

## آمار مبتنی بر شواهد: رویکرد مناسب توصیف و تحلیل داده های یک پژوهش

محمد رضا بانسی<sup>۱\*</sup>، سعیده حاجی مقصودی<sup>۲</sup>، اعظم رستگاری<sup>۲</sup>، محمد رضا محمودی<sup>۳</sup>

۱. دکترای آمار زیستی، استادیار گروه آمار زیستی، مرکز تحقیقات مدل سازی در سلامت، دانشکده بهداشت، دانشگاه علوم پزشکی کرمان، کرمان، ایران

۲. دانشجوی کارشناسی ارشد آمار زیستی، گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمان، کرمان، ایران

۳. دکترای تغذیه، استادیار گروه تغذیه، مرکز تحقیقات فیزیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمان، کرمان، ایران

• دریافت مقاله: ۸۹/۱۱/۱۰    آخرین اصلاح مقاله: ۸۹/۱۱/۲۰    • پذیرش مقاله: ۸۹/۱۲/۱۲

در سال‌های اخیر استفاده از آمار در زمینه پزشکی رشد قابل ملاحظه ای داشته است و جزء تفکیک ناپذیر هر مطالعه و تحقیق علمی به شمار می رود. امروزه برای انواع مطالعات و برای هر نوع داده ای روش آماری مناسب وجود دارد که منجر به تولید نتایج معتبر و صحیح می شود. بنابراین آشنایی با موارد فوق بر محققین و خوانندگان مطالعات پزشکی لازم است و این موضوع برای پژوهشگران، امری مهم تر به شمار می آید زیرا آنها می توانند با انجام یک مطالعه نامناسب و ارائه نتایج نادرست موجب گمراهی جمع کثیری از خوانندگانی که آشنایی زیادی با آمار، مفاهیم و کاربردهای آن ندارند شده و علاوه بر آن نیز باعث اتلاف زمان، انرژی و هزینه زیادی که از آغاز طراحی پروژه تا چاپ آن می شود، بشوند. اخیرا مقالات آموزشی اندکی از سوی محققان کشور برای درک بهتر پژوهشگران از آمار و کاربرد آن در روش شناسی پژوهشی انتشار پیدا کرده است، اما هنوز اشتباهات زیادی در مقالات یافت میشود. در این مقاله سعی شده است که روشهای صحیح و مناسب ارائه نتایج یک پژوهش، اهمیت و کاربرد فواصل اطمینان و اشتباهات رایج در تحلیل داده ها (مانند تحلیلهای تک متغیره، در نظر نگرفتن توان ازمون و استفاده از ضریب همبستگی برای سنجش توافق بین نمرات) ذکر شود.

کلید واژه‌ها: آمار مبتنی بر شواهد، توصیف، تحلیل

\* نویسنده مسؤول: گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمان، ابتدای هفت باغ، کرمان، ایران

## مقدمه

آمار به عنوان یکی از ابزارهای ضروری مورد نیاز یک پژوهش معتبر و علمی شناخته شده است. علم آمار برای کسب اطلاعات بیشتر در جهت برنامه ریزی و سیاست‌گذاری به منظور عرضه خدمات پزشکی و مراقبت‌های بهداشتی مناسب در جامعه مورد استفاده قرار می‌گیرد و در حال حاضر پزشکان و مدیران این عرصه توجه قابل ملاحظه‌ای به نقش علمی و کاربردی آن در موارد فوق دارند. بنابراین باید سعی شود پژوهش‌های انجام شده کمترین خطا را داشته و صحیح‌ترین روش برای تحلیل داده‌ها و ارائه نتایج استفاده شود. خصوصاً به دلیل حساسیت مطالعات در حیطه پزشکی، این دقت باید بیشتر باشد.

یکی از نقاط ضعف محققان کشور در زمینه علوم پزشکی، عدم آشنایی به مفاهیم آماری و نحوه به‌کار بردن مناسب‌ترین آزمونهای آماری به هنگام تحلیل داده‌ها می‌باشند. توجه داشته باشید که هر قدر که در روش کار خود دقت هم داشته باشید، اگر نتوانید آن را به خوبی تحلیل کنید و آزمونهای آماری مناسبی برای آن بکار نبرید، نمی‌توانید نتایج قابل قبولی برای ارائه گزارش نهایی داده‌ها به دست بیاورید.

امروزه انتشار مجلات علوم پزشکی چه در داخل کشور و چه در خارج کشور روز به روز در حال افزایش است. مجلات که خود به عنوان یک منبع اطلاعاتی قوی برای انتشار نتایج تحقیق بکار می‌روند، اگر در انتشار نتایج مقالات دقت کافی و لازم به خرج ندهند، هم خود به اشتباه رفته و هم سایرین را به راه نادرست هدایت خواهند کرد. با توجه به چنین نیازی در این مقاله سعی شده است که به طور خلاصه روش‌های صحیح و مناسب آماری برای تحلیل و ارائه نتایج یک پژوهش ذکر شود تا شاید نقاط ابهامی که محققین و خوانندگان در ارائه و گزارش نتایج یک پژوهش با آن مواجه هستند برطرف شود. البته فرض بر این است که مطالعه با چارچوب استاندارد و صحیح، طراحی و داده‌ها جمع‌آوری و سازماندهی شده و آماده تحلیل و تفسیر می‌باشد. به‌علاوه در این مقاله بعضی

اشتباهات آماری که به طور مکرر در مقالات دیده می‌شوند و راه صحیح برخورد با آنها ذکر شده است.

در این نوشتار به بیان مباحثی همچون آمار توصیفی، بررسی نرمال بودن داده‌ها، مفهوم P-Value و فاصله اطمینان، آزمونهای آماری متداول، تحلیلهای چند متغیره، روشهای تصادفی سازی توان آزمون در تحلیلهای آماری و بررسی توافق پرداخته و سعی شده که این مفاهیم با بیانی ساده و کاربردی برای محققین بیان گردد.

## آمار توصیفی

در این قسمت ابتدا به بررسی شاخص‌های توصیفی مناسب می‌پردازیم. سپس در مورد نحوه صحیح گزارش درصد‌ها و گروه‌بندی کردن متغیرهای کمی صحبت شده است. گام اول برای نوشتن نتایج یک پژوهش توصیف داده‌ها در قالب شاخص‌ها و نمودارهای مناسب می‌باشد. این امر کمک میکند که خواننده یک دید کلی نسبت به توزیع متغیرهای مورد مطالعه و نتایج آن پیدا کند(۱). شاخص مناسب برای توصیف داده‌ها بر اساس نوع متغیر انتخاب می‌شود. متغیرها به دو دسته متغیرهای عددی یا کمی (متغیرهایی که قابل اندازه‌گیری هستند، مثل نمره فرد در امتحان) و متغیرهای غیر عددی یا کیفی (متغیرهایی که بر اساس آنها نمونه‌های مورد بررسی گروه‌بندی می‌شوند) تقسیم می‌شوند.

در حالتی که متغیر مورد نظر عددی باشد، باید شاخص‌های میانگین و انحراف معیار را می‌توان گزارش کرد. البته وجود داده‌های پرت مقدار میانگین را تحت تأثیر قرار می‌دهد. در صورت چولگی داده‌ها چارک‌ها و کمترین و بیشترین مقدار داده و یا میانگین پیراسته (Trimmed mean) را می‌توان گزارش داد. نمودارهای مناسب برای متغیرهای کمی شامل هیستوگرام و نمودار جعبه‌ای می‌باشند.

بسیاری مواقع دو شاخص انحراف معیار و خطای معیار در جای مناسب خود گزارش نمی‌شوند. توجه کنید که انحراف معیار یک شاخص توصیفی است که پراکندگی داده‌های نمونه

توصیه میشود هر گاه فراوانی کمتر از ۲۰ باشد از گزارش فراوانی نسبی پرهیز شود. در مواقعی که دو متغیر کیفی وجود دارد از جدول متقاطع استفاده می شود. به عنوان مثال در حالتی که هر دو متغیر دارای دو حالت باشند این جداول دارای دو ردیف و دو ستون هستند (۴ خانه در کل). در جداول متقاطع بهتر است که متغیر وابسته در ردیف و متغیر مستقل در ستون قرار گیرد. طبیعتاً میتوان بر اساس فراوانی روی سطر ستون و یا فراوانی کل به محاسبه درصد پرداخت. ولی کدامیک از این ۳ روش صحیح تر است؟ ملاک اصلی برای درصد گیری آن است که، توزیع متغیر وابسته را در سطوح متغیر مستقل بدست می آوریم. مثلاً اگر ۱۰۰ دانشجوی با روش آموزشی سنتی و ۱۰۰ دانشجو با روش آموزشی مجازی آموزش دیدند و هدف محقق بررسی درصد قبولی دانشجویان باشد (جدول شماره ۱).

جدول ۱: توزیع وضعیت قبولی دانشجویان بر حسب نوع آموزش

نوع آموزش/ وضعیت قبولی	قبولی	مردودی	جمع کل
آموزش سنتی	۹۰	۱۰	۱۰۰
آموزش مجازی	۹۵	۵	۱۰۰
جمع کل	۱۸۵	۱۵	۲۰۰

با توجه به داده ها به نظر شما کدام یک از سه تفسیر زیر صحیح است؟ الف) تنها ۱۰٪ از کسانی که با روش آموزش سنتی دوره دیده اند مردود شده اند (درصد بندی روی سطر) ب) حدود ۶۷٪ از کسانی که مردود شده اند با روش سنتی آموزش دیده اند (درصد بندی روی ستون). ج) در کل ۵٪ نمونه ها کسانی بوده اند که با روش سنتی آموزش دیده و مردود شده اند (درصد بندی روی کل). در تفسیر الف و ج خواننده احتمالاً متقاعد میشود که روش سنتی مناسب است. در تفسیر ب درصد بندی بر روی متغیر وابسته (قبولی در امتحان) انجام شده که صحیح نیست. در اینجا به خواننده القا میشود که روش سنتی در مقایسه با روش مجازی بسیار ناکارآمد است. البته توجه داشته باشید که تفسیر ج نیز صحیح نیست چون در طراحی مطالعه ۱۰۰ نفر در هر روش انتخاب شده است.

را حول میانگین آن‌ها نشان می دهد. فرض کنید میانگین نمرات دانشجویان دو کلاس در درس آمار حیاتی ۱۵/۵ باشد. به نظر شما اگر حداقل و حداکثر نمره به ترتیب ۵ و ۲۰ (در کلاس اول) و ۱۳ و ۱۷ (در کلاس دوم) باشد آیا عدد ۱۵/۵ نماینده مناسبی برای وضعیت هر دو کلاس است؟ طبیعتاً پاسخ منفی است. برای اینکه ببینیم میانگین به دست آمده تا چه حد نماینده مناسبی از وضعیت نمرات دو کلاس است باید پراکندگی نمرات را نیز در نظر بگیریم. اگر نمرات کلاس اول دارای پراکندگی کمتر و تجمع بیشتر حول میانگین باشند میانگین در این کلاس انعکاس دهنده وضعیت بهتری از نمرات است (۲،۳،۴،۵).

از سوی دیگر هدف اصلی هر مطالعه ای تعمیم نتایج نمونه به جامعه است. به طور مثال می خواهیم از میانگین نمونه به عنوان بر آوردی از میانگین جامعه استفاده کنیم اما مشکل اینجاست که میانگین هر نمونه با میانگین نمونه های دیگری از همان جامعه متفاوت است که در اصطلاح به آن تغییرات نمونه ای میانگین می گویند. این تغییرات را می توان به وسیله شاخص خطای معیار اندازه گیری و گزارش نمود. به بیان دیگر شاخص خطای معیار در قسمت تحلیل اطلاعات نقش دارد. اگر پژوهشگر به دنبال پاسخ به این سؤال باشد که برآورد انجام شده از نمونه برای جامعه چقدر صحیح است می تواند آزمایش را به تعداد بسیار تکرار کند و برای هر کدام میانگین را محاسبه نماید. سپس برای میانگین ها، انحراف معیار محاسبه کند و این عدد همان خطای معیار خواهد بود که در واقع میزان تغییرپذیری میانگین را به ما نشان می دهد. از خطای معیار برای ساختن فاصله اطمینان استفاده می شود (۶،۷). در ادامه در مورد فاصله اطمینان بیشتر صحبت شده است ولی قبل از آن شاخص های توصیفی مناسب برای متغیر های کیفی مرور شده اند.

اگر متغیر مورد بررسی کیفی باشد، فراوانی و درصد و همچنین نمودارهای دایره ای یا میله ای (که به ترتیب جهت ارائه درصد و فراوانی مطلق به کار می روند) باید آورده شوند.

باشد در این صورت دو واحد تفاوت وجود دارد از آنجا که مطالعه انجام شده فقط روی یک نمونه انجام شده است نمی توان انتظار داشت که با تکرار آن بر روی یک نمونه دیگر دقیقاً دو واحد اختلاف مشاهده گردد این تغییرات که ناشی از نمونه برداریهای متفاوت است را می توان با ارائه فاصله اطمینان (Confidence Interval) نمایش داد(۶). در ساختن فواصل اطمینان از تفاوت مشاهده شده (در اینجا دو واحد) و خطای معیار که در قسمت قبل توضیح داده شد استفاده می شود. منظور از فاصله اطمینان ۹۵٪ آن است که اگر مطالعه را ۱۰۰ بار تکرار کنیم در ۹۵ بار برآوردی که از تفاوت میانگین دو گروه بدست خواهد آمد، در فاصله اطمینان ارائه شده قرار خواهد گرفت. حال اگر فاصله اطمینان بدست آمده (۶ و ۲-) باشد بدین معنی است که میانگین نمرات در روش سنتی می توانسته از ۲ واحد کمتر تا ۶ واحد بیشتر از روش مجازی باشد. این بدان معنی خواهد بود که تفاوتی بین این دو روش آموزش وجود ندارد چون عدد صفر نیز در این بازه قرار گرفته است. پس می بینید که فاصله اطمینان نه تنها به خواننده اجازه می دهد در مورد معنی داری آماری قضاوت کند بلکه مقدار تفاوت بین گروهها را نیز مشخص می کند(۹).

به این نکته نیز توجه داشته باشید که معنی داری آماری نمی تواند لزوماً مهم تلقی شود(۹). مثلاً اگر تفاوت میانگین نمرات در دو روش ۰/۵ واحد باشد و مقدار خطای معیار عدد کوچکی باشد مقدار P-Value زیر ۰/۰۵ خواهد شد و ما نتیجه می گیریم که اختلاف معنی داری وجود دارد. ولی ۰/۵ نمره ممکن است آن مقدار از نظر آموزشی اهمیت نداشته باشد. با استفاده و ارائه فاصله اطمینان می توان معنی دار بودن را نه تنها از لحاظ آماری بلکه از نظر بالینی نیز مورد بحث قرار داد(۹).

### بررسی نرمال بودن توزیع داده ها

در اکثر روشهای آماری (مثلاً در آزمون t دو نمونه مستقل برای مقایسه میانگین در دو گروه) فرض بر این است که توزیع

نکته دیگری که بد نیست به آن اشاره شود این است که در برخی از پژوهشها هنگام تحلیل به دلایل متعدد محققان سعی در طبقه بندی کردن متغیرهای کمی دارند. اگر چه این روش باعث سادگی در تحلیل می شود اما همیشه بهترین نتیجه ممکن را در پی نخواهد داشت (۸) باید توجه نمود که در صورت لزوم استفاده از طبقه بندی، برای نقاط برش باید دلایل بیولوژیکی و آماری قابل توجیه وجود داشته باشد.

### مفهوم و کاربرد P-Value و فاصله اطمینان در تحلیل های آماری

همانطور که در قسمت قبل اشاره شد در مطالعات تحلیلی محقق به دنبال تعمیم نتایج نمونه به کل جامعه هدف می باشند (۹). فرض کنید محقق میایل است دو روش متفاوت آموزش مجازی و سنتی را باهم مقایسه کند. بدین منظور تعدادی دانشجوی، درس را به صورت آموزش مجازی و تعدادی به صورت سنتی فرا گرفته اند. قاعداً با استفاده از آزمون t دو نمونه مستقل، میانگین نمرات در هر گروه مقایسه می شود. برای گزارش نتایج معمولاً از P-Value استفاده می شود. P-Value به ما نشان می دهد که تفاوت مشاهده شده بین دو گروه تا چه اندازه ناشی از شانس و تصادف بوده است(۱۰). توجه کنید که هر مطالعه روی تعدادی نمونه با خصوصیات مشخص و در یک مقطع زمانی و مکانی خاص انجام می شود. حتی اگر پژوهشگر سعی کند تمامی شرایط را تحت کنترل داشته باشد امکان تغییر پذیری های تصادفی که بتواند نتایج را تحت تأثیر قرار دهند وجود دارد(۱۱). اگر مقدار P-Value کم باشد(معمولاً کمتر از ۰/۰۵) می پذیریم که تفاوت مشاهده شده نمی تواند تصادفی باشد و این بدان معنی است که دو گروه با هم اختلاف معنی داری دارند(۱۰).

نکته مهمی که در اکثر مطالعات جای خالی آن دیده می شود آن است که نه تنها وجود تفاوت بین دو گروه، بلکه مقدار آن نیز از اهمیت خاصی برخوردار است(۹). مثلاً اگر میانگین نمرات دانشجویان در روش سنتی و مجازی ۱۶ و ۱۴ شده

آنها و قسمتهای مهمی که در نتایج باید ارائه شوند آورده شده است (۱،۲،۱۳).

جدول ۲: آزمونهای پارامتری و آزمونهای ناپارامتری معادل همراه با نتایج قابل گزارش در نتایج پژوهش

آزمون ناپارامتری معادل	نتایج	آزمون
Mann-Whitney	حجم نمونه، میانگین و انحراف معیار در هر گروه، فاصله اطمینان و P-Value	t-test
Wilcoxon	میانگین و انحراف معیار تفاوت قبل و بعد، فاصله اطمینان و P-Value	Paired t-test
Kruskal-Wallis	حجم نمونه، میانگین و انحراف معیار در هر گروه، فاصله اطمینان و P-Value برای تفاوتی گروهها به صورت دو به دو به دنبال آزمونهای تعقیبی (Post hoc)	ANOVA
Spearman correlation	مقدار ضریب همبستگی (r) و P-Value	Pearson Correlation
	جدول مقاطع، فراوانیها و درصدیهای مشاهده شده در خانه‌های جدول و P-Value	Chi-square

مزیت استفاده از روش‌های تحلیل چند متغیره نسبت به تک متغیره

آزمونهای آماری که در قسمت قبل و در جدول ۲ مورد بحث قرار گرفتند به عنوان روش‌های تک متغیره شناخته می‌شوند. در اکثر مطالعات متغیرهای زیادی اندازه گیری می‌شوند و هدف بررسی تأثیر آنها روی یک پیامد خاص است. در روش تک متغیره اثر هر عامل روی پیامد مطالعه بدون در نظر گرفتن اثر بقیه متغیرها انجام شود. البته روشهای تک متغیره علی رغم سادگی چندان مناسب نیستند چون اثر سایر عوامل را تعدیل نمی‌کنند. زمانی که بخواهیم اثر چند متغیر

داده‌ها نرمال باشد. در صورتیکه توزیع داده‌ها نرمال نباشد می‌توان با استفاده از تبدیلات آماری (مثلاً لگاریتمی) تقریباً آن را به صورت نرمال درآورد. البته در این صورت تفسیر نتایج مشکل خواهد بود (۱۲). راه حل دیگر استفاده از روشهای ناپارامتری است.

در اکثر مواقع محققین از آزمون Kolmogorov-Smirnov برای بررسی نرمالیتی استفاده می‌کنند. توجه داشته باشید که تصمیم‌گیری تنها براساس P-Value می‌تواند گمراه‌کننده باشد. به عنوان مثال فرض کنید حجم نمونه برابر ۱۰ باشد در این صورت مقدار P-Value احتمالاً خیلی بیشتر از ۰/۰۵ خواهد بود و شما نتیجه خواهید گرفت که توزیع داده‌ها نرمال است. اما دلیل معنی دار نشدن آزمون می‌توانسته پایین بودن توان آزمون باشد (مفهوم توان آزمون در ادامه شرح داده شده است). این در حالی است که اگر شما هیستوگرام را رسم کنید ممکن است شاهد یک توزیع چوله باشید. به عنوان مثالی دیگر، فرض کنید حجم نمونه ۱۰۰۰ باشد اگر انحراف کوچکی نیز از نرمالیتی وجود داشته باشد آزمون Kolmogorov-Smirnov به دلیل توان بالا معنی دار خواهد شد در حالی که اگر هیستوگرام آن را رسم کنید شکلی تقریباً متقارن خواهید دید. پس به یاد داشته باشید که همواره به هیستوگرام داده‌ها نیز توجه کنید (۱۲).

آزمونهای متداول آماری و موارد قابل گزارش در نتایج پژوهش

هنگامی که یک آزمون آماری با نرم افزار انجام میشود حجم زیادی از اطلاعات در خروجی نرم افزار مشاهده خواهد کردید. توجه کنید که لازم نیست تمام خروجی در متن مقاله ارایه شوند. به عنوان مثال بعد از استفاده از آزمون t مقدار اماره و درجه آزادی را گزارش میشود ولی فاصله اطمینان نه!! گزارش این اماره‌ها هیچ ضرورتی ندارد. در جدول ۲ پر استفاده‌ترین روش‌های پارامتری به همراه معادل ناپارامتری

جدول ۳-ب: بررسی ارتباط بین قبولی در امتحان با جنسیت در آزمون تک متغیره بر اساس وضعیت تأهل

	متأهل		مجرد	
	قبول	مردود	قبول	مردود
جنسیت	۱۲۰۰	۱۲۰	۱۸۰۰	۵۰
	۱۸۰۰	۱۲۶	۱۲۰۰	۱۵
Odds ratio=۱/۴۲		Odds ratio=۲/۲۲		
CI:(۱/۱۰, ۱/۸۵)		CI:(۱/۲۴, ۳/۹۸)		
P-Value=۰/۰۰۷		P-Value=۰/۰۰۶		

جدول ۳-ج: بررسی ارتباط بین قبولی در امتحان با جنسیت و تأهل در آزمون چند متغیره

Contributory Factor	Odds ratio	P_Value	CI
جنسیت (مرد در مقابل زن)	۱/۵۴	<۰/۰۰۰۱	(۱/۲۲, ۱/۹۶)
وضعیت تأهل (متأهل در مقابل مجرد)	۴/۱۳	<۰/۰۰۰۱	(۳/۱۱, ۵/۴۸)

هنگام استفاده از روش طبقه بندی نکات زیر را باید مد نظر قرار داد. طبقه بندی یک روش برای حذف اثر مخدوش کننده در زمان تجزیه و تحلیل آماری می باشد یعنی بعد از جمع آوری اطلاعات در زمان تحلیل هنگام پی بردن به حضور عامل مخدوش کننده در صورتی که اطلاعات مربوط به متغیرها به شکل جداگانه در تمام سطوح این عامل موجود باشد می توان با تحلیل در هر یک از سطوح، به وجود ارتباط واقعی پی برد. این روش معایبی دارد که استفاده از آن را محدود می کند که می توان به موارد زیر اشاره کرد: الف) به دلیل تحلیل های مجزا در سطوح مختلف عامل مخدوش کننده، حجم نمونه کاهش یافته و منجر به کم شدن توان آزمون می شود، ب) هنگامی که متغیر مخدوش کننده پیوسته باشد نمی توان از روش مذکور استفاده کرد، ج) در صورت وجود بیش یک متغیر مخدوش کننده استفاده از این روش باعث پیچیدگی مراحل تحلیل می شود.

مستقل را بر روی متغیر پاسخ مورد بررسی قرار دهیم می توان از تحلیل های چند متغیره استفاده نمود. این تحلیل ها در بیان روابط صحیح بین متغیرها به ویژه در مواقعی که متغیر مخدوش کننده وجود داشته باشد مناسب هستند. به عنوان مثال فرض کنید هدف مطالعه ای بررسی رابطه بین متغیرهای جنسیت و نتیجه یک آزمون (قبولی یا مردودی) می باشد؛ روش معمول مورد استفاده در این حالت، آزمون کای- دو است که با توجه به جدول ۳-الف ارتباط معنی دار بین دو متغیر را نتیجه نمی گیریم. بعد از بررسی بیشتر ثابت می شود که بین وضعیت تأهل و نتیجه آزمون ارتباط معنی دار وجود دارد، پس می توان نتیجه گرفت با احتمال زیاد متغیر وضعیت تأهل به عنوان یک متغیر مخدوش کننده، رابطه بین جنسیت و نتیجه آزمون را تحت تأثیر قرار داده است. ابتدا برای حذف اثر متغیر مخدوش کننده وضعیت تأهل، از روش طبقه بندی استفاده می کنیم یعنی با استفاده از آزمون کای- دو ارتباط بین دو عامل ذکر شده را به طور جداگانه در سطوح مختلف عامل مخدوش کننده وضعیت تأهل (مجرد و متأهل) بررسی می کنیم در این حالت با توجه به اطلاعات جدول ۳-ب ارتباط بین جنسیت و وضعیت قبولی معنی دار است و این بدان معنی است که جنسیت هم در افراد متأهل و هم در مجردها عامل مؤثری بر قبولی افراد بوده است.

جدول ۳-الف: بررسی ارتباط بین قبولی در امتحان با جنسیت در آزمون تک متغیره

	قبول	مردود
مرد	۳۰۰۰	۱۷۰
زن	۳۰۰۰	۱۴۱
Odds ratio=۱/۲۱		CI:(۰/۹۶, ۱/۵۲)
P-Value=۰/۱۱		

به روش آموزشی دوم تخصیص داده می شود حال ممکن است تعداد افراد تخصیص یافته در دو گروه یکسان نباشد مثلاً گروه اول و دوم به ترتیب ۶۰ و ۴۰ نفر را شامل شوند از آنجا که برابری حجم نمونه در گروه‌ها بیشترین توان آزمون را در پی خواهد داشت بنابراین این عدم توازن در تعداد نمونه، میتواند منجر به کاهش توان آزمون شود. این مسأله یکی از معایب تصادفی سازی به روش شیر و خط می باشد.

روش‌هایی نظیر استفاده از جدول اعداد تصادفی و بلوک جای‌گشت‌های تصادفی برای تخصیص افراد به گروه‌ها وجود دارد که این مشکل را بر طرف می سازند. با استفاده از روش بلوک جای‌گشت‌های تصادفی حجم نمونه در گروه‌های مختلف برابر می شود؛ اما ممکن است توزیع متغیرهای مخدوش کننده در گروه‌ها یکسان نباشد. به عنوان مثال فرض کنید خواستار مقایسه میانگین نمره امتحان پایان دوره دانشجویان در دو روش آموزشی باشیم. روش اول استفاده از وسایل کمک آموزشی و دیگری تدریس به روش رایج می‌باشد حال اگر با استفاده از روش بلوک جای‌گشت‌های تصادفی یا جدول اعداد تصادفی ۷۰ دانشجو که از لحاظ وضعیت اشتغال همزمان با تحصیل (شاغل، غیر شاغل) متفاوت هستند را به دو گروه تقسیم کنیم؛ اگر چه تعداد افراد در دو روش آموزشی یکسان خواهد شد اما ممکن است افراد شاغل در یک گروه بیش از گروه دیگر قرار بگیرند در صورتی که وضعیت اشتغال عامل اثر گذاری بر نمره باشد این مسأله نتایج مطالعه را تحت تأثیر قرار خواهد داد. این یکی از معایب روشهای مذکور است برای رفع این مشکل می توان از روش **Minimization** استفاده کرد. این تکنیک به گونه ای افراد را بین دو گروه تقسیم میکند که نسبت افراد شاغل در دو گروه یکسان شود تا اثر این متغیر کنترل شود. پس توجه داشته باشید در صورت استفاده از روشهای تصادفی سازی بنا بر شرایط مطالعه روشی را به کار برد که حداکثر توان آزمون را دارا باشد و توزیع متغیرهای مخدوش کننده در دو گروه یکسان شود.

با توجه به معایب فوق روش طبقه بندی مناسب ترین راه برای مهار اثر متغیر مخدوش کننده نیست. راه حل دیگر برای حل مشکل مذکور استفاده از روشهای **Multifactorial** از جمله رگرسیون لجستیک (**Logistic Regression**) می باشد. در این روش‌ها اثر متغیرها به صورت همزمان بررسی می شود. برای تحلیل مثال اخیر از طریق رگرسیون لجستیک، وضعیت قبولی را به عنوان متغیر وابسته و متغیرهای جنسیت و وضعیت تاهل را به عنوان متغیرهای مستقل وارد مدل می کنیم که نتایج این تحلیل در جدول ۳-ج آورده شده است. مشاهده میکنید که هم جنسیت و هم وضعیت تاهل روی قبولی افراد تاثیر داشته است (۱،۲،۱۴).

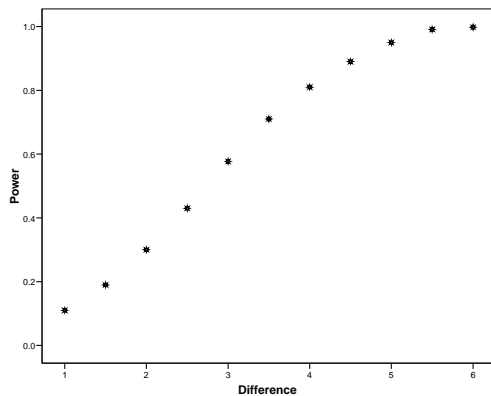
روشهای مختلفی برای حذف اثر متغیر مخدوش کننده وجود دارد بعضی از این روشها مانند تحلیل چند متغیره، هنگام آنالیز داده ها استفاده می شوند و برخی دیگر مانند جورسازی (**Matching**) و تصادفی سازی (**Randomization**) زمان طراحی مطالعه مورد استفاده قرار می گیرند که به دلیل اهمیت و کاربرد زیاد مبحث تصادفی سازی در ادامه به آن پرداخته می شود.

### تصادفی سازی

یک اصل مهم در مطالعات تجربی که بررسی وجود تفاوت بین گروه‌ها می‌پردازد تصادفی سازی است زیرا مزیت مطالعات تجربی از بین بردن اثر متغیرهای مخدوش کننده می‌باشد. وقتی که افراد بر اساس این روش به گروههای مختلف تقسیم شوند، قادر خواهیم بود ویژگی‌هایی که نتایج مطالعه را تحت تأثیر قرار داده و جز اهداف مطالعه نیستند (متغیرهای مخدوش کننده) را تعدیل کرده و بیشترین همگنی را در گروه‌ها ایجاد کنیم.

آسان ترین راه تصادفی سازی روش شیر و خط است فرض کنید هدف بررسی دو روش آموزشی بر نمره ۱۰۰ دانشجو باشد در این روش بعد از پرتاب سکه در صورت ظاهر شدن خط، دانشجو به روش آموزشی اول و در صورت مشاهده شیر،

## توان آزمون



نمودار ۱: اختلاف میانگین دو گروه در برابر توان آزمون

## ضریب توافق

یکی از مباحثی که در محافل آموزشی مطرح است ارزشیابی اساتید توسط خودشان و دانشجویان و بررسی توافق بین آنها میباشد. در این گونه مسایل محققین معمولاً از ضریب همبستگی پیرسون و یا آزمون تی زوجی استفاده می‌کنند. ولی هر دو این روش‌ها می‌تواند گمراه کننده باشد. برای روشن شدن موضوع به مثال‌های زیر توجه کنید. فرض کنید داده‌های جدول ۴-الف نمرات ارزیابی استاد از خودش و دانشجو از استاد باشد. در این حالت اگر ضریب همبستگی محاسبه شود مقدار آن  $0/631$  و مقدار  $P\text{-Value}$  برابر  $0/028$  است که نشان دهنده ارتباط نسبتاً قوی و مستقیم بین نمرات است. این در حالی است که تفاوت نمرات نشان دهنده آن است که نمرات اساتید به خودشان همواره حداقل به اندازه ۶ نمره بالاتر از دانشجویان بوده است (نمونه یک). حال فرض کنید در یک ارزشیابی دیگر نمرات به صورت جدول ۴-ب باشد (نمونه دو).

تعیین حجم نمونه از مسایلی است که باید در هر مطالعه ای مورد نظر قرار گیرد. فرض کنید بنا به دلایلی (کم بودن حجم جامعه مورد مطالعه یا پایین بودن هزینه سرشماری) به جای نمونه گیری از سرشماری استفاده شده است. در چنین مواقعی اگر اختلاف معنی داری مشاهده نشود ممکن است به دلیل پایین بودن توان آزمون (قدرت ما در کشف تفاوتی که واقعا وجود داشته است) باشد.

فرض کنید می‌خواهیم دو روش آموزشی را روی دو گروه از دانشجویان اعمال و نمرات دانشجویان را مقایسه کنیم. اگر اختلاف معنی داری بین دو گروه مشاهده نشود به طور قطعی نمی‌توان گفت که دو روش یکسان بوده اند. در این حالت باید توان آزمون را محاسبه کرد و اگر توان آزمون کم بود می‌توان نتیجه گرفت که دلیل عدم مشاهده اختلاف معنی دار، توان کم آزمون بوده است. در مثال فوق فرض کنید تعداد دانشجویان در دو کلاس به ترتیب ۲۳ و ۲۶ باشد. اگر میانگین و انحراف معیار نمرات در کلاس اول  $14/5 (\pm 4/1)$  و در کلاس دوم  $16/5 (\pm 5/6)$  باشد مقدار  $P\text{-Value} = 0/17$  به دست می‌آید که نشان دهنده عدم وجود اختلاف معنی دار بین دو روش آموزشی است. این در حالی است که توان این آزمون فقط  $0/3$  است.

نمودار شماره ۱ نشان دهنده رابطه بین توان آزمون و تفاوت میانگین در دو گروه میباشد وقتی که تعداد دانشجو و انحراف معیار ثابت نگه داشته شوند؛ مشاهده می‌شود که با افزایش اختلاف، توان آزمون نیز افزایش پیدا می‌کند و با کاهش اختلاف، توان کاهش می‌یابد و مبین این مسأله است که مطالعاتی که به صورت سرشماری انجام میشوند توان کمی برای کشف تفاوت‌های کوچک و توان بالاتری برای کشف تفاوت‌های بزرگ دارند (۱۵).



جدول ۴-الف: بررسی توافق بین ارزشیابی استاد از خود و دانشجو از استاد (نمونه یک)

نمره استاد	۲۰	۱۹	۱۸	۱۷	۱۹	۱۸	۱۶	۱۹	۲۰	۱۹	۱۸	۱۷
نمره دانشجو	۱۴	۱۳	۱۱	۱۰	۱۲	۱۱	۱۰	۱۲	۱۰	۱۱	۱۱	۱۰
تفاوت	۶	۶	۷	۷	۷	۷	۶	۷	۱۰	۸	۷	۷

جدول ۴-ب: بررسی توافق بین ارزشیابی استاد از خود و دانشجو از استاد (نمونه دو)

نمره استاد	۲۰	۱۵	۱۳	۱۹	۱۲	۱۸	۱۳	۱۹	۲۰	۱۴	۱۹
نمره دانشجو	۱۲	۱۹	۱۸	۱۱	۱۶	۱۲	۱۹	۱۱	۱۲	۱۹	۱۳
تفاوت	۸	-۴	۵-	۸	-۴	۶	-۶	۸	۸	-۵	۶

با در نظر گرفتن شرایط هر آزمون به تحلیل داده ها پرداخت. در مواردی که انتظار می رود متغیرهایی غیر از متغیرهای موجود بر ارتباط بین متغیرها اثر گذارند، باید با استفاده از تحلیل های چند متغیره اثر مخدوش کننده ها را حذف کرد چرا که گاهی اوقات ارتباطات واقعی با تحلیلهای تک متغیره قابل مشاهده نیستند. در تفسیر و ارائه نتایج باید علاوه بر **P-Value**، فاصله اطمینان نیز گزارش داده شود. در هنگام تخصیص افراد به گروههای مختلف توجه داشته باشید از روشهایی استفاده شود که افراد را به صورت تصادفی و با در نظر گرفتن اثر متغیرهای مخدوش کننده به گروهها تخصیص دهند.

به یاد داشته باشید که در انجام یک پژوهش تعیین حجم نمونه و توان نقش بسزایی دارند، در صورتی که حجم نمونه و توان آزمون درست انتخاب شده باشند نتایج به دست آمده در صورت انجام تحلیل مناسب مورد قبول خواهند بود، چرا که توان آزمون پایین باعث عدم کشف ارتباط معنی دار می گردد. در مورد بحث توافق ضریب همبستگی بالا بین دو اندازه کمی بیان کننده توافق داده ها نیست چرا که ممکن است داده های با توافق پایین، همبستگی بالایی داشته باشند.

حرف آخر آنکه با توجه به اهمیت آمار در پژوهشها، انتظار می رود که محققین قادر به انجام پژوهشهایی با روش تحلیل مناسب و صحیح و ارائه نتایج معتبر و قابل استناد جهت استفاده متخصصین، دانشجویان و مسئولین باشند که به منظور

اگر از آزمون دو نمونه زوجی برای بیان توافق داده های دو گروه استفاده شود مقدار **P-Value** برابر ۰/۳۷ می شود که نشان دهنده عدم تفاوت معنی دار در دو گروه است و مقدار ضریب همبستگی ۰/۸۷- است که همبستگی قوی بین نمرات را بیان می کند در حالی که با توجه به اطلاعات ارائه شده، هر جا که استاد به خود نمره بالاتری داده است دانشجو به او نمره کمتری داده است و برعکس.

برای سنجش توافق با استفاده از نمودار بلاند و آلتمن اختلاف بین دو اندازه در مقابل میانگین دو اندازه رسم می شود. در صورت وجود توافق انتظار داریم که اختلاف اندازه ها به صفر نزدیک باشد. در این حالت اگر نقاط حول خطی که روی محور  $Y=0$  متمرکز باشند توافق برقرار است (۱۶، ۱۷). توضیحات بیشتر در رفرنسهای ارائه شده موجود است.

## بحث و نتیجه گیری

آنچه که درانتشار یک مقاله حائز اهمیت است بهره گیری از روشهای صحیح آماری و ارائه نتایج مناسب و مورد نیاز آن پژوهش می باشد. توصیف داده ها با استفاده از شاخصها و نمودارهای مناسب به صورت یک اصل مهم باید مورد توجه قرار گیرد، توجه به نوع متغیر مورد بررسی می تواند در ارائه توصیف صحیح داده ها کمک کننده باشد. در مرحله تحلیل داده ها ابتدا باید شرایط استفاده از آزمونهای مناسب را مد نظر قرار داده و پس از آن از آزمونها متناسب با هدف پژوهش

به جامعه استفاده شود.

ارائه نوآوری های پزشکی و عرضه خدمات بهداشتی و مراقبتی

#### References:

1. Cehrei A, Haghdoost AA, Fereshtehezad M, Bayat A. [Data analysis in medical research using SPSS software]. 1<sup>st</sup> ed. Tehran: Karimi; 2010. [In Persian]
2. Nasirian M, Sadeghi M, Haghdoost AA. [Principals in data analysis and presentation of results]. Journal of Isfahan University of Medical Sciences 2009; 100: 646-59. [In Persian]
3. Bahrapour A. [Inference in Biostatistics]. Kerman: Vadiat; 2002. [In Persian]
4. Mohammad K, Malekafzali H, Nehaptian W. [Statistical methods and health indicators]. 12th ed. Tehran: Salman; 2004. [In Persian]
5. Ayatollahi M.T. [Principals and Methods in Biostatistics]. Tehran: Sepehr; 2000. [In Persian]
6. Streiner DL. Maintaining standards: Differences between the standard deviation and standard error, and when to use each. Can J Psychiatry 1996; 41(8): 498-502.
7. Altman DG, Bland JM. Standard deviations and standard errors. BMJ 2005; 331(7521): 903.
8. Baena-Díez JM, Byram AO, Grau M, Gómez-Fernández C, Vidal-Solsona M, Ledesma-Ulloa G. Obesity is an independent risk factor for heart failure: Zona Franca cohort study. Clin Cardiol 2010; 33(12): 760-4.
9. Gardner MJ, Altman DG. Confidence intervals rather than P values: Estimation rather than hypothesis testing. BMJ 1986; 292(6522): 746-50.
10. Davies HTO, Crombie IK. What are confidence intervals and p-values? London: Hayward Medical Communications; 2009.
11. Altman DG. Why we need confidence intervals. World Journal of Surgery 2005; 29(5): 554-6.
12. Drezner Z, Turel O, Zerom D. A modified Kolmogorov-Smirnov test for normality. Munich Personal RePEc Archive; 2008.
13. Chan YH. Biostatistics 102: Quantitative data-parametric & non-parametric tests. Singapore Medical Journal 2003; 44(6): 280-5.
14. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 1996; 49(8): 907-16.
15. Motulsky HJ. Prism 4 statistics guide: Statistical analyses for laboratory and clinical researchers. San Diego; GraphPad Software Inc; 2003.
16. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 307-10.
17. Chan YH. Biostatistics 104: correlational analysis. Singapore Med J 2003; 44(12): 614-9.