایج برای افراد

به طور کلی، دامنه تغییرات، شاخص خوبی برای بیان تغییرات مجموعه‌ای از مشاهدات نیست. زیرا می‌تواند تحت تأثیر شدید هر مقدار بزرگ یا کوچک قرارگیرد. البته این دامنه ممکن است با افزایش حجم نمونه نیز بزرگ‌تر شود. اگر داده‌ها منطقی‌اً دارای توزیع نرمال باشند، یعنی دلیلی بر غیرنرمال بودن آن‌ها وجود نداشته باشد، فاصله دو انحراف معیار در دو طرف میانگین (فاصله اطمینان) تقریباً ۹۵ درصد مشاهدات را می‌پوشاند؛ اما در سایر توزیع‌ها (به غیر از توزیع نرمال)، دامنه‌های درصدی به صورت گسترده‌تر مورد استفاده قرار می‌گیرند.

هر چند تجزیه و تحلیل آماری با متوسط آثار سر و کار دارد، اما در بسیاری از شرایط، در نظر گرفتن نحوه پاسخ افراد نیز مهم است. مثلاً اغلب دانستن این که چه تعداد از بیماران با یک روش درمانی خاص بهبود نیافته‌اند و نیز دانستن متوسط اثر (میانگین یا میانه) از نظر بالینی بسیار مهم است. یک متوسط اثر نباید به صورت کاربرد برای همه بیماران، تفسیر شود.

### ارائه نتایج آزمون‌های فرضیه

آزمون‌های فرضیه مقدار مشاهده شده آماره آزمون را به دست می‌دهند که به منظور استخراج مقادیر  $p$  مربوط باید با مقادیر جداول توزیع مربوط (نرمال،  $t$ ، کای دو و غیره) مقایسه شوند. مطلوب‌تر آن است که مقادیر مشاهده شده برای آماره‌های آزمون (نظیر  $t$ ،  $z$ ،  $\chi^2$ ) به همراه مقادیر  $p$  گزارش شوند. مقادیر کمی که مورد

آزمون قرار می‌گیرند، مثل مقادیر میانگین، نسبت‌ها و ضرایب همبستگی، باید بدون در نظر گرفتن معناداری یا غیرمعناداری آزمون، بیان شوند. از طرف دیگر، باید دقیقاً مشخص شود که کدام داده‌ها مورد تجزیه و تحلیل قرار گرفته‌اند. توصیه می‌شود از به کار بردن نمادهایی نظیر ستاره برای نمایش سطح احتمال پرهیز شود، ولی چنانچه مجبور به ارائه این نمادها باشیم یکسان بودن این نمادها در سرتاسر مقاله الزامی است.

دلیل مشخصی برای مرسوم بودن استفاده از مقادیر  $p$  به صورت  $p < 0.05$ ، یا  $p < 0.01$  یا  $p < 0.001$  وجود ندارد و می‌توان گفت که تنها به طور قراردادی از چنین نمادهای خاصی استفاده می‌شود. ارائه مقادیر دقیق  $p$  حداکثر با دو رقم معناداری مثلاً  $p = 0.18$  یا  $p = 0.03$  مورد تأکید نگارندگان است. لذا مشخص کردن سطوح  $p$  های کم‌تر از  $0.001$  ضروری نیست. نامیدن مقادیر  $p > 0.05$  به صورت «غیر معنادار» نیز توصیه نمی‌شود، زیرا می‌تواند نتایجی را که کاملاً از نظر آماری معنادار نیستند پنهان کند و فقط از یک تأثیر واقعی خبر دهد. وقتی مقادیر واقعی  $p$  ارائه می‌شوند تشخیص بین کوچک‌تر از ( $<$ ) و بزرگ‌تر از ( $>$ ) مهم است. مقادیر  $p$  بین دو محدوده نیز باید با ترتیبی منطقی بیان شود. برای مثال در مواقعی که  $p$  بین  $0.01$  و  $0.05$  است، می‌نویسیم  $0.01 < p < 0.05$  و ضرورتی به تکرار مقادیر  $p$  ارائه شده در جداول در خود متن نیست.

### شکل‌ها (نمایش گرافیکی)

نمایش‌های گرافیکی به خوبی قادرند نتایج یک تحقیق را به خوانندگان منتقل کنند. بدین جهت استفاده از نمودارهایی که مشاهدات افراد را نشان دهند، توصیه می‌شود. ترجیحاً بهتر است نقاط مربوط به یک فرد خاص در مواقع مختلف در روی منحنی به هم متصل گردند، یا نمادهایی برای تأکید ارتباط این نقاط به کار گرفته شوند. روش مفید دیگر، رسم تفاضل‌ها برای هر فرد در موقعیت‌های مختلف است.

روش متداول رسم نمودار میل‌های برای خطا با نشان دادن یک خطای معیار (SE) در بالا و پایین

همواره نتایج به دست آمده از ماشین حساب یا کامپیوتر باید گرد شوند لازم است نویسنده میانگین‌ها، انحراف معیارها و سایر آماره‌های محاسبه شده را با توجه به دقت داده‌های اصلی ارائه کند البته توجه به این نکته ضروری است که محاسبات تنها به هنگام ارائه نتایج باید گرد شوند. لذا از گرد کردن محاسبات در حین تجزیه و تحلیل یا قبل از آن شدیداً اجتناب شود. به صورت متداول، معمولاً میانگین‌ها را با یک رقم اعشار بیش‌تر از داده‌های اولیه ارائه می‌کنند، اما ممکن است انحراف معیارها یا خطای معیارها با ارقام اعشاری بیش‌تری ارائه شود. در اکثر مواقع، دو رقم اعشار بیش‌تر از تعداد ارقام اعشاری داده‌های خام مناسب است. معمولاً درصدها را با یک رقم اعشاری نشان می‌دهند و حتی بیش‌تر اوقات نیازی به یک رقم اعشار نیست. فقط در مواردی که سرجمع‌های کل یا حاشیه‌ای حائز اهمیت باشند، برای نشان دادن درصد باید تا یک رقم و بسیار به ندرت تا دو رقم اعشار بیش‌تر از داده‌های خام استفاده کرد. در ارائه آماره‌های آزمون، نظیر «تی»، کای اسکور و مقدار  $\chi^2$  دو رقم اعشار کفایت می‌کند [۲۰۱].

### واژه‌های تکنیکی گوناگون

هر چند ارائه تعاریف تمام واژه‌های آماری در این مقاله میسر نیست، اما توضیحات زیر فقط مرتبط با واژه‌هایی هستند که مکرر به‌طور اشتباه و گمراه‌کننده مورد استفاده قرار می‌گیرند.

از واژه «همبستگی» ترجیحاً نباید به‌عنوان یک واژه عمومی برای توصیف هر رابطه استفاده کرد. این واژه یک معنای خاص تکنیکی به‌عنوان اندازه‌گیری ارتباط دارد و به همین جهت باید در کارهای آماری در نظر گرفته شود [۶].

از واژه «بروز» باید برای توصیف میزان وقوع موارد جدید یک ویژگی خاص، در یک مطالعه نمونه‌ای و یا جمعیت، مثل تعداد گزارش‌های جدید سرطان در یک مدت معین استفاده شود. نسبت کل موارد در یک نمونه، شامل کلیه موارد قدیم و جدید که در حال حاضر دارای مشخصه‌ای خاص هستند، شیوع نام دارد [۷].

میانگین، تنها ۶۷ درصد فاصله اطمینان را بیان می‌کند، حال آن‌که فواصل اطمینان ۹۵ درصد اولویت دارند. نمایش چنین اطلاعاتی در نمودارها، منوط به همان جنبه‌هایی است که در بالا مورد بحث قرار گرفت. در مواقعی که داده‌ها پیچیده‌تر از آن هستند که بتوان آن‌ها را در جدول نمایش داد، استفاده از نمودار با ارزش‌ترین روش است. از سوی دیگر، نموداری که مثلاً تنها دو یا سه میانگین یا خطای معیار یا فواصل اطمینان آن‌ها را نشان دهد، اغلب باعث اتلاف جا می‌شود، زیرا یا باید اطلاعات بیش‌تری، مثل داده‌های خام اضافه شوند، و یا به جای آن، خلاصه مقادیر در متن یا جدول گذاشته شوند. همچنین قرار دادن جداول از این نظر ارجحیت دارد که احتمالاً مقادیر داده‌ها توسط سایرین در آینده مورد تجزیه و تحلیل قرار گیرد (مثلاً برای تجزیه و تحلیل افزوده).

دیگرام‌های پراکنش درباره دو متغیر باید کلیه مشاهدات را نشان دهند، حتی اگر این امر بدین معنا باشد که تعدیل مختصری در نقاط همانند صورت گرفته است. این کار نیز می‌تواند با جایگزین کردن علائم نمایشی به جای اعداد واقعی از نقاط همانند، بیان شود [۲۰۱].

### جداول

نتایج عددی در ستون‌ها بسیار آسان‌تر از سطرها قابل رؤیت هستند. بنابراین توصیه می‌شود اطلاعات مختلف - مثل میانگین‌ها و خطاهای معیار - را در ستون‌های مجزا نوشته، تعداد مشاهدات برای هر نتیجه در یک جدول ارائه شود. اگر سطرهای جداولی که درباره بیماران خاص، مناطق جغرافیایی و غیره، اطلاعاتی را ارائه می‌کنند، بر طبق یکی از متغیرهای ارائه شده مرتب شوند، آسان‌تر خوانده می‌شوند [۲۰۱].

### دقت عددی

باید به این موضوع توجه داشت که دقت‌های کاذب، نه فقط چیزی به ارزش مقاله نمی‌افزایند، بلکه حتی از اعتبار آن کاسته، انگیزه و تمایل به خواندن آن مقاله را (قابلیت خواندن) تحت تأثیر قرار می‌دهند. از آن جا که

حتی اگر یک اثر (اختلاف) واقعی بزرگ موجود باشد، در صورت وجود مشاهدات کم، یک نتیجه غیرمعنادار کاملاً محتمل است. بر عکس، اگر حجم نمونه خیلی بزرگ باشد، اگر فقط یک اثر واقعی کوچک وجود داشته باشد، ممکن است یک نتیجه معنادار آماری رخ بدهد. بنابراین، معناداری آماری را نباید مترادف با اهمیت بالینی (اختلاف بین دو جامعه یا دو روش درمانی) تلقی کرد.

تفسیر نتایج آزمون فرضیه عمدتاً از مطلب فوق تبعیت می‌کند. یک نتیجه معنادار لزوماً به معنای وجود اثر واقعی یا اختلاف نیست؛ زیرا همواره خطر مشاهده یک یافته مثبت کاذب وجود دارد. این خطر برای مقادیر  $p$  کوچک‌تر کاهش می‌یابد. به علاوه یک نتیجه غیرمعنادار (که به‌طور قراردادی به‌صورت  $p > 0.05$  نمایش داده می‌شود) به معنای عدم وجود اثر (اختلاف بین دو جامعه یا روش درمانی) نیست، بلکه بدین معنا است که داده‌ها تنها با این‌که اثری وجود نداشته باشد سازگار هستند. هرچند سطح  $0.05$  یک نقطه برش مناسب است، لکن مقادیر  $p$  بین  $0.04$  تا  $0.06$  که تفاوت چندانی ندارند، باید به تفسیری مشابه و نه کاملاً متفاوت منجر شوند. بنابراین، تعیین هر نتیجه با  $p > 0.05$  به‌عنوان غیرمعنادار ممکن است مخاطب را گمراه کند. از این رو پیشنهاد و توصیه می‌شود که نتایج آزمون‌های فرضیه به همراه مقادیر واقعی  $p$  ارائه گردند.

از آن‌جا که فواصل اطمینان، درجه عدم قطعیت مرتبط به نتیجه (مثل تفاوت بین دو میانگین) را نشان می‌دهند، خواه این اختلاف از نظر آماری معنادار باشد یا نباشد، کاملاً در تفسیر نتایج، خصوصاً برای مطالعات کوچک مفید هستند. استفاده از فواصل اطمینان، همراه با نتایج غیرمعنادار می‌تواند آموزنده باشد.

### تعدد آزمون‌های فرضیه

در بسیاری از پروژه‌های تحقیقاتی، برخی از آزمون‌های فرضیه به مقایسه‌های مهمی می‌پردازند که در پروتکل تحقیق از ابتدا در نظر گرفته شده‌اند. آزمون‌های فرضیه‌هایی که از قبل راجع به انجام آن‌ها تصمیم‌گیری

واژه «ناپارامتری» به نوع خاصی از تجزیه و تحلیل‌های آماری - مثل آزمون من یو ویتنی - اطلاق می‌شود که استفاده از روش‌های متداول پارامتری در آن امکان‌پذیر نیست. یادآوری این نکته ضرورت دارد که واژه ناپارامتری ویژگی مربوط به خود مشاهدات نیست و فقط به روش آماری مربوط می‌شود [۱۱].

از واژه «پارامتر» نباید به جای «متغیر» که برای اشاره به اندازه‌گیری یا ویژگی که به مشاهدات بر می‌گردد، استفاده کرد. پارامترها خصوصیات توزیع‌ها و روابط در جمعیت‌هایی هستند که با تجزیه و تحلیل آماری یک نمونه از مشاهدات برآورد می‌شوند.

صدک‌ها - وقتی که دامنه مقادیر یک متغیر به گروه‌های مساوی تقسیم می‌شود، نقاط برش میانه، ثلث‌ها، چارک‌ها، پنجم‌ها و غیره هستند. به خود گروه‌ها باید به‌عنوان نصف‌ها، ثلث‌ها، چارک‌ها و... اشاره شود.

«حساسیت» قدرت یک آزمون تشخیصی برای شناسایی یک بیماری است که به واقع وجود دارد (یعنی نسبت نتایج مثبت در کسانی که واقعاً بیمارند). «ویژگی» قدرت یک آزمون تشخیصی برای شناسایی عدم بیماری است وقتی که واقعاً بیماری وجود ندارد (یعنی نسبت نتایج منفی در کسانی که واقعاً بیمار نیستند). همچنین به مطلب (پیشگویی و آزمون‌های تشخیصی) در ذیل مراجعه شود [۸].

### دستورالعمل‌های آماری برای بحث و تفسیر

#### تفسیر آزمون‌های فرضیه

یک آزمون فرضیه با استفاده از مقدار  $p$ ، مقبولیت داده‌های مشاهده شده را وقتی که تعدادی فرضیه صفر (مثل عدم تفاوت بین گروه‌ها) صحیح است، ارزیابی می‌کند. مقدار  $p$  احتمالی است که در صورت درست بودن فرضیه صفر، داده‌های مشاهده شده شانس و تصادفی روی داده باشد (یعنی فقط به واسطه تغییرات نمونه باشد). اگر  $p$  بزرگ باشد (بزرگ‌تر از مقدار مشخص شده سطح معناداری) داده‌ها با فرضیه صفر سازگارند که در نتیجه، فرضیه مورد بحث نمی‌تواند رد شود.

رابطه علیتی بین متغیرهای مورد مطالعه را فراهم نمی‌کند. استنتاج علیت در کارآزمایی‌های تصادفی آسان‌تر است. در مقایسه متغیرهایی که هر دو با زمان تغییر می‌کنند، باید خیلی دقت کرد، زیرا دستیابی به ارتباطات آشکار کاذب آسان است.

### پیشگویی و آزمون‌های تشخیصی

حتی اگر تجزیه و تحلیل رگرسیون، یک ارتباط معنادار آماری را بین دو متغیر نشان دهد، در بعضی موارد به هنگام استفاده از معادله رگرسیون به منظور پیشگویی سطح کمی متغیر (Y) از روی متغیر دیگر (X) ممکن است بی‌دقتی قابل توجهی وجود داشته باشد. دقت چنین پیشگویی‌هایی را نمی‌توان با ضریب همبستگی یا رگرسیون ارزیابی کرد. اما نیاز به محاسبه فاصله پیشگویی برای مقدار Y پیشگویی شده متناظر با مقدار X خاص دارد. خط رگرسیون را فقط باید برای پیشگویی متغیر Y از روی متغیر X به کار برد و بر عکس آن صادق نیست.

یک آزمون تشخیصی با حساسیت و ویژگی بالا الزاماً نمی‌تواند آزمون مفیدی برای مقاصد تشخیصی باشد، مخصوصاً وقتی که شیوع بیماری در جمعیت مرجع مورد بررسی پایین باشد. در چنین شرایطی، محاسبه نسبت افراد با نتایج آزمون مثبت که واقعاً بیمارند و به آن «ارزش اخباری مثبت» می‌گویند، مفید است. باید توجه داشت که توافقی بر سر تعریف «میزان مثبت کاذب» یا «میزان منفی کاذب» وجود ندارد و همیشه باید به وضوح و روشنی مشخص کرد که چه چیزی محاسبه می‌شود. به بهترین نحو و با کمک یک جدول 2×2 می‌توان نشان داد که نتایج آزمون را با وضعیت بیماری واقعی بیماران مرتبط می‌کند.

یک مشکل تشخیصی مشابه با متغیرهای پیوسته مطرح می‌شود. طبقه‌بندی مقادیر خارج از دامنه طبیعی به عنوان «غیرطبیعی»، برای یک متغیر عادی است، اما اگر شیوع ناهنجاری واقعی کم باشد، بیش تر مقادیر خارج از دامنه نرمال، نرمال خواهند بود. تعریف ناهنجاری باید براساس هر دو معیار بالینی و آماری باشد.

نشده، کمکی هستند؛ خصوصاً اگر پس از رؤیت نتایج اتخاذ شده باشند. تشخیص این دو مورد مهم است و به آزمون‌هایی که فرضیات آن‌ها از قبل مشخص شده‌اند، اهمیت بیش‌تری داده می‌شود. سایر آزمون‌ها باید تنها به عنوان آزمون‌های اکتشافی (برای تشکیل فرضیات جدید و به منظور بررسی در مطالعات آینده) در نظر گرفته شوند. این بدان جهت است که وقتی تعداد بسیار زیادی آزمون فرضیه در تجزیه و تحلیل یک مطالعه انجام می‌شوند، پس از مقایسه چند زیرگروه یا بررسی تعداد زیادی متغیر، برخی نتایج مثبت کاذب را می‌توان انتظار داشت که تنها به وسیله شانس مطرح می‌گردند و ممکن است مشکلات قابل توجهی را در تفسیر نتایج ایجاد کنند. قطعاً هر چه آزمون‌های بیش‌تری انجام شود، احتمال یافتن نتایج معنادار بیش‌تر می‌شود، اما تعداد یافته‌های مورد انتظار مثبت کاذب نیز افزایش می‌یابد. یک راه کاهش خطر نتایج مثبت کاذب، در نظر گرفتن سطح کوچک‌تر مقدار p به عنوان معیار معناداری آماری است. مشکل پیچیده‌تر زمانی به وقوع می‌پیوندد که آزمون‌های معناداری روی داده‌های همبسته انجام شود. تجزیه و تحلیل داده‌های مکرر (اندازه‌گیری‌های تکراری) مثالی از این نوع است؛ وقتی آزمونی مشابه روی داده‌هایی برای یک متغیر یکسان که از افراد مشابه در زمان‌های مختلف جمع‌آوری شده، انجام می‌شود. مشکل دیگر مربوط به مواقعی است که تجزیه و تحلیل‌های جداگانه دو یا چند متغیر همبسته به صورتی انجام شود که گویی آن متغیرها مستقل هستند. در این موارد، هر اثباتی نمی‌تواند اهمیت شواهد را خیلی افزایش دهد؛ زیرا آزمون‌ها مربوط به داده‌های مشابه هستند. برای مثال، فشار خون سیستولیک و دیاستولیک بسیار شبیه به هم عمل می‌کنند، همان‌طور که راه‌های دیگر ارزیابی پاسخ بیماران عموماً شبیه هم هستند. در چنین مواردی، تفسیر بسیار دقیق نتایج مورد نیاز است.

### ارتباط و علیت

یک ارتباط معنادار آماری که از همبستگی یا تجزیه و تحلیل کای دو حاصل شده به تنهایی دلیل درستی برای

## ضعف‌ها

چنانچه محقق نسبت به ضعف‌های تحقیق خود آگاهی دارد، بهتر است ضعف‌های طراحی و اجرای تحقیق را مشخص کند. این امر موجب می‌شود تا آثار احتمالی این ضعف‌ها را روی نتایج تحقیق و تفسیر آن‌ها مشاهده کند، تا این‌که آن‌ها را نادیده گرفته، تصور کند که کسی متوجه آن‌ها نخواهد شد.

## نتیجه‌گیری

هدف از اعمال روش‌های آماری، ارائه یک دلیل واقعی از شواهد علمی ناشی از یک تحقیق است. در نظر گرفتن و انجام تجزیه و تحلیل آماری مناسب در طراحی مطالعات، و ارتباط دادن یافته‌ها با روشی روشن و هدف‌دار، نیازمند مهارت و تجاربی است که به راحتی امکان‌پذیر نیست. متأسفانه محققین علوم پزشکی به دلیل گرفتاری‌های شغلی، کم‌تر فرصت می‌یابند تا موارد فوق را که در دروس دوره‌های مختلف دانشگاهی کم‌تر به آن پرداخته می‌شود، فراگرفته، آن‌ها را در نگارش مقالات و گزارش‌های تحقیقاتی خود به کار برند. علی‌رغم نیاز مبرم محققین به آگاهی درباره طراحی، تجزیه و تحلیل، و تفسیر داده‌های خود - که در بررسی مقالات ارسالی به مجلات علمی پزشکی کاملاً مشهود است و بسیاری از اعضای هیأت‌های تحریریه و داوران این مجلات به کرات بر آن اذعان کرده‌اند - هنوز بسیاری از محققین از تبحر کافی در آن بهره‌مند نیستند. دستورالعمل‌های فوق به مؤلفین کمک می‌کند تا ضمن استفاده از آن‌ها و اصلاح مقالات خود مطابق استانداردهای این مجلات، در چالش‌های آماری گرفتار نشوند. کماکان بر خود لازم می‌دانیم تا توصیه‌های قبلی مبنی بر استفاده از یک آمارشناس مجرب را قویاً تکرار کنیم. استفاده از این دستورالعمل‌ها - اگرچه دستورالعمل‌های بسیار مفیدی بوده و براساس تجارب داوران و هیأت تحریریه این مجلات که بیش از همه با ایرادات مقالات ارسالی سروکار دارند، تهیه شده باشند - هیچگاه نمی‌تواند جای استفاده از نظر افراد باتجربه را در بر طرف کردن ایرادات

متدولوژیک تحقیق بگیرد. باتجربه‌ترین محققین نیز از مشورت با افراد خبره در زمینه‌های مختلف تحقیقات خصوصاً روش‌های آماری، تجزیه و تحلیل، و تفسیر داده‌ها بی‌نیاز نیستند. بنابراین توجه محققین را به این مطلب معطوف می‌داریم که موفقیت در ارائه مقالات علمی در مجلات پزشکی، بیش از آن‌که به دانسته‌ها، مدرک تحصیلی و سایر خصوصیات علمی محققین بستگی داشته باشد، بر تجارب آنان و کسانی که در این زمینه طرف مشورت آنان قرار می‌گیرند، متکی است.

## منابع

1. رودباری مسعود، صانعی سیدحسن. د ترجمه آمار با اطمینان، آلتمن داگلاس جی، همچنین روبرت، برایانت ترور ان، گاردنر مارتین جی. انتشارات دانشگاه علوم پزشکی زاهدان. ۱۳۸۳.
2. Altman DG. Practical Statistics for medical research. London: Chapman & Hall. 1997.
3. Armitage P, Berry G. Statistical methods in medical research. 3<sup>rd</sup> ed. Oxford: Blackwell Science, 1994.
4. Bland M. An introduction to medical Statistics. 3<sup>rd</sup> ed. Oxford University press, 2000.
5. Campbell MJ, Machin D. Medical Statistics. A commonsense approach. 3<sup>rd</sup> ed. Chichester: John Wiley, 1999.
6. صانعی سیدحسن. آمار حیاتی کاربردی. چاپ دوم. تهران: انتشارات دانشمند. ۱۳۸۲.
7. صدری غلامحسین، محجوب حسین. اصول اپیدمیولوژی و روش‌های آماری در اپیدمیولوژی. معاونت آموزشی و پژوهشی دانشگاه علوم پزشکی و خدمات بهداشتی درمانی همدان. ۱۳۸۰.
8. Petrie A, Sabin C. Medical Statistics at a glance. Oxford: Blackwell. 2000.
9. آیت‌اللهی سیدمحمدتقی. در ترجمه اصول و روش‌های آمار زیستی، واین، دانیل. چاپ دوم. تهران: انتشارات امیر کبیر. ۱۳۶۸.
10. Kirkwood BR, Sterne JAC. Medical Statistics. Oxford: Blackwell. 2003.
11. سرفراز علی‌اکبر، غفرازادگان کامران. در ترجمه آمارپزشکی، ساندرز بت داوسون، تراپ روبرت. مشهد: انتشارات دانشگاه فردوسی. ۱۳۷۷.
12. محمد کاظم، ملک‌افضلی حسین، نهایتپیان وارنگس. روش‌های آماری و شاخص‌های بهداشتی. تهران: انتشارات ابلاغ. ۱۳۷۸.
13. صانعی سیدحسن. تجزیه و تحلیل داده‌های بقاء. تهران: اندیشمند. ۱۳۸۰.