

راهنمای انتخاب آزمون‌های آماری صحیح در پروپوزال‌ها و مقالات علوم پزشکی

محمد رضا بانسی^{۱*}، فرزانه ذوالعلی^۲، الهام محبی^۳

۱. دکترای آمارزیستی، دانشیار، مرکز تحقیقات مدل‌سازی در سلامت، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
۲. دکترای تخصصی اپیدمیولوژی، استادیار، مرکز منطقه‌ای آموزش نظام مراقبت HIV/ایدز- مرکز همکار سازمان بهداشت جهانی، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
۳. دامپزشک، کارشناس ارشد بهداشت عمومی، مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران

● دریافت مقاله: ۹۲/۹/۲۶ آخرین اصلاح مقاله: ۹۲/۱۱/۲۱ ● پذیرش مقاله: ۹۲/۱۱/۲۸

زمینه و هدف: هدف اصلی انجام پژوهش در حوزه علوم پزشکی، پاسخ به سؤال یا مشکلی است که محقق می‌خواهد با استفاده از آن سلامت افراد جامعه را ارتقا بخشد. هدف اولیه رسیدن به پاسخ صحیح و با کمترین خطا از دغدغه‌های مهم محققین می‌باشد و هدف ثانویه نشر نتایج حاصل از این پژوهش‌ها برای تعمیم به جامعه و استفاده از آن‌ها در بعد وسیع‌تر است. رسیدن به این اهداف و سایر اهداف جانبی مستلزم به کارگیری علم آمار زیستی در روند انجام پژوهش می‌باشد. با وجود اهمیت بسیار زیاد آمار در تحقیقات، پژوهشگران با این علم آشنایی کمی دارند و یا به علت برخی پیچیدگی‌ها از آن دوری می‌کنند. آمار در سطوح متفاوت تحقیق از جمله نوشتن پروپوزال و تفسیر نتایج مقالات کمک‌کننده و حضور متخصصین آمار و اپیدمیولوژی در آماده‌سازی مقالات جهت چاپ بسیار مهم است. در مقاله حاضر سعی شد با زبانی شیوا و با ذکر مثال‌هایی کاربردی، مفاهیم برخی از آزمون‌های کاربردی در حوزه علوم پزشکی در اختیار محققان قرار گیرد.

کلید واژه‌ها: انتخاب آزمون آماری، آزمون‌های پارامتری، آزمون‌های ناپارامتری

*نویسنده مسؤول: مرکز تحقیقات مدل‌سازی در سلامت، معاونت تحقیقات و فناوری، چهارراه سمیه، خیابان جهاد، دانشگاه علوم پزشکی کرمان، کرمان، ایران

● Email: rbaneshi@yahoo.com

● تلفن: ۰۳۴۱-۲۲۶۳۷۲۵ ● شماره: ۰۳۴۱-۲۲۶۳۷۲۵

مقدمه

تحلیل داده‌ها در فرایند هر پژوهش علمی به ویژه حوزه علوم پزشکی از اهمیت زیادی برخوردار است. از طرف دیگر، با گسترش روزافزون پزشکی مبتنی بر شواهد، تصمیم‌گیری بر اساس نتایج ارایه شده برای استفاده بالینی یکی از چالش‌های پزشکان محسوب می‌شود. ارزیابی نتایج علاوه بر تسلط بیولوژیک بر بیماری یا مسأله موردنظر، نیازمند داشتن قدرت تحلیل اطلاعات با استفاده از پیش‌زمینه آماری است. توجه به نوع آزمون به کار برده شده و قضاوت در مورد آن یکی از چالش‌ها و نیازهای محققان و به ویژه محققان بالینی می‌باشد، اما با توجه به مفاهیم و پیش‌فرض‌های آزمون‌های آماری (که در برخی موارد پیچیده است)، تلاش برای یادگیری و دستیابی به دانش آمار با استقبال کمی روبه‌رو شده است (۱-۳).

در حال حاضر اشکالات زیادی در قسمت تحلیل آماری طرح‌های پژوهشی دانشگاه‌های علوم پزشکی مشاهده می‌شود؛ بدین معنی که تناسبی بین اهداف مطالعه و آزمون انتخابی وجود ندارد. این اشکال حتی در بعضی مقالات به چاپ رسیده نیز مشاهده می‌گردد.

گرچه طیف وسیعی از آزمون‌ها و تحلیل‌های آماری در حوزه علوم پزشکی وجود دارد، در مقاله حاضر سعی شد پرکاربردترین و اصولی‌ترین آزمون‌های آماری با زبانی ساده و شیوا بر اساس نوع داده‌ها و نوع مطالعات برای خوانندگان ارایه گردد.

با توجه به توضیحات فوق، در مقاله حاضر ابتدا به عوامل مهم در انتخاب آزمون پرداخته و سپس توضیح داده شد که برای پاسخ به چه هدفی از کدام آزمون آماری استفاده شود. در این مقاله فقط روی اصلی‌ترین آزمون‌های آماری تمرکز گردید و از جزئیات متدولوژیک و پیش‌فرض‌های هر کدام از آزمون‌ها صحبت نشد، بلکه فقط هدف آن بود که محققین بتوانند به صورت تقریبی روش آماری صحیحی را برای تحلیل داده‌های خود مشخص نمایند.

عوامل مؤثر بر انتخاب آزمون مناسب

انتخاب آزمون آماری مناسب به عوامل مهمی مانند تعیین نوع متغیر وابسته یا پیامد (کمی یا کیفی)، تعداد گروه‌های مورد مقایسه (مقایسه دو گروه یا بیشتر از دو گروه)، استفاده از روش‌های همسان‌سازی گروه‌های مورد مقایسه (به عبارت دیگر مستقل و یا وابسته بودن مشاهدات) و نرمال بودن توزیع داده‌ها بستگی دارد (۴، ۵).

آزمون‌های آماری کاربردی

۱- آزمون‌های اصلی (مقایسه میانگین و یا نسبت در دو جامعه): دو آزمون آماری بسیار کاربردی در حوزه علوم پزشکی Independent two-sample t و آزمون χ^2 می‌باشد (۵-۷).

آزمون Independent two-sample t برای مقایسه میانگین متغیرهای کمی (عددی) در دو گروه مستقل استفاده می‌شود (۸، ۹). به عنوان مثال به منظور بررسی ارتباط مصرف آسپیرین و مدت زمان لخته شدن خون در گروهی از افراد مبتلا به افزایش چربی خون مطالعه‌ای انجام شده است. در طراحی مطالعه به گروهی از بیماران قرص آسپیرین داده و گروه دیگری از بیماران به عنوان گروه شاهد در نظر گرفته شدند که هیچ درمانی دریافت نکردند. با توجه به این‌که مقایسه متغیر کمی (مدت زمان لخته) بین دو گروه مستقل (گروه مداخله و شاهد) انجام می‌شود، باید آزمون Independent two-sample t استفاده گردد.

در آزمون‌هایی که به مقایسه میانگین‌ها می‌پردازند (مانند Paired t، Independent two-sample t، Variance analysis و...)، لازم است توزیع متغیر وابسته (مدت زمان لخته شدن خون) از توزیع طبیعی یا نرمال پیروی کند. اگر این فرض برقرار نبود از روش‌های دیگری در تحلیل داده‌ها استفاده می‌شود که به نام روش‌های ناپارامتری مشهور هستند. این امر بدین معنی است که برای آزمون Independent two-sample t یک آزمون ناپارامتری وجود دارد که وابسته به فرض نرمال

بودن داده‌ها نیست و این آزمون Mann-Whitney می‌باشد (۱۰-۱۲).

از آزمون χ^2 برای مقایسه نسبت‌ها و یا شیوع در دو گروه مستقل استفاده می‌شود. به طور مثال محقق‌ی تمایل دارد نسبت افسردگی را در افرادی که امگا-۳ مصرف می‌کنند و گروهی که امگا-۳ مصرف نمی‌کنند با یکدیگر مقایسه کند. با توجه به این‌که افسردگی متغیری کیفی است (افسردگی دارد/ ندارد) و دو جامعه (مصرف و عدم مصرف امگا-۳) در سؤال پژوهش تعریف شده‌اند، آزمون χ^2 مناسب می‌باشد. با توجه به این‌که این آزمون با میانگین سر و کار ندارد، نرمال بودن توزیع متغیر وابسته مفهومی نداشته و آزمون ناپارامتری معادلی برای آزمون χ^2 وجود ندارد، اما پیشنهاد می‌شود در مواردی که حجم نمونه کافی نیست (کمتر از ۲۰ نمونه) از آزمون Fisher's exact استفاده گردد (۱۳).

به هر حال توجه به این نکته ضروری است که گروه‌ها در صورتی قابل مقایسه هستند که از نظر متغیرهای مخدوش کننده (Confounders) شبیه باشند. به عنوان مثال، محقق‌ی از اعضای هیأت علمی پرسیده است «آیا در سال گذشته تجربه تصادف داشته‌اید؟» (بلی/ خیر). حال او مایل است به بررسی فراوانی وقوع تصادف در خانم‌ها و آقایان بپردازد. روش تحلیل با توجه به مطالب گفته شده، آزمون χ^2 می‌باشد. به فرض آن‌که در نتیجه آزمون مردان بیشتر تصادف کرده باشند، سؤالی که پیش می‌آید این است که آیا مردان به دلیل مرد بودنشان بیشتر تصادف می‌کنند یا عوامل دیگری در میزان تصادف آنان مؤثر است؟ آیا قبول دارید که اگر تعداد ساعات رانندگی افزایش یابد، احتمال تصادف کردن هم بیشتر می‌شود؟ و آیا قبول دارید مردان بیش از زنان رانندگی می‌کنند؟ پس شاید تعداد ساعات رانندگی با تصادف کردن ارتباط دارد و نه جنسیت فرد.

به بیان دیگر، اگر دو گروه هیأت علمی مرد و زن انتخاب شوند؛ به طوری که متوسط ساعات رانندگی‌شان در طول هفته برابر باشد (و البته از نظر عوامل مخدوش کننده دیگر مثل تعداد سال‌های رانندگی و... هم مثل هم باشند) آیا اگر جنسیت

معنی‌دار شود، می‌توان پذیرفت که مردان به دلیل مرد بودنشان بیشتر خطر و تصادف می‌کنند.

برای این‌که دو گروه از نظر عوامل مخدوش کننده همسان شوند، روش‌های مختلفی وجود دارد. اگر محقق‌ی بخواهد دو روش آموزشی را مقایسه کند، باید گروهی که آموزش اول را دریافت می‌کند از نظر عوامل مخدوش کننده شبیه گروهی باشد که آموزش دوم را دریافت می‌کند. یکی از روش‌های کنترل عوامل مخدوش کننده، استفاده از روش همسان‌سازی است. به طور مثال اگر در گروه اول ۴۰ درصد دختر هستند، در گروه دوم نیز باید همین تعداد دختر انتخاب شود. برای متوسط نمره آمار دو گروه قبل و بعد از انجام یک مداخله آموزشی متوسط نمره آمار دوباره محاسبه خواهد شد. به این نوع همسان‌سازی که گروهی از افراد با گروه دیگر همسان می‌شوند، همسان‌سازی گروهی (Frequency matching) می‌گویند. چنانچه از همسان‌سازی گروهی در روش اجرای پژوهش استفاده شود، می‌توان از آزمون‌های ذکر شده (χ^2)، Independent two-sample t و Mann-Whitney استفاده کرد. تفسیر نتایج با انجام همسان‌سازی گروهی دقیق‌تر خواهد بود (۴، ۶، ۷، ۱۴).

۲- تعمیم آزمون‌های اصلی برای مقایسه بیش از دو گروه (مقایسه میانگین و یا نسبت در بیش از دو گروه)

فرض بر این است که محقق‌ی قصد دارد اثر سه نوع مداخله را بر روی کاهش اندازه زخم (میلی‌متر) ناشی از سالک بررسی کند. سه گروه که از نظر متوسط اندازه زخم اولیه شبیه بوده‌اند (با استفاده از همسان‌سازی گروهی)، انتخاب شدند. پیش‌تر مشاهده شد که آزمون Independent two-sample t به مقایسه میانگین در دو گروه می‌پردازد. آزمون تعمیم یافته‌ای که امکان مقایسه میانگین در بیش از دو گروه را فراهم می‌کند، تحلیل واریانس (Variance analysis) می‌باشد (۱۵، ۱۶). البته آزمون تحلیل واریانس فقط به این فرضیه پاسخ می‌دهد که آیا میانگین در سه جامعه برابر است یا خیر؟ اگر فرض برابری میانگین‌ها رد شود، به آزمون‌های تعقیبی مانند Tukey Post-Hoc یا

می‌کنند را با یکدیگر مقایسه کند، آزمون χ^2 پیشنهاد می‌شود. در صورت رد شدن فرض برابری شیوع افسردگی در سه گروه، برای پاسخ به این سؤال که کدام دو گروه تفاوت داشته‌اند روش‌های پیشرفته‌ای مانند تجزیه کردن جدول به جداول کوچک‌تر و یا مدل‌های رگرسیونی وجود دارد، اما یک راه‌حل ساده همان مقایسه گروه‌ها به صورت دو به دو و Bonferroni correction می‌باشد.

۳- معادل‌های زوجی تکنیک‌های اصلی

روش دیگر همسان‌سازی که دقیق‌تر و البته مشکل‌تر است، همسان‌سازی فردی (Individual matching) نام دارد. چنانچه محقق بخواهد دو روش آموزشی را مقایسه کند (جنسیت و سخت‌کوشی باید کنترل شوند) و دانشجویی با نمره آمار ۱۵ و نمره اپیدمیولوژی ۱۷ در گروه آموزش اول قرار گرفت، باید دانشجویی با مشخصات مشابه در گروه دوم قرار گیرد. در واقع برای هر فرد در گروه اول فردی مشابه در گروه دوم قرار دارد. مشاهده می‌گردد که هرچه متغیرهایی که باید همسان شوند افزایش یابد، کار سخت‌تر می‌شود.

یک حالت خاص همسان‌سازی فردی آن است که فرد خودش یک بار در گروه اول و یک بار در گروه دوم قرار گیرد (هر فرد دو بار مورد مطالعه قرار گیرد). این روش مقایسه قبل و بعد (Before-after) نامیده می‌شود. به طور مثال به کل دانشجویان تا میان‌ترم با روش اول و سپس تا پایان ترم با روش دوم آموزش داده شود. سپس متوسط نمرات میان‌ترم و پایان ترم مقایسه گردد. به طور طبیعی باید مطمئن باشیم آموزش اول روی نمره پایان ترم تأثیری نداشته است. اگر در مطالعه‌ای امکان مقایسه قبل و بعد وجود داشته باشد با توجه به آن‌که افراد با خودشان مقایسه می‌شوند به طور خودکار همسان‌سازی فردی از نظر همه مخدوش‌کننده‌های احتمالی صورت گرفته است.

چنانچه از همسان‌سازی فردی و یا طراحی قبل و بعد استفاده شود، آزمون‌های گروه اصلی قابل استفاده نیست. در خانواده اصلی لازم است گروه‌ها مستقل از هم باشند، اما در

Scheffe Post-Hoc نیاز است تا به طور دقیق مشخص شود اختلاف بین کدام گروه‌ها وجود داشته است.

حال به نظر شما آیا مشکلی پیش می‌آید اگر گروه ۱ با ۲ سپس ۱ با ۳ و در نهایت ۲ با ۳ به شکل دو به دو مقایسه شوند؟ یعنی سه بار آزمون Independent two-sample t انجام گیرد. پاسخ «بلی» است و به طور قطع باید از تحلیل واریانس استفاده شود. از نظر آماری مقایسه‌های دو به دو منجر به افزایش خطای نوع اول می‌شود. اگر فرض برابری میانگین‌ها رد شده باشد؛ خطای نوع اول یعنی چقدر احتمال دارد که میانگین‌ها به طور واقعی برابر بوده و به اشتباه برابری آن‌ها رد شده باشد. این مقدار به طور معمول برابر با ۰/۰۵ در نظر گرفته می‌شود. وقتی دو یا سه آزمون انجام شد، این مقدار به ۰/۱۰ و ۰/۱۵ افزایش می‌یابد. یک راه‌حل آن است که مقدار خطای نوع اول قابل قبول از ابتدا برابر با ۰/۰۵ تقسیم بر تعداد آزمون‌ها در نظر گرفته شود. به طور مثال اگر قرار است پنج مقایسه صورت گیرد از ابتدا مقدار خطای نوع اول یا سطح معنی‌داری برابر با ۰/۰۱ در نظر گرفته شود و این بدان معنی است که $P = ۰/۰۲$ و یا $P = ۰/۰۳$ را دیگر نباید معنی‌دار تلقی کرد. این روش تصحیح را روش Bonferroni correction می‌گویند.

حال فرض بر آن است که با چارچوبی مشابه آزمون ANOVA توزیع داده‌ها نرمال نباشد، در این صورت از چه آزمونی استفاده می‌شود؟ آزمون Kruskal-Wallis برای این منظور بسیار مناسب است (۱۷، ۱۱). با توجه به این‌که آزمون Kruskal-Wallis آزمون تعقیبی ندارد، برای پیدا کردن این‌که کدام گروه‌ها با هم تفاوت دارند باید گروه‌ها با استفاده از روش Mann-Whitney دو به دو مقایسه و نتایج با استفاده از روش Bonferroni correction تصحیح شوند.

از آزمون χ^2 می‌توان برای مقایسه اختلاف نسبت صفت کیفی در بیش از دو جامعه استفاده کرد. به عنوان مثال اگر محقق بخواهد شیوع افسردگی (دارد/ ندارد) در سه گروه از افراد که به ترتیب ۰/۵، ۱/۵ و ۲/۵ گرم در روز امگا-۳ مصرف

طول زمان تغییری کرده است؟ آیا سه هفته بعد از آن نمره نسبت به قبل از آموزش بالاتر می‌رود؟ و یا آموزش فقط تأثیر کوتاه مدتی دارد و در مدت سه هفته عملکرد فرد مثل قبل از آموزش می‌باشد؟

این بدان معنی است که در یک گروه سه بار اندازه‌گیری انجام می‌شود. پس باید تکنیک Paired t به شکلی تعمیم داده شود که امکان مقایسه در بیش از دو مقطع زمانی فراهم گردد. این آزمون آنالیز واریانس با داده‌های تکراری (Repeated measures of variance) نام دارد (۱۳، ۹، ۸). مثال ذکر شده ساده‌ترین حالت است که تنها تغییر در طول زمان اندازه‌گیری می‌شود، اما باید توجه داشت که روش آماری با افزایش تعداد گروه‌ها و همچنین متغیرهای مورد بررسی پیچیده‌تر شده و باید از آزمون‌های دیگری استفاده کرد. معادل ناپارامتری روش تحلیل واریانس با داده‌های تکراری آزمون Friedman می‌باشد (۱۷، ۱۱، ۱۰، ۶).

در صورتی که متغیر وابسته اندازه‌گیری شده در طول زمان کیفی باشد، باید آزمون McNemar را توسعه داد که نتیجه آن آزمون GEE (Generalized estimation equation) است (۱۵). به طور مثال اگر شدت افسردگی گروهی از بیماران (خفیف/ شدید) قبل از مصرف دارو و سپس دو و شش ماه بعد از مصرف دارو تعیین شود تا ببینیم آیا شیوع افسردگی شدید در طول زمان کم شده است یا نه، از این آزمون استفاده می‌شود. البته آزمون GEE برای متغیرهای غیر کیفی نیز کاربردهایی دارد که خارج از بحث این مقاله است. خلاصه‌ای از آزمون‌ها همراه با مثال‌های کاربردی به اختصار در جدول ۱ و شکل ارائه شده است.

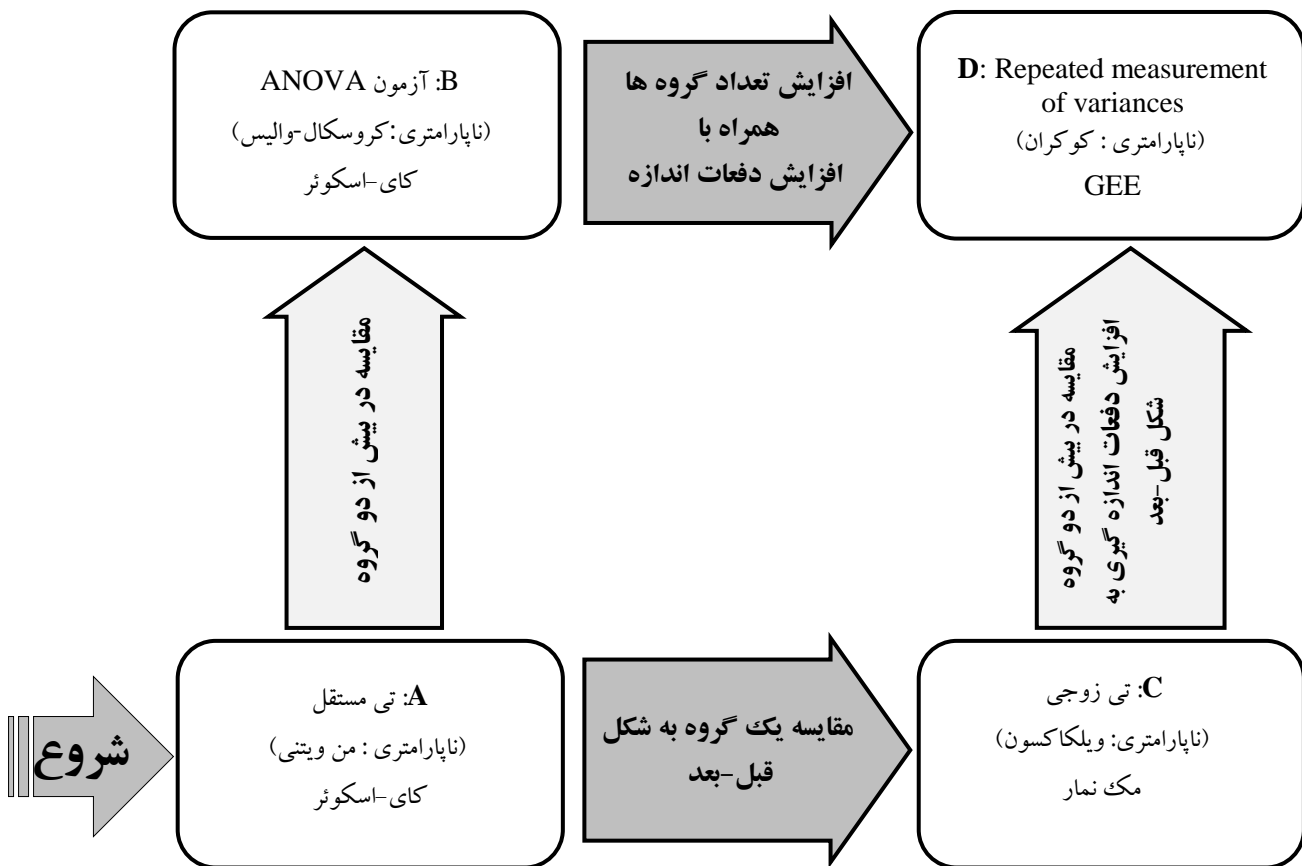
این‌جا گروه‌ها شبیه به هم انتخاب می‌شوند. این شباهت و یا وابستگی بین نمونه‌ها روی انتخاب آزمون تأثیر دارد. در این حالت به جای آزمون Independent two-sample از آزمون Paired t و به جای آزمون χ^2 از آزمون McNemar استفاده می‌شود (۱۷، ۱۱).

اگر متغیر وابسته کمی باشد، باید از آزمون Paired t استفاده شود. چنانچه مطالعه‌ای به بررسی ارتباط مصرف آسپیرین و مدت زمان لخته شدن خون در گروهی از افراد مبتلا به هایپرلیپیدمی انجام و مدت زمان لخته شدن خون قبل و بعد از مصرف آسپیرین اندازه‌گیری شود، باید از آزمون Paired t استفاده گردد. برای بررسی متغیرهای غیر نرمال در دو گروه وابسته (قبل- بعد) از آزمونی ناپارامتری به نام آزمون Willcoxon استفاده می‌شود (۱۳، ۱۱).

در صورتی که نوع مطالعه قبل- بعد و متغیر وابسته از نوع کیفی باشد، از آزمون McNemar برای مقایسه قبل و بعد از مداخله و متغیر وابسته استفاده می‌شود. برای مثال محقق به مقایسه افسردگی (دارد/ ندارد) در یک گروه از افراد قبل و بعد از مصرف امگا-۳ می‌پردازد.

تعمیم آزمون‌های زوجی برای اندازه‌گیری‌های بیش از دو بار در طول زمان

در قسمت قبل در مورد آزمون‌های مناسب وقتی متغیر وابسته دو بار در طول زمان اندازه‌گیری شده باشد (به طور معمول قبل و بعد از انجام یک مداخله) صحبت شد. حال اگر بخواهیم اندازه‌گیری را بیش از دو بار انجام دهیم چگونه؟ مثلاً محقق مایل است تأثیر یک روش تقویت حافظه را نشان دهد. نمره نمونه‌های مورد مطالعه قبل از آموزش و سپس بلافاصله بعد از آموزش و در نهایت سه هفته بعد از آموزش اندازه‌گیری شده است. محقق می‌خواهد بداند آیا متوسط نمره حافظه در



شکل ۱: چگونگی انتخاب آزمون آماری مناسب برای مطالعات حوزه علوم پزشکی، A: بررسی پیامد در دو گروه مستقل، B: بررسی پیامد در بیش از دو گروه، C: بررسی پیامد در یک گروه قبل و بعد از مداخله و D: بررسی پیامد با بیش از دو بار اندازه‌گیری

جدول ۱: مثال‌های کاربردی با توجه به نوع آزمون در زمینه فیزیولوژی، بهداشت و پزشکی بالینی

مثال کاربردی			کاربرد	نوع آزمون	گروه آزمون
فیزیولوژی	بهداشت	پزشکی بالینی			
بررسی تأثیر مکمل روی بر قد کودکان	بررسی تأثیر میزان درآمد خانوار ابتلا به تب کریمه کنگو (بلی/خیر)	مقایسه اندازه زخم سالک در دختران و پسران دبستانی شهر کرمان	مقایسه اختلاف میانگین در دو گروه با یک بار اندازه‌گیری	Independent t Mann- Whitney	اصلی
مقایسه شیوع نارسایی دریچه میترا در منطقه ۱ و ۲ شهر کرمان	مقایسه شیوع آنمی در زنان سنین باروری شاغل و خانه‌دار	مقایسه شیوع سندرم متابولیک در زنان و مردان بالای ۴۰ سال	مقایسه اختلاف نسبت در دو گروه با یک بار اندازه‌گیری	χ^2	
بررسی کاهش کلسترول در یک گروه از مبتلایان به کم‌کاری تیروئید قبل و بعد از آموزش حرکات هوازی	مقایسه معدل دانش‌آموزان قبل و بعد از تغییر روش آموزشی	مقایسه اندازه زخم سالک قبل و بعد از مصرف عصاره آلونهورا	مقایسه اختلاف میانگین در دو گروه با یک بار اندازه‌گیری به صورت قبل و بعد از مداخله	Paired t (Willcoxon)	زوجی (قبل-بعد)

مقایسه میزان مرجعه برای ماموگرافی در کارکنان معاونت های دانشگاه علوم پزشکی قبل و بعد از کلاس های آموزشی	مقایسه شدت افسردگی در کارکنان زندان کرمان قبل و بعد از بهبود محیط کار	شیوع لیشمانیوز در دانش آموزان مدارس ابتدایی منطقه ۱ بم قبل و بعد از تجویز واکسن	مقایسه اختلاف نسبت در دو گروه با یک بار اندازه گیری به صورت قبل و بعد از مداخله	McNemar	
بررسی اثربخشی استفاده از پوشش های مختلف بر کاهش درماتیت تماسی	مقایسه میانگین تعداد تصادفات رانندگی در فصل های مختلف	مقایسه میزان هموگلوبین در دختران شهرستان های استان اصفهان	مقایسه اختلاف میانگین در بیش از دو گروه با یک بار اندازه گیری	Kruskal- (Wallis ANOVA)	گسترده (بیش از دو گروه)
مقایسه شیوع افسردگی در مردان راننده، معلم و کارگر	مقایسه شیوع پوکی استخوان در شهر کرمان، زاهدان و یزد	مقایسه شیوع کم کاری تیروئید در شهرستان های کرمان، رفسنجان و بافت	مقایسه اختلاف نسبت در بیش از دو گروه با یک بار اندازه گیری	χ^2	
مقایسه اندازه زخم سالک در افراد ۱۸-۱۰، ۳۰-۱۹، ۴۰- و ۳۱ ساله در هفته های ۴، ۸ و ۱۲ پس از کرایوتراپی	بررسی میزان اکسیژن بیولوژیکی مورد نیاز (BOD) در ۷ تصفیه خانه تهران، یک ماه و یک سال پس از بهسازی	اثربخشی مصرف روزانه چای سبز بر BMI: کارآزمایی بالینی تصادفی شده (مقایسه با گروه شاهد در هفته های ۴، ۲، ۶)	مقایسه اختلاف میانگین در بیش از دو گروه با بیش از یک بار اندازه گیری	Repeated measures of variances (Friedman)	بیش از دو گروه و بیش از یک بار اندازه گیری
مقایسه شیوع افسردگی قبل از زایمان و سه و نه ماه پس از زایمان در روستای علی آباد	مقایسه میزان مصرف شیر در کارکنان بخش رادیولوژی بیمارستان های دانشگاه علوم پزشکی کرمان (سه بیمارستان) ۲، ۴، ۶ هفته پس از اطلاع رسانی	مقایسه میزان بیماری های تنفسی در کارکنان کارخانه سیمان قبل و بعد از یک و سه سال استفاده از ماسک	مقایسه اختلاف نسبت در بیش از دو گروه و با بیش از یک بار اندازه گیری	GEE	

BOD: Biochemical oxygen demand; BMI: Body mass index; GEE: Generalized estimation equation

هم انتخاب شوند، همان آزمون های ساده ای که در قسمت های قبل ارایه گردید برای تحلیل داده ها کفایت می کند. چنانچه محقق در زمان طراحی مطالعه همسان سازی را انجام نداده باشد، راه حل چیست؟ در روش هایی که تا این جا به آن پرداخته شد تنها ارتباط یک متغیر مستقل با یک متغیر وابسته بررسی گردید. اگر دو گروه از نظر متغیرهای مخدوش کننده شبیه به هم انتخاب نشوند، باید خود آن متغیرها به عنوان متغیرهای مستقل وارد تحلیل شوند. به عبارت دیگر باید تأثیر همزمان چندین متغیر مستقل بر روی یک متغیر وابسته بررسی گردد.

گروهی از تکنیک های آماری که در آن به بررسی نقش چندین متغیر مستقل روی یک متغیر وابسته پرداخته می شود،

محققین باید توجه داشته باشند که در بیشتر پژوهش ها ممکن است بیش از یک آزمون استفاده شود. ضمن این که قالب های ذکر شده در مقاله حاضر در شرایط خاص قابل تغییر بوده و می توان با توجه به سؤال پژوهش از آن ها استفاده کرد.

آزمون های چند متغیره

فرض کنید محقق می خواهد دو روش جراحی قلب را مقایسه کند و به مقایسه درصد فوت بیماران در دو روش بپردازد. متغیر فوت متغیر وابسته و متغیر نوع درمان متغیر مستقل نامیده شد. احتمال دارد متغیرهایی مانند سن بیمار، مصرف سیگار و سابقه بیماری قلبی بتواند روی فوت بیمار تأثیرگذار باشد. اگر دو گروه درمانی از نظر این عوامل شبیه به

هم آزمون Covariance analysis و هم تحلیل واریانس دو طرفه جزء گروه مدل‌های رگرسیونی خطی محسوب می‌شوند. هدف از مدل‌های رگرسیونی، بررسی تأثیر هم‌زمان نقش متغیرهای مستقل بر متغیر وابسته می‌باشد.

بحث و نتیجه‌گیری

همان‌گونه که برای همه محققین واضح است، علم آمار یکی از ارکان تحقیق می‌باشد و عدم آگاهی از مفاهیم آماری بر روی نتیجه‌گیری نهایی محقق به شدت اثر می‌گذارد. این امر در مطالعات حوزه علوم پزشکی به علت ارتباط نتایج مطالعات بر سلامت افراد جامعه از اهمیت بسزایی برخوردار است؛ چرا که نتایج حاصل از پژوهش در بیشتر موارد وجود یا عدم وجود رابطه در زمینه سبب‌شناسی بیماری یا مشکلات بهداشتی، پیشگیری، تشخیص و درمان را ثابت می‌کند و در صورت تعمیم این نتایج اشتباه به جامعه هدف گاه خساراتی غیر قابل جبران بر افراد جامعه وارد خواهد شد.

گرچه مفاهیم و تفسیر نتایج آماری به ویژه آمار زیستی به نسبت دشوار و گاه پیچیده است، اما دانستن برخی مفاهیم و آزمون‌ها به صورت کلی می‌تواند محقق را در رسیدن به پاسخ سؤال پژوهش یاری دهد و از تفسیر نادرست نتایج پیشگیری کند. در مقاله حاضر تلاش شد با زبانی ساده و شیوا و فارغ از بحث‌های تخصصی و پیچیده، روند انتخاب آزمون‌های آماری در هنگام نگارش پروپوزال و همچنین درک بهتر و تفسیر صحیح بخش آزمون‌های آماری مقالات توسط محققین تسهیل گردد. لازم به ذکر است گرچه انتخاب آزمون در ابتدای پژوهش از اهمیت ویژه‌ای برخوردار است، اما در طول اجرای تحقیق و بسته به نوع تحقیق ممکن است آزمون‌های مختلفی در تحلیل نتایج نهایی استفاده شود.

با وجود این‌که دانستن آزمون‌ها در روند خواندن و ارزیابی مقالات بسیار کمک کننده می‌باشد، اما حضور متخصص آمار زیستی و اپیدمیولوژی در هنگام طراحی (تخمین حجم نمونه و تعیین و بحث در خصوص متغیرهای مرتبط با هدف پژوهش،

گروه تکنیک‌های چند متغیره رگرسیونی می‌نامند (۱۸، ۱۹). اگر متغیر وابسته کمی باشد، نوع رگرسیون خطی است. توجه به این نکته ضروری به نظر می‌رسد که اگر فقط یک متغیر مستقل وجود داشته باشد نیز از رگرسیون تک متغیره استفاده می‌شود. به بیان دیگر، Independent two-sample و تحلیل واریانس هم نوعی رگرسیون خطی تک متغیره محسوب می‌شوند. در حالتی که متغیر وابسته کیفی دو حالتی باشد (مثل نتیجه درمان جراحی به صورت فوت یا زنده مانده بیمار) آن را رگرسیون لجستیک (Logistic regression) می‌نامند که خارج از این بحث می‌باشد (۱۸، ۲۰).

بسته به متغیرهای مستقل موجود در مدل رگرسیون خطی، اسامی اختصاصی‌تری هم برای آن در نظر گرفته شده است (۲۱). پیش‌تر مثالی مطرح شد با این مضمون که محقق مایل است اثر سه نوع مداخله (متغیر مستقل) را بر روی کاهش اندازه زخم (متغیر وابسته) ناشی از سالک بررسی کند. اگر سه گروه از نظر متوسط مساحت زخم اولیه شبیه انتخاب نشوند باید خود متغیر زخم اولیه را وارد تحلیل کرد. یعنی به جای آزمون تحلیل واریانس از نوعی آزمون رگرسیونی استفاده نمود که یک متغیر کمی را هم وارد فرایند تحلیل کرده باشد. چنین آزمونی، تحلیل کواریانس (Covariance analysis) نام دارد. دقت شود که متغیر وابسته از نوع کمی (مساحت بعد از دریافت دارو) و دو متغیر مستقل یکی کیفی (نوع درمان) و دیگری مخدوش کننده کمی (اندازه زخم اولیه) است.

محقق می‌خواهد تأثیر سه روش آموزشی را بر روی نمره دانشجویان مقایسه کند. سه گروه از دانشجویان پرستاری انتخاب می‌شوند. در یک گروه استاد درس می‌دهد، در گروه دوم دانشجویان درس را ارایه می‌کنند و در روش سوم کلاس به صورت پرسش و پاسخ برگزار می‌شود. برای مقایسه نمرات پایان ترم در سه گروه از تحلیل واریانس استفاده گردید. حال اگر متغیر جنسیت دانشجوی هم در تحلیل وارد شود، از تحلیل واریانس دو طرفه استفاده می‌شود. در این‌جا متغیر وابسته از نوع کمی و دو متغیر مستقل از نوع کیفی بود.

و...) در تیم تحقیق ضروری می‌باشد و باعث حرکت پژوهش در راستای پاسخ به سؤال مورد نظر محقق خواهد شد.

References:

- Okeh UM. Statistical problems in medical research. *East Afr J Public Health* 2009; 6(1):1-7.
- Sathian B, Sreedharan J. Importance of biostatistics to improve the quality of medical journals. [cited 2012 May 5]. Available from: http://67.227.236.94/wmcpdf/Article_WM003332.pdf
- West CP, Ficalora RD. Clinician attitudes toward biostatistics. *Mayo Clinic Proc* 2007; 82(8): 939-43.
- Mohammad K, Malekafzali H. Statistical methods and health indices. Tehran: Salman; 2000. [In Persian]
- Baneshi MR, Haji Maghsoudi S, Rastegari A. Evidence-based statistics: appropriate approach for description and analysis of research data. *Strides Dev Med Educ* 2011; 8(1): 4-13. [In Persian]
- Ayatollahi MT. Principals and methods in biostatistics. Tehran: Sepehr; 2000. [In Persian]
- Cadarso-Suárez C, González-Manteiga W. Statistics in biomedical research. *Arbor* 2007; 183(725): 353-61.
- Shahravan A, Baneshi MR. Statistics for dental researchers; Part 1: Introduction. *Journal of Oral Health and Oral Epidemiology* 2012; 1(1): 1-2. [In Persian]
- Shahravan A, Ghassemi AR, Baneshi MR. Statistics for Dental Researcher, Part 2: Descriptive Statistics. *Journal of Oral Health and Oral Epidemiology* 2013; 1(2): 53-59. [In Persian]
- Behboodian J. Nonparametric statistics. 5th ed. Shiraz: University of Shiraz; 2008. [In Persian]
- Govindarajulu Z. Nonparametric Inference. Hackensack: World Scientific; 2007.
- Govindarajulu Z. A brief survey of nonparametric statistics. *Commun Stat Theory Methods* 1976; 5(5): 429-53.
- Kirkwood B, Sterne JA. Exact test for 2*2 table: *Essentials of Medical Statistics* Blackwell. 2nd ed. Oxford: John Wiley and Sons Ltd; 2003: 69-71.
- Gehan EA. The role of the biostatistician in cancer research. *Biom Pharmacother* 2001; 55(9): 502-9.
- Rosner B. Fundamentals of biostatistics experiments. Boston: Thomson; 2006.
- Sonnerup BUO, Scheible M. Minimum and maximum variance analysis. *Analysis methods for multi-spacecraft data* 1998; 1: 185-220.
- Bahrampour A. Inference in biostatistics. Kerman: Vadiat; 2002. [In Persian]
- Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 2nd ed. USA: Wiley & Sons, Inc; 2013.
- Kutner MH, Nachtsheim C, Neter J. Applied Linear Regression Models. 4th ed. Irwin: McGraw-Hill, 2004; 10-12.
- Neter J, Wasserman W, Kutner MH. Applied linear statistical models. 5th ed. Irwin: McGraw-Hill; 2004.
- Seber GA, Lee AJ. Linear regression analysis. 2nd ed. New Jersey: Wiley; 2003.

A Guide to Selecting the Appropriate Statistical Tests for Proposals and Articles in Medical Sciences

Mohammad Reza Baneshi^{1}, Farzaneh Zolala², Elham Mohebbi³*

1. *Ph.D in Biostatistics, Associate Professor, Department of Epidemiology and Biostatistics, Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran.*
2. *Ph.D in Epidemiology, Assistant Professor, Department of Epidemiology and Biostatistics, Regional Knowledge Hub, and WHO Collaborating Centre for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran*
3. *DVM/MPH, Researcher, Research Center for Social Determinants of Health, Institute for Futures Studies for Health, Kerman University of Medical Sciences, Kerman, Iran*

• **Received:** 17 Dec, 2013

• **Received Corrected Version:** 10 Feb, 2014

• **Accepted:** 17 Feb, 2014

Background & Objective: The main purpose of medical researches is to answer a research question or to solve a problem to promote the health of a society. The first objective is to answer the research question correctly with minimal errors. The second objective is the publication of the results in order to generalize them to a population and use in a wider dimension. To achieve these objectives, using biostatistics is necessary. Despite the importance of biostatistics in medical research, researchers have limited understanding of it or due to its complications they refrain from its use. Statistics help the researcher in different levels of research including writing a proposal and interpretation of other papers. Moreover, biostatisticians and epidemiologists also play a very important role in the preparation of manuscripts for publication. The present article has eloquently described the most important statistical tests in medical research with applied examples.

Key Words: Selecting statistical tests, Parametric tests, Non-parametric tests

* **Correspondence:** Research Center for Modeling in Health, Somayeh Square, Jahad Street, Kerman University of Medical Sciences, Kerman, Iran

• **Tel:** (+98) 341 226 3725

• **Fax:** (+98) 341 226 3725

• **Email:** rbaneshi@yahoo.com