مجلهٔ غدد درون‌ریز و متابولیسم ایران
دانشگاه علوم پزشکی و خدمات بهداشتی ـ درمانی شهید بهشتی
سال سوم ـ شمارهٔ ۱ ـ صفحه‌های ۷۰-۵۷ (بهار ۱۳۸۰)

رهنمودهای آماری برای نویسندگان مقاله‌های مجله‌های پزشکی

دکتر محسن جانقربانی

چکیده: در سال‌های اخیر فعالیت‌های پژوهشی، تعداد مجله‌های پزشکی و تعداد مقاله‌هایی که به مجله‌های علوم پزشکی کشور ارسال می‌شود به طور چشمگیری افزایش یافته است. در بیشتر این مقاله‌ها تجزیه و تحلیل‌هایی ارایه می‌شود که بدون یاری آمارشناسان و اپیدمیولوژیست‌ها انجام شده است. اگر چه تقریباً تمام پژوهشگران پزشکی تا حدودی با مبانی آمار آشنا هستند، ولی راه آسانی برای آشنایی کامل آنها با مفاهیم و اصول آماری وجود ندارد. همچنین درخصوص نحوهٔ طراحی، تحلیل و نوشتن یک پروژهٔ کامل کمک چندانی در اختیار آنها نیست. تا حدودی به این دلایل بسیاری از مقاله‌هایی که در مجله‌های پزشکی منتشر می‌شوند، از نظر آماری ضعیف یا حتا ممکن است غلط باشند. در این مقاله سعی شده به روش‌های آماری اشاره شود تا خوانندهٔ آگاه با دسترسی به داده‌های اصلی بتواند یافته‌های گزارش شده را تأیید نماید. در صورت امکان باید یافته‌ها سنجیده شده و با شاخص‌های مناسب سنجش خطا یا ابهام نظیر دامنه‌های اطمینان ارایه شود و از تکیه بر آزمون فرضیهٔ آماری نظیر مقدار P به طور انحصاری که اطلاعات کمّی مهمی را بیان نمی‌کند، اجتناب گردد. این رهنمود قصد دارد به نویسندگان کمک کند تا بدانند چه چیز از نظر آماری مهم تلقی می‌شود و چگونه آن را در مقاله‌های خود ارایه نمایند. در این مقاله راجع به نحوهٔ انتخاب روش‌های آماری مناسب بحث نشده است؛ چنین اطلاعاتی را می‌توان با مشورت با آمارشناسان بدست آورد، ولی این مقاله توجه خوانندگان را به سوءاستفاده از روش‌های آماری جلب می‌نماید و تلاش می‌کند اطلاعات کلی و توصیه‌هایی را دربارهٔ جنبه‌های مهم طراحی، تجزیه و تحلیل و نحوهٔ ارایه یافته‌ها به آنها پیشنهاد کند.

واژگان کلیدی: تحلیل آماری، رهنمودها، مقاله‌های پزشکی، پژوهش پزشکی، طراحی مطالعه

مقدمه

در سال‌های اخیر فعالیت‌های پژوهشی و تعداد مجله‌های علمی پزشکی و نیز تعداد مقاله‌هایی که برای این مجله‌ها ارسال می‌شوند به طور چشمگیری افزایش یافته است.

در اغلب مقاله‌هایی که در مجله‌های پزشکی منتشر می‌شوند، تجزیه و تحلیل‌هایی ارایه می‌شود که بدون کمک آمارشناس یا اپیدمیولوژیست انجام شده است. اگرچه تقریباً تمام پژوهشگران پزشکی تا حدودی با مبانی آمار آشنا هستند، با مفاهیم و اصول آمار چندان آشنا نمی‌باشند و نیز آشنایی چندانی با نحوهٔ طراحی، تجزیه و تحلیل و نوشتن یک مقالهٔ پژوهشی ندارند. به این دلایل، بیشتر مقاله‌هایی که در مجله‌های پزشکی منتشر می‌شوند، از نظر آماری ضعیف یا حتا غلط می‌باشند.[1] در بازنگری‌های متعدد از مقاله‌های منتشر شده،

دانشکده بهداشت،
دانشگاه علوم پزشکی و خدمات بهداشتی ـ درمانی اصفهان

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

خطاهای آماری زیادی مشاهده می‌شود که موجب نگرانی می‌باشد.

هیچکدام از مجله‌های پزشکی که در سطح کشور منتشر می‌شوند، به نویسندگان توصیه‌های آماری نمی‌کنند. از این رو، با آگاه نمودن پژوهشگران پزشکی از اصول مهم آماری و با مشخص کردن اطلاعاتی که باید در یک مقاله منظور شود، می‌توان به آنها کمک نمود.[۱،۲]

در این مقاله سعی می‌شود رهنمودهای جامعی در این خصوص ارایه گردد. البته درخصوص رهنمودهای نوشتن مقاله‌های پزشکی کارگاه‌های مقاله‌نویسی و روش تحقیق در سطح دانشگاه‌های سراسر کشور برگزار می‌شود، مقاله‌های متعددی نیز منتشر شده است[۳-۱۱] و خوانندگان محترم می‌توانند به آنها مراجعه کنند، ولی تا کنون رهنمودهای آماری به زبان فارسی منتشر نشده است. در سال‌های اخیر که تعداد مجله‌های پزشکی فارسی به سرعت رو به فزونی است و مقاله‌هایی که در این مجله‌ها منتشر می‌شوند به طور روزافزونی در حال افزایش می‌باشند، نیاز به ارتقای کیفیت این مقاله‌ها به شدت احساس می‌گردد.

تصمیم بر اینکه این رهنمودها تا چه حد وارد جزئیات شوند و چه مطالبی را در نظر گیرند، مشکل است. در اینجا سعی نمی‌شود یک مجموعه قوانین تدوین گردد، بلکه تلاش می‌شود اطلاعاتی کلی و توصیه‌هایی دربارهٔ جنبه‌های مهم طراحی آماری، تجزیه و تحلیل و نحوهٔ ارایهٔ یافته‌ها داده شود.

در این مقاله فرض شده که خوانندگان تا حدودی با روش‌ها و نظریه‌های آماری آشنا هستند، زیرا قبل از انجام تحلیل‌های آماری قدری آشنایی با آمار ضروری است. این رهنمودها باید به آنهایی که اطلاعات محدودی در زمینهٔ آمار دارند، نشان بدهد که موضوع خیلی وسیع‌تر از منحصراً آزمون‌های معنی‌دار آماری می‌باشد و تفسیر صحیح داده‌ها چقدر می‌تواند مهم باشد. فقدان توصیه‌های دقیق در برخی موارد نشان می‌دهد که تحلیل آماری خوب به قضاوت، درک عمومی و نیز ارایهٔ روش‌های رسمی نیاز دارد. همان گونه که در مورد پزشکی صدق می‌کند، در آمار نیز نوعی هنر وجود دارد. در این مقاله سعی شده دامنهٔ رهنمودها عمدتاً به روش‌های آماری متداولتر محدود گردند.

ممکن است خوانندگان با اطلاعات یا توصیه‌های یک قسمت آشنا نباشند و متوجه آن نشوند. هرچند، تقریباً تمام مباحث را می‌توانند به طور جامعتر در درسنامه‌های آمار زیستی پیدا کنند،[۱۲-۱۵] ولی به شدت توصیه می‌شود از راهنمایی‌های آمارشناسان بهره جویند. بعلاوه چون که خطاهای طراحی را نمی‌توان بعداً اصلاح کرد، باید از همان اول در هنگام طراحی پروژهٔ پژوهشی به توصیه‌های حرفه‌ای عمل نمود و تا موقع تحلیل داده‌ها صبر نکرد.

این رهنمودها برای کمک به نویسندگانی تهیه شده که می‌دانند چه چیز از نظر آماری مهم است و چگونه آنها را در مقاله‌شان ارایه نمایند. در اینجا راجع به شیوهٔ انتخاب روش آماری مناسب به تفصیل بحث نمی‌شود. بهتر است چنین اطلاعاتی را با مشورت با یک آمارشناس بدست آورد، ولی سعی می‌شود توجه را به سوءاستفاده‌های خاصی که از روش‌های خاص آماری می‌شود، جلب نمود. ساختار این رهنمودها طبق ساختار متداول مقاله‌های پژوهشی پزشکی شامل روش بررسی، یافته‌ها (تحلیل و ارایه) و بحث (تفسیر) می‌باشد.

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

قسمت روش بررسی

اصول کلی

مهمترین نکته در روش بررسی این است که به طور روشن شرح داده شود چه کاری انجام شده است، که شامل طرح پژوهشی (تجربی، کارازمایی یا مطالعهٔ مشاهده‌ای) و گرداوری اطلاعات است. هدف باید ارایهٔ اطلاعات کافی برای درک کامل روش مطالعه باشد و در صورت تمایل، دیگران بتوانند آن را تکرار کنند. نویسندگان در طراحی پژوهش خود باید نکات زیر را در نظر گیرند.

- هدف از پژوهش و فرضیه‌های اصلی؛
- نوع افراد تحت مطالعه، بیان معیارهای ورود و خروج از مطالعه؛
- جمعیت هدف و نحوهٔ انتخاب افراد تحت مطالعه (روش نمونه‌گیری)؛
- تعداد افراد مطالعه شده (تعیین حجم نمونه)؛
- انواع روش‌های مشاهده و سنجش بکار رفته.

هر نوع مطالعه (مثلاً مطالعه‌های مشاهده‌ای و کارازمایی‌های بالینی) به اطلاعات اضافهٔ خاصی نیز نیاز دارند.

مطالعه‌های مشاهده‌ای

طرح مطالعه باید به روشنی توضیح داده شود. مثلاً شیوهٔ انتخاب گروه شاهد و همسان‌سازی باید به طور مفصل شرح داده شود. همچنین باید به روشنی بیان شود آیا مطالعه مورد ـ شاهدی، مقطعی یا همگروهی است. بخصوص روش انتخاب افراد و میزان بالای مشارکت آنها مهم است، زیرا معمولاً از یافته‌های نمونه راجع به جمعیت‌ها قیاس می‌شود. گزارش هر اقدامی که برای تشویق به مشارکت در مطالعه انجام شده، سودمند است.

کارازمایی‌های بالینی

رژیم‌های درمانی (شامل مراقبت از بیمار و معیارهای تغییر یا توقف درمان) باید به طور مفصل شرح داده شوند. روش تخصیص درمان‌ها به افراد باید بوضوح بیان شود. بویژه روش‌های خاص تقسیم تصادفی (شامل هر نوع طبقه‌بندی) و نحوهٔ انجام آن باید شرح داده شود. عدم تخصیص تصادفی به عنوان یک نقص طرح محسوب می‌شود و دلیل آن باید ذکر شود.

روش‌های کورسازی و دیگر احتیاط‌ها که برای اطمینان غیر سوگرای پاسخ بیماران انجام شده، باید توضیح داده شود. معیارهای اصلی مقایسهٔ درمان‌ها، همان گونه که در منشور کارازمایی موافقت شده، باید فهرست شود. برای کارازمایی‌های ضربدری[i] الگوی دقیق درمان‌ها و هر دورهٔ درگیر سازی[ii] یا شستشو (پاکسازی)[iii] باید توضیح داده شود. موضوع‌های اخلاقی شامل موافقت آگاهانه، رازداری و احترام به حقوق بشر باید نوشته شود.

روش‌های آماری

تمام روش‌های آماری که در مقاله استفاده شده باید مشخص شوند. اگر از چندین روش استفاده شده باید به طور قطعی روشن شود در کجا از چه روشی استفاده شده است و اگر لازم باشد در قسمت یافته‌ها این مطلب روشن شود. لازم نیست روش‌های متداول نظیر آزمون t، مجذور کای، ویلکاکسون[iv] و من‌ویتنی،[v] همبستگی (r) و

---

i - Crossover
ii - Run-in
iii - Wash out
iv- Wilcoxon
v - Mann- whitney

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

رگرسیون خطی شرح داده شوند، ولی روش‌هایی که بیش از یک شکل دارند نظیر آزمون‌های t (جفتی و غیرجفتی)، آنالیز واریانس و همبستگی رتبه‌ای باید بدون ابهام مشخص شوند. روش‌های پیچیده‌تر را باید تا حدودی شرح داد و اگر روش‌ها معمول نباشند، باید مرجع دقیقی، ترجیحاً یک کتاب درسی (با شماره صفحه)، ارایه گردد. شرح مختصری راجع به اینکه چرا از روش تحلیلی خاصی استفاده شده است نیز ممکن است سودمند باشد، بویژه اگر از روش آشناتری استفاده نشده باشد. ذکر نام برنامهٔ رایانه‌ای یا برنامهٔ آمادهٔ رایانه‌ای[i] استفاده شده، مفید است (مثلاً Epi-Info, SPSS). با این حال روش‌های آماری بکار رفته حتماً باید مشخص شوند.

قسمت یافته‌ها

اطلاعات توصیفی

توصیف مناسب داده‌ها باید قبل و همراه با تحلیل‌های آماری اصلی باشد. به طور کلی، متغیرهایی که برای اعتبار و تفسیر تحلیل‌های آماری بعدی مهم می‌باشند، باید به تفصیل شرح داده شوند. این کار را می‌توان با روش‌های نموداری نظیر نمودارهای پراکندگی یا هیستوگرام‌ها یا با استفاده از آمارهای خلاصه انجام داد. متغیرهای پیوسته (مانند وزن یا فشار خون) را می‌توان با استفاده از میانگین و انحراف معیار یا میانه و یک دامنهٔ صدک، مثلاً دامنهٔ بین چارک (صدک ۲۵ تا ۷۵) خلاصه کرد. اگر توزیع متغیرهای پیوسته نامتقارن باشد، روش دوم ارجح است. محاسبهٔ میانگین و انحراف معیار برای داده‌های کیفی رتبه‌ای (مانند مراحل بیماری از مرحلهٔ یک تا چهار) غلط است؛ بجای آن باید نسبت‌ها را گزارش نمود.

انحراف‌ها از طراحی اولیه مطالعه باید شرح داده شوند، مثلاً در کارازمایی‌های بالینی تعداد حذف‌شدگان با دلیل آن (اگر معلوم باشد) و بویژه تخصیص درمان مهم است. در بررسی‌های مشاهده‌ای، اگر میزان پاسخ اهمیت حیاتی داشته باشد، ارایه اطلاعات دربارهٔ ویژگی‌های کسانی که پاسخ نداده‌اند در مقایسه با آنهایی که شرکت کرده‌اند، با ارزش است. اگر منظور تعمیم نتایج به جمعیت خاصی باشد، معرّف بودن نمونهٔ مورد مطالعه باید بررسی شود.

مقایسهٔ توزیع ویژگی‌های گروه‌های مختلف، نظیر گروه‌های درمان در یک کارازمایی بالینی، در آغاز مطالعه مفید است. وجود تفاوت‌هایی بین گروه‌ها حتا اگر از نظر آماری معنی‌دار نباشد، واقعی هستند و باید به طور صحیح در تحلیل درنظر گرفته شوند (به تحلیل‌های پیچیده مراجعه کنید).

فرض‌های اصلی

تمام روش‌های تحلیلی نظیر آزمون‌های t، همبستگی، رگرسیون و آنالیز واریانس تا حدودی به فرض‌های خاصی دربارهٔ توزیع متغیر(های) مورد تحلیل وابسته‌اند. از نظر فنّی، این فرض‌ها عبارتند از اینکه داده‌ها «توزیع نرمال» دارند و اگر دو گروه یا بیشتر با هم مقایسه می‌شوند، بی‌ثباتی در هر دو یا چند گروه مشابه است. به طور قطعی نمی‌توان ابراز کرد تا چه حدی این فرض‌ها ممکن است نقض شده باشند بدون اینکه تحلیل بی‌اعتبار شده باشد، ولی اطلاع از اینکه توزیع به شدت چوله (نامتقارن) است یا بی‌ثباتی بین گروه‌ها به طور

i - Package

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

قابل توجهی متفاوت است، ضروری است. شاید لازم باشد قبل از تحلیل، تبدیل‌هایی صورت گیرد (به تبدیل داده‌ها مراجعه کنید)، یا از روش‌های بدون توزیع استفاده شود که به فرض توزیع بستگی ندارد (به این روش‌ها اغلب روش‌های غیر پارامتری گویند)، مثلاً آزمون U من ـ ویتنی[i] یک آزمون بدون توزیع مشابه با آزمون t دو نمونه است.

روش‌های بدون توزیع برای داده‌های کوچکی که نمی‌توان برای آنها فرض‌ها را به اندازهٔ کافی معتبر ساخت نیز مناسب می‌باشند. ممکن است فرض نرمال بویژه مهم باشد (مثلاً اگر دامنهٔ مقادیر بصورت ۹۵٪ «نرمال» یا دامنهٔ مرجع محاسبه شده باشد)، در چنین مواردی باید نشان داد که فرض توزیعی صحیح است.

آزمون‌های فرضیه

هدف اصلی از آزمون‌های فرضیه (به آنها آزمون‌های معنی‌داری نیز گفته می‌شود) ارزیابی تعداد محدودی فرضیه‌های از پیش ساخته شده می‌باشد. آزمون‌های دیگری که پس از وارسی مقدماتی داده‌ها انجام می‌شوند، گمان کاذبی ایجاد می‌کنند، زیرا در چنین شرایطی مقدار P محاسبه شده خیلی کوچک است. مثلاً آزمودن تفاوت بین کوچکترین و بزرگترین مجموعه‌ای از چندین میانگین یا نسبت بدون ارایهٔ دلیل قانع‌کننده برای آزمودن آن تفاوت خاص، معتبر نیست؛ برای مقایسهٔ جفتی بین چندین گروه روش‌های خاص «مقایسهٔ چندگانه» وجود دارد. با وجود این، هر گاه سه گروه یا بیشتر که رتبهٔ طبیعی دارند (نظیر گروه‌های سنی یا مراحل سرطان) با هم مقایسه می‌شوند، باید داده‌ها با روشی تحلیل گردند که به طور اختصاصی روند بین گروه‌ها را ارزیابی می‌کند.

معمولاً از آزمون‌های فرضیهٔ دو دامنه استفاده می‌شود. اگر از آزمون یک دامنه استفاده شود، باید مشخص گردد و دلیل آن توجیه شود. راجع به نحوهٔ ارایه و تفسیر یافته‌های آزمون‌های فرضیه بعداً بحث می‌شود. علاوه بر آزمون‌های فرضیه، استفاده از دامنهٔ اطمینان به شدت توصیه می‌شود.

دامنهٔ اطمینان

بیشتر مطالعه‌ها با براورد برخی کمیّت‌ها (نظیر اختلاف میانگین‌ها یا خطر نسبی) سر و کار دارند. محاسبهٔ دامنهٔ اطمینان حول چنین براوردهایی پسندیده است. دامنهٔ اطمینان، دامنهٔ مقادیری است که مثلاً با ۹۵٪ اطمینان شامل مقدار واقعی است. رابطهٔ نزدیکی بین یافته‌های آزمون فرضیه و دامنهٔ اطمینان همراه با آن وجود دارد. اگر تفاوت بین درمان‌ها در سطح ۵٪ معنی‌دار باشد،آنگاه ۹۵٪ دامنهٔ اطمینان همراه آن تفاوت، شامل صفر نخواهد بود. دامنهٔ اطمینان بیشتر از آزمون فرضیه اطلاعات دارد، زیرا دامنهٔ مقادیر اثر واقعی را نشان می‌دهد که با مشاهدات نمونه سازگار است (به تفسیر آزمون‌های فرضیه نیز مراجعه نمایید). دامنهٔ اطمینان دقت یک براورد را نشان می‌دهد. یک دامنهٔ اطمینان وسیع دلالت بر فقدان اطلاعات می‌کند، خواه تفاوت از نظر آماری معنی‌دار باشد یا نباشد و هشداری است علیه تفسیر بیش از حد یافته‌های مطالعه‌های کوچک.

در یک بررسی مقایسه‌ای، باید به جای یافته‌های هر گروه به طور مجزا، دامنه‌های اطمینان تفاوت‌های بین گروه‌ها گزارش شود.

---

i- Mann-Whitney u test

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

مشاهده‌های جفتی

هر گاه سنجش‌های دو گروه مختلف با هم مقایسه می‌شوند، لازم است مشخص شود که این سنجش بین مشاهده‌های فرد (مثلاً افرادی که درمان‌های مختلف دریافت می‌کنند) یا مشاهده‌های جفتی یعنی مقایسهٔ بین سنجش‌هایی که در افراد مشابه در شرایط مختلف انجام شده (نظیر قبل و پس از درمان) می‌باشد. مثلاً هر گاه داده‌ها فرد باشند، باید از آزمون t دو نمونه استفاده کرد، ولی برای داده‌های جفتی باید از آزمون t جفتی استفاده نمود. همین طور، آزمون غیر پارامتری u من ـ ویتنی برای داده‌های فرد و آزمون ویلکاکسون برای داده‌های جفتی است، و آزمون مجذور کای معمولی برای جدول‌های ۲×۲، برای داده‌های فرد و آزمون مک نمار برای داده‌های جفتی است. همیشه باید مشخص گردد از چه نوع آزمونی استفاده شده است. همین طور، روش محاسبهٔ دامنهٔ اطمینان نیز برای مشاهده‌های فرد و جفتی متفاوت است.

هرگاه سه دسته مشاهده یا بیشتر وجود داشته باشد، تمایز مشابهی باید انجام شود. تمام روش‌های آماری را که در این قسمت گفته شد می‌توان به بیش از دو گروه و بویژه، آزمون‌های t دو نمونه‌ای و جفتی را می‌توان به اشکال مختلف آنالیز و واریانس تعمیم داد.

سنجش‌های تکراری

یک طرح مطالعه ممکن است شامل ثبت سنجش‌های پی در پی متغیر(های) مشابهی در یک فرد در لحظه‌های زمانی مختلف باشد. اغلب چنین داده‌هایی با محاسبهٔ میانگین‌ها و انحراف معیارها در هر زمان تحلیل می‌شوند و در نمودار با خطی نشان داده می‌شوند که این میانگین‌ها را به هم وصل می‌کند، شکل این منحنی میانگین، ایدهٔ خوبی از شکل تک تک منحنی‌ها ارایه نمی‌کند، مگر اینکه تک تک پاسخ‌ها خیلی مشابه باشند. تحلیل برخی ویژگی‌های نمودار بخصوص ممکن است با ارزش‌تر باشد، نظیر زمانی که برای رسیدن به اوج یا طول مدت بالای یک سطح معین لازم است. این مطلب می‌تواند به اجتناب از مسایل همراه با آزمون‌های فرضیه‌های متعدد کمک کند (به آزمون‌های فرضیه‌های متعدد مراجعه کنید).

وقتی گروه‌های افراد با هم مقایسه می‌شوند، نباید سنجش‌های تکراری یک متغیر مشابه روی یک فرد تحت شرایط تجربی مشابه، که به بررسی‌های تکراری معروف است، به عنوان مشاهده‌های مستقل در نظر گرفته شوند. اگر تعداد تکرارها برای تمام افراد یکسان باشد تحلیل چندان مشکل نیست. بویژه اگر از آزمون‌های t به عنوان داده‌های غیر تکراری استفاده شود، باید از آنالیز و واریانس استفاده کرد. اگر تعداد تکرارها بین افراد متفاوت باشد، یک تحلیل کامل می‌تواند خیلی پیچیده شود. اگر تعداد مشاهده‌های بین افراد خیلی متفاوت باشد، استفاده از بزرگترین و کوچکترین سنجش از یک مجموعه سنجش‌ها (نظیر حداکثر فشارخون در حین آبستنی) می‌تواند گمراه‌کننده باشد.

تبدیل داده‌ها

بسیاری از متغیرهای زیست ـ پزشکی توزیع‌های چولهٔ مثبت با برخی مقادیر خیلی زیاد دارند. برای اینکه اطلاعات برای تحلیل مناسب گردد، ممکن است به تبدیل ریاضی نیاز باشد. در چنین شرایطی غالباً از تبدیل لگاریتمی استفاده می‌شود، هرچند گاهی ممکن است تبدیل‌های دیگر نظیر ریشهٔ دوم یا معکوس مناسب‌تر باشند.

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

مطلوب آن است که یافته‌ها پس از تحلیل برای گزارش کردن دوباره به مقیاس اصلی تبدیل گردند. در حالت عادی تبدیل لگاریتمی باید از آنتی لگاریتم (پاد لگاریتم) میانگین لگاریتم داده‌ها (که به میانگین هندسی معروف است) استفاده نمود. به هرحال، از انحراف معیار یا خطای معیار نباید آنتی لگاریتم (پاد لگاریتم) گرفت، بجای آن می‌توان از پادلگاریتم دامنهٔ اطمینان در مقیاس لگاریتمی استفاده نمود تا دامنهٔ اطمینان مقیاس اصلی را نشان دهد. اگر نمونهٔ منفردی وجود داشته باشد، می‌توان از روش مشابه به سایر تبدیل‌ها استفاده کرد، ولی برگرداندن تبدیل دامنهٔ اطمینان برای اختلاف بین میانگین‌های نمونه،تنها برای تبدیل لگاریتمی مفهوم دارد.

اگر از تبدیل استفاده شده است، بازبینی اینکه اثر مورد نظر حاصل شده است (نظیر توزیع تقریباً نرمال) اهمیت دارد. نباید فرض کرد که مثلاً تبدیل لگاریتمی لزوماً برای تمام متغیرهایی که به طور مثبت چوله هستند، مناسب است.

موارد خارج از حد (استثناها)[i]

موارد خارج از حد، مشاهده‌هایی متفاوت با بقیهٔ داده‌ها هستند. در حدی که در پژوهشگران این تردید را بوجود می‌آورند که یک خطای فاحش روی داده است یا این مقادیر به جامعهٔ دیگری تعلق دارند. مشاهده‌هایی که به شدت بی‌ثبات هستند، نباید از تحلیل حذف شوند، مگر اینکه دلیل دیگری وجود داشته باشد که نسبت به اعتبار آنها شک کنیم. حذف چنین استثناهایی باید گزارش شود. از آنجا که استثناها می‌توانند اثر مهمی روی تحلیل آماری داشته باشند، برای ارزیابی اینکه چقدر نتیجه‌گیری به این مقادیر وابسته است، تحلیل داده‌ها با و بدون چنین مشاهده‌هایی مفید است.

i - Outliers

همبستگی

ارایهٔ نمودار پراکندگی داده‌هایی که ضریب همبستگی آنها ارایه می‌شود، مطلوب است. هر چند اگر چندین متغیر وجود داشته باشد، این کار امکان‌پذیر نیست. وقتی چندین متغیر بررسی می‌شود، نمایش همبستگی‌های بین تمام جفت‌های متغیر در یک جدول (ماتریکس همبستگی) مفیدتر است تا اینکه فقط بزرگترین یا مقادیر معنی‌دار بیان شوند.

برای داده‌هایی که توزیع نامناسب دارند، بجای محاسبهٔ همبستگی معمول «حد حاصل ضرب» پیرسون (r)، می‌توان همبستگی رتبه‌ای را محاسبه کرد. همچنین برای متغیرهایی که به مقادیر خاص بالا یا پایین محدود هستند (مثلاً وزن هنگام تولد زیر ۲۵۰۰ گرم) یا برای متغیرهای طبقه‌ای می‌توان از همبستگی رتبه‌ای استفاده نمود. اگر رابطهٔ بین متغیرها خطی نباشد یا مقادیر متغیر بجای اینکه نامحدود باشند توسط پژوهشگر انتخاب شده باشند، نیز همبستگی رتبه‌ای ارجح است.

ضریب همبستگی خلاصهٔ مفیدی از میزان همبستگی خطی بین دو متغیر کمّی است، ولی یکی از روش‌های آماری است که به کرّات از آن سوءاستفاده می‌شود. در شرایط متعددی نباید از همبستگی استفاده کرد. محاسبهٔ ضریب همبستگی ساده برای داده‌هایی که بیش از یک مشاهده روی برخی یا تمام افراد دارند، غلط است، زیرا چنین مشاهده‌هایی از هم مستقل نیستند. برای مقایسهٔ دو روش اندازه‌گیری یک متغیر مشابه، همبستگی مناسب نیست، زیرا توافق را اندازه نمی‌گیرد، بلکه همبستگی را ارزیابی می‌کند. استفاده از همبستگی برای ربط دادن تغییر مقدار اولّیه با زمان می‌تواند نتایج گمراه کننده‌ای به همراه داشته باشد. محاسبهٔ

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

ضریب همبستگی برای داده‌هایی که شامل زیرگروه‌هایی است که می‌دانیم سطوح میانگین یکی یا هر دو متغیر با هم فرق دارند، می‌تواند گمراه‌کننده باشد (مثلاً ترکیب داده‌های مردان و زنان وقتی یکی از متغیرها قد باشد). تفسیر ضریب همبستگی در ذیل بحث شده است (همبستگی و علیت).

رگرسیون

نمایش خط رگرسیون مناسب همراه با نمودار پراکندگی داده‌های خام خیلی پسندیده است. ترسیم خط مناسب بدون ارایهٔ داده‌ها، اندکی بیشتر از خود معادلهٔ رگرسیون اطلاعات می‌دهد. ارایهٔ مقادیر شیب (با خطای معیار آن) و طول و عرض از مبدأ[i] و سنجشی از پراکندگی نقاط حول خط مناسب (انحراف معیار باقیمانده) مفید است. دامنهٔ اطمینان یک خط رگرسیون را می‌توان محاسبه کرد.

رگرسیون داده‌هایی که زیر گروه‌های مشخصی دارند، می‌تواند نتایج گمراه‌کننده‌ای بدهد. بویژه اگر میانگین گروه‌های متغیر وابسته (y) با هم فرق داشته باشند. با استفاده از تحلیل هموردایی (کوواریانس) نتایج قابل اطمینان‌تری را می‌توان بدست آورد. رگرسیون و همبستگی روش‌های متفاوتی هستند که برای منظورهای مختلفی بکار می‌روند و لزومی ندارد به طور خودکار با یکدیگر همراه باشند. تفسیر تحلیل رگرسیون در ادامه بحث شده است (پیشگویی و آزمون‌های تشخیصی).

داده‌های بقا

گزارش داده‌های بقا باید شامل نمایش نموداری یا جدولی، جدول عمر، با شرح کامل تعداد بیماران در خطر (مثلاً در خطر مرگ) در زمان‌های پیگیری متفاوت باشد. جدول عمر یا منحنی بقای احصایی به طور مؤثر با زمان‌های بقای «سایش شده» سر و کار دارد که وقتی بیمار در حین پیگیری گم می‌شود یا هنوز زنده است، مطرح می‌شود و معلوم است زمان بقای آنها حداقل چند روز است. وقتی فقط چند نفر (مثلاً ۵ نفر) در خطر هستند، برای پرهیز از تفسیر غلط قسمت دوم منحنی که قابل اطمینان نیست، شاید حذف منحنی بقا مفید باشد. از آنجا که توزیع زمان‌های بقا معمولاً به طور مثبت چوله است، در صورت وجود اطلاعات سایش شده (سانسور شده) محاسبهٔ میانگین زمان بقا روا نمی‌باشد.

مقایسهٔ گروه‌های درمانی که در زمان ثابت قراردادی زنده هستند، می‌تواند گمراه‌کننده باشد و معمولاً کمتر از مقایسهٔ جدول‌های عمر با روش‌هایی نظیر رتبهٔ لگاریتمی[ii] کارا می‌باشد.

وقتی تعداد موارد مرگ کافی باشد، با رسم نسبت کسانی که در آغاز هر فاصلهٔ زمانی زنده‌اند و در طی آن فاصله مرده‌اند، می‌توان نشان داد چگونه خطر مرگ با زمان تغییر می‌کند. با استفاده از مدل‌های رگرسیون مناسب داده‌های بقا می‌توان عوامل بیماری را که ممکن است روی پیش‌آگهی مؤثر باشند، تطبیق نمود (به قسمت بعد مراجعه کنید). مقایسهٔ بقای افرادی که به درمان پاسخ داده‌اند و آنهایی که پاسخ نداده‌اند، گمراه کننده است و هرگز نباید انجام شود.

تحلیل‌های پیچیده

در بسیاری از مطالعه‌ها ممکن است مشاهده‌های مورد نظر، تحت تأثیر متغیرهای متعدد دیگر قرار گیرند. این متغیرها می‌توانند هر چیزی باشند که بین افراد متفاوت هستند و می‌توانند روی پیامد

i - Intercept

ii - Log rank

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

تحت مطالعه اثر کنند. مثلاً در کارازمایی‌های بالینی ممکن است شامل ویژگی‌های بیمار یا علایم و نشانه‌ها باشند. با روش‌های رگرسیون چند گانهٔ مناسب می‌توان تمام یا برخی از هم‌متغیرها[i] را با هم ترکیب کرد تا متغیر پیامد یا پیشگویی را شرح داد، خواه این متغیر، پیوسته (مانند فشارخون)، کیفی (مانند ترمبوز پس از عمل جراحی) یا طول بقا باشد. حتا در کارازمایی‌های تصادفی شده نیز لازم است پژوهشگران مطمئن شوند که پس از تطبیق خودبخودی برای چندین عامل خطر، هنوز اثر درمان وجود دارد.

در استفاده از روش‌های چند متغیره برای روبرو شدن همزمان با بیش از یک متغیر پیامد (وابسته)، واقعاً به کمک افراد خبره نیاز است و ورای بحث این رهنمود است. هر روش آماری پیچیده را باید به نحوی ارایه کرد که برای خواننده قابل فهم باشد.

قسمت یافته‌ها: ارایهٔ یافته‌ها

ارایهٔ آمار خلاصه

میانگین‌ها را نباید بدون سنجش پراکندگی یا دقت بیان کرد. برای نشان دادن پراکندگی باید از انحراف معیار و برای نمایش دقت میانگین نمونه از خطای معیار میانگین استفاده نمود. باید روشن کرد کدامیک ارایه می‌شود. استفاده از علامت ± برای چسبانیدن خطا یا انحراف معیار به میانگین (نظیر ۱/۹±۱۴/۲) موجب سردرگمی می‌شود و باید از آن پرهیز کرد. در حال حاضر، چندین مجلهٔ پزشکی اجازهٔ استفاده از این علامت را نمی‌دهند. ترجیح داده می‌شود میانگین‌ها مثلاً به صورت «میانگین (خطای معیار)» (۱/۹) ۱۴/۲ یا «میانگین (انحراف معیار)» (۷/۴) ۱۴/۲ ارایه گردد. دامنه‌های اطمینان روش خوبی برای ابراز معقول عدم اطمینان از میانگین‌های نمونه، نسبت‌ها و دیگر آمارها است. مثلاً ۹۵٪ دامنهٔ اطمینان از میانگین واقعی تقریباً از دو خطای معیار زیر میانگین مشاهده شده تا دو خطای معیار بالای آن است. دامنه‌های اطمینان بجای استفاده از علامت ± به طور روشن‌تر بصورت ۱۰/۴ تا ۱۸ ارایه می‌شوند.

هرگاه مقایسه‌های جفتی انجام می‌گیرد (مانند وقتی که از آزمون‌های t جفتی استفاده می‌شود) ارایه میانگین و انحراف معیار تفاوت بین مشاهده‌ها یا خطای معیار اختلاف میانگین‌ها (هر کدام مناسب‌تر است) اهمیت دارد.

برای داده‌هایی که با روش‌های بدون توزیع تجزیه و تحلیل گردیده‌اند، ارایهٔ میانه و یک دامنهٔ مرکزی، مثلاً ۹۵٪ مشاهده‌ها، از استفاده از میانگین و انحراف معیار مناسبتر است (به اطلاعات توصیفی مراجعه کنید). دامنه‌های اطمینان غیرپارامتری را نیز می‌توان محاسبه نمود. همین طور، اگر داده‌های تبدیل شده را تجزیه و تحلیل کرده‌ایم، میانگین و انحراف معیار داده‌های خام احتمالاً سنجش‌های خوبی برای گرایش‌های مرکزی و پراکندگی داده‌ها نمی‌باشند و نباید ارایه گردند.

اگر درصدها ارایه می‌شوند، همیشه باید مخرج کسر را مشخص کرد. برای نمونه‌های کوچک درصدها چندان کمکی نمی‌کنند. اگر درصدها با هم مقایسه می‌شوند، فرق بین یک اختلاف مطلق از یک اختلاف نسبی اهمیت دارد. مثلاً یک کاهش از ۲۵ درصد به ۲۰ درصد را می‌توان به صورت ۵٪ یا ۲۰ ٪ کاهش بیان کرد.

---

i- Covariates

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

یافته‌های افراد

دامنه نشانگر خوبی از پراکندگی یک مجموعه مشاهده‌ها نمی‌باشد، زیرا قویاً تحت تأثیر یک مقدار فوق‌العاده می‌باشد و با اندازهٔ نمونه افزایش می‌یابد. اگر توزیع داده‌ها نسبتاً طبیعی (نرمال) باشد، ۹۵٪ مشاهده‌ها در فاصلهٔ دو انحراف معیار در هر طرف میانگین قرار دارند، ولی برای دیگر توزیع‌ها از دامنهٔ صدک بیشتر استفاده می‌شود (به اطلاعات توصیفی مراجعه کنید).

هر چند تحلیل آماری با متوسط اثرات سر و کار دارد، در بسیاری از شرایط در نظر گرفتن نحوهٔ پاسخ افراد نیز حایز اهمیت است. مثلاً اینکه بدانیم چند بیمار با درمان بهبود نیافته‌اند و نیز متوسط منفعت آن اغلب از نظر بالینی خیلی مهم است. یک متوسط اثر را نباید به صورت اعمال به تمام افراد تفسیر کرد (به سنجش‌های تکراری نیز مراجعه کنید).

ارایهٔ یافته‌های آزمون‌های فرضیه

آزمون‌های فرضیه مقادیر آزمون‌های آماری مشاهده شده را بدست می‌دهند، که با مقادیر جدول‌بندی شدهٔ توزیع مناسب (طبیعی (نرمال)، t، مجذر کای و غیره) جهت استخراج مقادیر P همراه با آن، مقایسه شده‌اند. ارایه مقادیر مشاهده شدهٔ آزمون‌های آماری پسندیده‌تر از ارایهٔ تنها مقادیر P است. نتایج کمّی مورد آزمون نظیر مقادیر میانگین‌ها و نسبت‌ها، ضریب‌های همبستگی باید ارایه گردند، خواه آزمون معنی‌دار باشد یا نباشد. باید دقیقاً مشخص نمود کدام داده‌ها تحلیل شده‌اند. اگر برای نشان دادن سطح احتمال از علایمی نظیر ستاره‌ها استفاده می‌شود، باید این علایم تعریف شوند و بهتر است در سراسر مقاله مشابه باشند.

مقادیر P به طور قراردادی به صورت P<۰/۰۵، <۰/۰۱، یا <۰/۰۰۱ ارایه می‌شوند، ولی بجز آشنایی با استفاده از این مقادیر خاص دلیلی برای استفاده از آنها وجود ندارد. مقادیر دقیق P (نه بیش از دو رقم معنی‌دار) نظیر P = ۰/۱۸ یا ۰/۰۳ مفیدتر است. بعید است لازم باشد سطوح پی کمتر از ۰/۰۰۰۱ مشخص گردد. نامیدن هر مقدار با P>۰/۰۵ به صورت «غیر معنی‌دار» توصیه نمی‌شود، زیرا ممکن است یافته‌هایی را که کاملاً از نظر آماری معنی‌دار نیستند، محو کند یا اثر واقعی را نشان ندهد (به تفسیر آزمون‌های فرضیه مراجعه کنید). هر گاه مقادیر P ارایه می‌شود تمایز > (کمتر از) از < (بیشتر از) اهمیت دارد. مقادیر P بین دو محدوده باید به ترتیبی منطقی ارایه شود، مثلاً ۰/۰۱ < P<۰/۰۵ به طوری که P بین ۰/۰۱ و ۰/۰۵ باشد. لازم نیست مقادیر P که در جدول‌ها ارایه می‌شود در متن تکرار شود. تفسیر آزمون‌های فرضیه و مقادیر P در ذیل بحث شده است (تفسیر آزمون‌های فرضیه).

شکل‌ها (نمایش نموداری)

نمایش نموداری یافته‌ها به خوانندگان کمک می‌کند و شکل‌هایی که مشاهده‌های خاصی را نشان می‌دهند، توصیه می‌شوند. نقاط روی یک نمودار که به فرد مشابهی در شرایط متفاوت مربوط است یا علایمی که برای نشان دادن نقاط مربوطه بکار می‌روند، ترجیحاً باید به هم وصل شوند. راه مفید دیگر این است که تفاوت بین شرایط را برای هر فرد رسم کنیم.

جدول‌ها

خیلی ساده‌تر می‌توان یافته‌های عددی را بجای سطرها در ستون‌ها ملاحظه نمود، و بهتر است

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

انواع مختلفی از اطلاعات (مانند میانگین‌ها و انحراف معیارها) را در ستون‌های مجزا نوشت. در یک جدول باید تعداد مشاهده‌های هر یافته را ارایه نمود. اگر سطرهای جدول‌هایی که اطلاعاتی دربارهٔ بیماران خاص، نواحی جغرافیایی و غیره می‌دهند. بر حسب سطح یکی از متغیرهای ارایه شده مرتب شده باشند، آسانتر خوانده می‌شوند.

دقت عددی

دقت کاذب به ارزش مقاله نمی‌افزاید و حتا از قابلیت خواندن و اعتبار آن می‌کاهد. نتایجی که از یک رایانه یا ماشین حساب به دست می‌آیند، معمولاً باید گرد (سر راست) شوند. هر گاه میانگین‌ها، انحراف معیارها و دیگر آمارها ارایه می‌شوند، نویسنده باید دقت داده‌های اصلی را در نظر بگیرد. معمولاً میانگین‌ها را نباید با بیش از یک اعشار بیش از داده‌های خام نشان داد، ولی ممکن است لازم باشد انحراف معیارها یا خطاهای معیار با یک اعشار بیشتر ارایه شوند. بندرت لازم می‌شود درصدها با بیش از یک اعشار نشان داده شود و حتا غالباً یک اعشار هم لازم نیست. با نمونه‌های کمتر از ۱۰۰، استفاده از اعشار دقّت را تضمین نمی‌کند و باید از آن پرهیز کرد. شایان ذکر است که این توصیه‌ها تنها برای ارایهٔ یافته‌ها است (اعداد نباید قبل یا در حین تجزیه و تحلیل گرد شوند). ارایهٔ مقادیر t، مجذور کای و r تا دو رقم اعشار کافی است.

سایر واژه‌های فنی

تعریف تمام واژه‌های آماری در اینجا غیرممکن است. توصیه‌های زیر راجع به واژه‌هایی است که غالباً به طور غلط یا گیج کننده استفاده می‌شوند.

از واژهٔ **همبستگی** نباید به عنوان یک واژهٔ کلی برای توصیف هر رابطه‌ای استفاده شود. سنجش همبستگی معنی فنی خاص دارد، که این معنی باید در کار آماری در نظر گرفته شود.

از واژهٔ **بروز** باید برای توصیف میزان رویداد موارد جدید یک خاصه در نمونه یا جمعیت مورد بررسی استفاده شود، نظیر تعداد موارد جدید سرطان در یک سال. سهمی از یک نمونه که در حال حاضر یک ویژگی را دارد **شیوع** است.

واژهٔ **غیرپارامتری** به تحلیل‌های آماری معینی اطلاق می‌شود که خصوصیت خود مشاهده‌ها نیست، نظیر آزمون‌های من ـ ویتنی.

از واژهٔ **پارامتر** نباید بجای «متغیر» برای اشاره به سنجش یا صفتی که مشاهده‌ها روی آن انجام می‌شوند، استفاده کرد. پارامترها خصوصیت‌های توزیع‌ها یا روابط در جمعیت هستند که با تحلیل آماری نمونه‌ای از مشاهده‌ها براورد می‌شوند.

صدک‌ها: وقتی دامنهٔ مقادیر یک متغیر به گروه‌های مساوی تقسیم شود، مرزهای آن میانه، سه یک، چارک، پنجک و غیره می‌باشند، به خود گروه‌ها باید نصف، ثلث، ربع، یک پنجم و غیره اطلاق شود.

حساسیت قدرت یک آزمون برای مشخص کردن یک بیماری است که واقعاً وجود دارد (یعنی نسبت موارد مثبت در بیماران).

ویژگی قدرت یک آزمون برای شناسایی عدم بیماری است وقتی که واقعاً بیماری وجود ندارد (یعنی نسبت موارد منفی در غیر بیماران). به آزمون‌های تشخیصی و ارزش اخباری نیز مراجعه کنید.

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

قسمت بحث: تفسیر

تفسیر آزمون‌های فرضیه

یک آزمون فرضیه، با احتمال P، مقبولیت داده‌های مشاهده شده را ارزیابی می‌کند. اگر «فرضیهٔ صفر» (برای مثال: اختلافی بین گروه‌ها وجود ندارد)، صحیح باشد مقدار P احتمال آن است که داده‌های مشاهده شده یا یک پیامد افراطیتر، تصادفی (شانسی) روی داده باشد (یعنی، تنها ناشی از بی‌ثباتی نمونه‌گیری باشد). اگر P بزرگ باشد، داده‌ها به طور قابل قبول با فرضیهٔ صفر سازگارند، که در نتیجه نمی‌تواند رد شود. بنابراین، P احتمال فقدان اثر واقعی نمی‌باشد. حتا اگر یک اثر حقیقی بزرگ باشد، اگر تعداد مشاهده‌ها کم باشند یک یافتهٔ غیر معنی‌دار کاملاً محتمل است. بر عکس اگر اندازهٔ نمونه خیلی بزرگ باشد و تنها اثر واقعی کوچکی وجود داشته باشد، یک یافتهٔ از نظر آماری معنی‌دار ممکن است روی دهد. بنابراین، معنی‌داری آماری را نباید مترادف با اهمیت بالینی تلقی کرد.

تفسیر یافته‌های آزمون‌های فرضیه عمدتاً از مطالب فوق پیروی می‌کنند. یک یافتهٔ معنی‌دار لزوماً دلالت بر یک اثر واقعی نمی‌کند و همیشه خطر یک یافتهٔ مثبت کاذب وجود دارد. این خطر برای مقادیر P کوچکتر کاهش می‌یابد. به علاوه، یک یافتهٔ غیر معنی‌دار (به طور قراردادی P>۰/۰۵) بدین معنی نیست که اثری وجود ندارد، بلکه داده‌ها تنها با اینکه اثری وجود نداشته باشد، سازگارند. در تفسیر مقادیر P قدری انعطاف پذیری پسندیده است. سطح ۰/۰۵ مرز راحتی است، ولی مقادیر P ۰/۰۴ و ۰/۰۶ که چندان با هم فرقی ندارند، بجای اینکه کاملاً متفاوت باشند، باید یکسان تفسیر گردند. تخصیص هر یافته با P>۰/۰۵ به عنوان غیر معنی‌دار ممکن است خواننده (و نویسندگان) را گمراه کند، از این رو پیشنهاد می‌شود (در ارایهٔ یافته‌های آزمون‌های فرضیه) مقادیر واقعی P نقل شود.

در تفسیر اطلاعات بویژه در مطالعه‌های کوچک دامنه‌های اطمینان فوق‌العاده مفیدند، زیرا میزان ابهام در رابطه با یافته‌ها (نظیر اختلاف بین دو میانگین) را نشان می‌دهند، خواه اختلاف از نظر آماری معنی‌دار باشد یا نباشد. بخصوص استفاده از دامنهٔ اطمینان همراه با یافته‌های غیر معنی‌دار ممکن است روشنگر باشد.

چندین آزمون فرضیه

در بسیاری از طرح‌های پژوهشی، در آغاز پژوهش تعدادی آزمون‌های فرضیه برای مقایسه‌های مهم در نظر گرفته می‌شود. آزمون‌های فرضیه که از قبل راجع به آنها تصمیم گرفته نشده مکمل آنها هستند، بویژه اگر یافته‌ها آنها را پیشنهاد کرده باشند. تمایز بین این دو نوع آزمون فرضیه اهمیت دارد و به آزمون‌های فرضیه که از قبل تنظیم شده‌اند اهمیت خیلی بیشتری داده می‌شود. دیگر آزمون‌ها باید تنها به عنوان آزمون‌های اکتشافی در نظر گرفته شوند (برای ایجاد فرضیه‌های جدید که با مطالعه‌های بیشتر بررسی شوند). یک دلیل این کار این است که وقتی در تحلیل یک مطالعه تعداد بسیار زیادی آزمون‌های فرضیه انجام می‌شود، احتمالاً زیر گروه‌های زیادی با هم مقایسه می‌شوند که در این حالت می‌توان انتظار داشت تعداد یافته‌های مثبت کاذب که ناشی از شانس می‌باشند، مطرح گردد که می‌توانند در تفسیر یافته‌ها مسایل قابل توجهی ایجاد کند. به طور کلی، هر قدر آزمون‌های بیشتری انجام شود، احتمال اینکه یافته‌های معنی‌دار حاصل گردد بیشتر می‌شود، ولی احتمال تعدد یافته‌های مثبت کاذب نیز

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

افزایش می‌یابد. یک راه برای کاهش یافته‌های مثبت کاذب این است که سطح P کوچکتری را به عنوان معیار معنی‌داری آماری در نظر بگیریم. مشکل پیچیده‌تر وقتی مطرح می‌شود که آزمون‌های معنی‌داری روی داده‌های وابسته به هم (مرتبط) انجام شود. مثالی از این نوع داده‌ها در تحلیل داده‌های پشت سر هم (سریال) (سنجش تکراری) است، یعنی وقتی که آزمون مشابهی روی داده‌های متغیر یکسانی که از افراد مشابه در زمان‌های گوناگون گرداوری شده انجام می‌شود.

مشکل دیگر این است که روی دو یا چند متغیر مرتبط با هم تحلیل‌های مجزا به طوری انجام شود که اگر آنها مستقل از هم بودند انجام می‌شد. هر اثباتی وزن مدارک را زیاد افزایش نمی‌دهد، زیرا آزمون‌ها به داده‌های مشابهی مربوط بوده‌اند. مثلاً فشارخون‌های سیستولی و دیاستولی، مثل بسیاری از راههای دیگر ارزیابی پاسخ کلی بیمار، خیلی شبیه بهم عمل می‌کنند. در چنین مواردی تفسیر بسیار دقیق نتایج ضروری است.

همبستگی و علیت

یک همبستگی که از نظر آماری معنی‌دار است (که از تحلیل همبستگی یا مجذور کای حاصل شده است) بخودی خود دلیل مستقیمی دال بر یک رابطهٔ علیتی بین متغیرهای مورد نظر نمی‌باشد. در مطالعه‌های مشاهده‌ای علیّت را تنها بر اساس روش‌های غیرآماری می‌توان برقرار کرد، در کارازمایی‌های تصادفی شده استنتاج علیتی آسانتر است. در مقایسهٔ متغیرهایی که هر دو با زمان تغییر می‌کنند باید خیلی دقت کرد، زیرا به سادگی باعث همبستگی‌های ظاهری و ساختگی می‌شوند.

پیشگویی و آزمون‌های تشخیصی

حتا وقتی تحلیل رگرسیون، رابطه‌ای از نظر آماری معنی‌دار بین دو متغیر نشان می‌دهد، اگر از معادلهٔ رگرسیون برای پیشگویی سطح عددی یک متغیر (Y) از متغیر دیگر (X) برای موارد خاصی استفاده شود، ممکن است بی‌دقتی قابل توجهی وجود داشته باشد. درستی چنین پیشگویی‌هایی را نمی‌توان از ضریب همبستگی یا رگرسیون ارزیابی کرد. بلکه باید دامنهٔ پیشگویی براورد شدهٔ مقدار Y در پاسخ به مقدار X خاص محاسبه شود. از خط رگرسیون تنها برای پیشگویی متغیر Y از متغیر X استفاده می‌شود و برعکس آن صادق نمی‌باشد.

یک آزمون تشخیصی با حساسیت و ویژگی زیاد لزوماً آزمون مفیدی برای مقاصد تشخیصی نیست، بویژه اگر در جمعیتی بکار رود که شیوع بیماری در آن خیلی کم باشد.

در چنین شرایطی محاسبهٔ نسبت افراد با نتایج آزمون مثبت که واقعاً بیمارند (که به آن ارزش اخباری مثبت[i] می‌گویند) مفید است. شایان ذکر است، توافق کلی دربارهٔ تعریف «میزان مثبت کاذب» یا «میزان منفی کاذب» وجود ندارد و همیشه باید دقیقاً مشخص کرد چه چیزی محاسبه گردیده است، و این کار را به بهترین نحو می‌توان با یک جدول ۲×۲ نشان داد که نتایج آزمون را با وضع واقعی بیماری بیماران ربط می‌دهد.

مسألهٔ تشخیصی مشابهی با متغیرهای پیوسته مطرح می‌شود. طبقه‌بندی مقادیر خارج از «دامنهٔ طبیعی» به عنوان «غیرطبیعی» برای یک متغیر معمول است، ولی اگر شیوع ناهنجاری واقعی کم باشد، بیشتر مقادیر خارج از دامنهٔ طبیعی، طبیعی

i- Positive predictive value

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]

خواهند بود. تعریف غیرطبیعی باید بر اساس هر دو معیار بالینی و آماری باشد.

نقاط ضعف

اگر کسی از نقاط ضعف پژوهش خود آگاه است، بهتر است نقاط ضعف طرح و اجرای پژوهش را مشخص کند و اثرات احتمالی آن را روی نتایج و تفسیرها در نظر بگیرد تا اینکه از آنها چشم‌پوشی کند به این امید که کسی متوجه آنها نشود.

نتیجه‌گیری

هدف از روش‌های آماری، ارایهٔ واقعیت‌های دلایل علمی ساده است که از یک کار پژوهشی بدست می‌آید. کسب مهارت‌ها و تجارب لازم برای طراحی مطالعه‌های صحیح، انجام تحلیل‌های آماری مستدل و ارایهٔ یافته‌ها به روشی روشن و هدفدار ساده نیست. امیدوارم این رهنمودها بتواند تحلیل‌های آماری را که در نشریات پزشکی گزارش می‌شوند بهبود بخشد.

سپاسگزاری

نگارنده از آقایان دکتر علی صادقی حسن‌آبادی، دکتر نصراله بشردوست، دکترمهدی تذهیبی، سرکار خانم دکتر زهرا محمدزاده و خانم دکتر مطهره ژیانپور به خاطر توصیه‌های ارزنده‌شان سپاسگزاری می‌کند.

References

۸. جانقربانی م. ارزیابی جنبه‌های طراحی، اجرا، تحلیل آماری و ارایهٔ مقاله‌های پزشکی. پژوهش در پزشکی سال ۲۱ شماره ۳، مهر ـ آذر ۱۳۷۶. ص ۱۸.

۹. هال جورج ام. ساختار یک مقالهٔ علمی. ترجمهٔ علیخانی کتایون مجلهٔ دانشکده پزشکی تهران ویژه‌نامهٔ فروردین ۱۳۷۸.

۱۰. عزیزی ف. چگونه می‌توان پژوهش‌های بالینی را انجام داد؟ ضمیمهٔ مجلهٔ دانشکده پزشکی دانشگاه علوم پزشکی شهید بهشتی، اسفند ۱۳۶۸.

۱۱. عزیزی ف. روش‌های یادگیری و تحقیق در علوم پزشکی. نشر جوان تهران، ۱۳۷۱.

1. Altman DG. Statistics in medical journals. Statistics in Medicine 1982; 1:59-71.
2. O'Fallon JR, Dubey SB, Salsburg DS, et al. Should there be statistical guidelines for medical research papers? Biometrics 1978; 34:687-95.
3. Bailar JC, Mosteller F. Guidelines for statistical reporting in articles for medical journals: amplifications and explanations. Ann Intern Med 1988; 108:266-73.
4. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. New Engl J Med 1997; 336:309-15.
5. Simon R, Wittes RE. Methodologic guidelines for reports of clinical trials. Canaer Treat Rep 1985; 69:1-3.
6. Joint Committee on Governmental Affairs of the Epidemiology Section (American Public Health Association) and the Society for Epidemiologic Research. The interagency regulatory liaison group "Guidelines for documentation of epidemiologic studies". Am J Epidemiol 1981; 114:614-8.
7. Lichtenstein MJ, Mulrow CD, Elwood PC. Guidelines for reading case-control studies. J Chron Dis 1987; 40:893-903.
12. Armitage P, Berry G. Statistical methods in medical research. 2nd ed. Oxford: Blackwell, 1987.
13. Colton T. Statistics in medicine. Boston: Little, Brown and Co, 1974.
14. Bland M. An introduction to medical statistics. Oxford University Press, 1987.
15. Bradford Hill A. A short textbook of medical statistics. 11th ed. London, Hodder and Stoughton, 1984.

[Downloaded from ijem.sbmu.ac.ir on 2026-11-15]