

شماره: ۸۰۰۵
تاریخ: ۱۳۹۸/۰۳/۲۲
پیوست: ندارد

بسم تعالیٰ

هفتادمین سال تاسیس دانشگاه فردوسی مشهد کرامی بااد

کواہی می شود جناب آقای دکتر محمدی جباری

در اولین سمینار استباط شواهدی که در تاریخ ۲۲ خرداد ۱۳۹۸ در دانشگاه فردوسی مشهد برگزار گردید،

شرکت نموده و مقاله ای با عنوان

"استباهات رایج آماری"

را با همکاری خانم نسرین تلمذی به صورت سخنرانی ارایه داده اند.

با آرزوی توفیق روز افزون

دیر علمی سمینار

دکتر سید محمود طاهری

دیر علمی سمینار
دکتر محمدی جباری

اولین سمینار استباط شواهدی

خرداد ماه ۱۳۹۸

isei1@um.ac.ir

شانی الکترونیکی:
<http://isei1.um.ac.ir/>



نشانی:
مشهد، میدان آزادی
پردیس دانشگاه فردوسی
دانشکده علوم ریاضی
صندوق پستی: ۹۱۷۷۵-۱۱۵۹
تلفن: ۰۵۱-۳۸۸۰۶۲۱۳
بست الکترونیکی:

اشتباهات رایج آماری

"تحقیقات خوب مستحق ارائه خوب هستند، و ارائه خوب علاوه بر جمع آوری و تجزیه و تحلیل داده‌ها به عنوان بخش مهمی از تحقیق به شمار می‌رود. زمانی که یک نوشه دیده می‌شود، خوب نوشته شدن آن تشخیص داده می‌شود، اجازه دهد ما نیز تشخیص دهیم که علم حق دارد به خوبی نوشته شود."

۱. مهدی جباری نوقابی^۱

گروه آمار، دانشگاه فردوسی مشهد، مشهد، ایران

۲. نسرین تلخی^۲

گروه اپیدمیولوژی و آمار زیستی، دانشگاه علوم پزشکی مشهد، مشهد، ایران

چکیده

امروزه محققان شاهد وجود اشتباهات آماری رایجی، در بسیاری از مقالات علمی بوده‌اند. به منظور اینکه نتایج ارائه شده در مقاله‌های علمی صحیح و قبل اتکا باشد، پیش از چاپ در نشریه‌های تخصصی مورد داوری قرار می‌گیرند اما با این حال در برخی از مقاله‌های پذیرفته و چاپ شده در نشریه‌های علمی-پژوهشی اشتباهات مختلفی مشاهده شده است که باعث می‌شود از اعتبار نتایج گزارش شده در مقالات کاسته شود. در این مقاله به معرفی برخی از اشتباهات رایج آماری که محققان در تجزیه و تحلیل و نوشتمن مقالات علمی دچار می‌شوند، پرداخته شده است. امید است نتایج حاصل از این مقاله در به کار بردن صحیح روش‌های آماری در مقالات علمی نقش مهمی داشته باشد. توصیه می‌شود افرادی که قصد دارند در این موارد مطالعه بیشتری داشته باشند، از متون مفصل‌تر و همچنین سایر منابع ذکر شده کمک بگیرند.

مقدمه

محققان نرخ بالایی از اشتباهات آماری را در تعداد زیادی از مقالات علمی حتی در بهترین ژورنال‌ها مشاهده کردند. در واقع، مشکل گزارش‌های ضعیف آماری، بسیار گسترده و به صورت بالقوه، جدی و شناخته نشده هستند، با وجود اینکه بیشتر اشتباهات، مربوط به مفاهیم پایه آماری هستند اما می‌توان با داشتن درک صحیحی از روش‌های آماری و مفاهیم مربوطه، از به کار بردن نادرست آن‌ها در تحقیقات علمی اجتناب ورزید. به منظور اینکه نتایج ارائه شده در مقاله‌های علمی صحیح و قبل اتکا باشد، پیش از چاپ در نشریه‌های تخصصی مورد داوری قرار می‌گیرند اما با این حال در برخی از مقاله‌های پذیرفته شده و چاپ شده در نشریه‌های علمی-پژوهشی مشاهده شده است که اشتباهات مختلفی وجود دارد که باعث می‌شود از اعتبار نتایج گزارش شده در مقالات کاسته شود. بنابراین لازم هست تا محققین مبتدی از اشتباهات رایجی که در زمان برنامه‌ریزی، تجزیه و تحلیل داده‌ها، روش تحقیق و نوشتمن یک مقاله علمی ممکن است با آن مواجه شوند، آگاهی یابند که این امر مستلزم این است که قبل از شروع تحقیق و نگارش مقاله خود بر تمامی مراحل مطالعه سلط کافی داشته باشند و همچنین با فنون روش تحقیق و مفاهیم آن

¹ E-mail: jabbarinm@um.ac.ir; jabbarinm@yahoo.com

² E-mail: talkhin961@mums.ac.ir; nasrin.talkhi70@gmail.com

آشنایی کامل داشته باشند. محققان می‌توانند در دو زمینه مرتکب اشتباه شوند، هر دو زمینه‌ای که مطرح خواهد شد، می‌توانند صحت نتایج گزارش شده در متون را تحت تأثیر قرار دهند. اولین زمینه، اشتباهاتی هستند که در طول فرایند تحقیق رخ می‌دهند (مانند اشتباهات در برنامه‌ریزی و یا اجرای مطالعه) و دومین زمینه، مواردی هستند که محقق به هنگام تجزیه و تحلیل داده‌ها، تفسیر و ارائه نتایج (در قالب یک مقاله یا پوستر...) مرتکب می‌شود.

نکته‌ای که در اینجا حائز اهمیت می‌باشد و می‌بایست بدان اشاره نمود، این هست که اغلب اوقات خطاهای برنامه‌ریزی و اجرای مطالعه توسط یک ویرایشگر مجله که فقط نسخه نهایی متن نگارش شده را مشاهده می‌کند، قابل شناسایی و تشخیص نمی‌باشند. از این رو هر مطالعه بایستی با بیشترین دقیقت و توجه به جزئیات انجام شود، زیرا خطاهای عمدی و غیرعمدی که در طول فرآیند تحقیق اتفاق می‌افتد، زیان‌بار خواهد بود و از این رو بیشترین اهمیت را دارا می‌باشند. آمار را می‌توان به دو حوزه گسترده تقسیم نمود: آمار توصیفی که مربوط به نحوه توصیف نمونه‌هایی از داده‌های جمع‌آوری شده در یک تحقیق است و آمار استنباطی که مربوط به نحوه برآورد (یا استنباط) ویژگی‌های جامعه از نمونه‌ای که انتخاب شده است. در روش‌های آمار استنباطی (کلاسیک یا بیزی) که مبتنی بر منطق لم نیمن پیرسون هستند نمی‌توان هر نوع تفسیری که محقق می‌خواهد را از نتایج آمار استنباطی داشت و استفاده نمود. بنابراین اگر محقق بخواهد تفسیرهای متفاوتی در مورد استنباط آماری داشته باشد بهتر است که از روش مناسب دیگری تحت عنوان استنباط شواهدی استفاده نماید. استنباط شواهدی در عرصه تفاسیر منطقی از آمار کلاسیک پیش است و در مقابل تفسیر با نظام استقرایی آمار کلاسیک ابزاری قوی در اختیار دارد که مشکل آن را حل می‌کند. مقاله حاضر به بررسی اشتباهات رایجی که به واسطه‌ی زمینه دوم اتفاق خواهد افتاد می‌پردازد که در ادامه مورد بررسی قرار می‌گیرند.

1- روش‌های آمارگیری

روش‌های آمارگیری به دو دسته سرشماری و نمونه‌گیری تقسیم می‌شوند. به منظور برآورد یک مشخصه در جامعه می‌توان از سرشماری یا نمونه‌گیری استفاده کرد. سرشماری از جامعه متناهی، بررسی‌ای است که تمام واحدهای جامعه را در بر می‌گیرد و باید گفت در بسیاری از موارد، اجرای سرشماری در یک جامعه متناهی، کاری شدنی است. به طور مثال در موسسه‌ای که 50 نفر کارمند دارد می‌توان با بررسی پرونده‌های استخدامی آن‌ها سال‌ها خدمات تمام آن‌ها را فهرست کرد که در واقع یک سرشماری انجام شده است. وقتی حجم جامعه متناهی و بزرگ باشد انجام سرشماری به دلیل وقت‌گیر و پرهزینه بودن غالباً عملی نیست. بنابراین برای بررسی یک مشخصه به نمونه‌گیری متولّ می‌شوند. نمونه بخشی از جامعه تحت بررسی است که با روشنی که از پیش تعیین شده است انتخاب می‌شود، به قسمی که می‌توان از این بخش، استنباطهایی درباره کل جامعه بدست آورد. تعیین روش آمارگیری سرشماری یا نمونه‌گیری یا انواع مختلف روش‌های نمونه‌گیری می‌تواند بر داده‌های بدست آمده و نتایج تجزیه و تحلیل تاثیر بسزایی بگذارد، به عبارتی دیگر اگر در یک مطالعه شرایط نمونه‌گیری تصادفی طبقه‌ای برقرار باشد بدیهی است که انتخاب نمونه از هر روش دیگری مثلاً خوش‌های نماینده واقعی جامعه را بدست نخواهد داد، به عنوان مثال اگر بخواهیم در مورد متوسط درآمد شهروندان شهر مشهد تحقیقی انجام بدهیم و دسته‌بندی افراد جامعه در شهر مشهد را بر حسب مناطق شهرداری درنظر بگیریم، در هر منطقه شهرداری افرادی با کمترین درآمد و بیشترین درآمد زندگی می‌کنند، بنابراین شرایط نمونه‌گیری تصادفی خوش‌های برقرار هست، واضح است که در اینجا اگر به روش نمونه‌گیری تصادفی طبقه‌ای نمونه انتخاب شود، نماینده واقعی جامعه را به دست نخواهد داد.

2- روش های تعیین حجم نمونه

در انجام هر تحقیق علمی، موضوع تعیین حجم نمونه بسیار اهمیت دارد و نتایج بدست آمده از تحقیق بستگی به انتخاب یک نمونه مناسب با حجم کافی خواهد داشت. حداقل حجم نمونه مورد نیاز باید براساس اهداف اصلی تحقیق برآورده شود. پژوهش‌ها و مطالعات کاربردی بر اساس اهداف اصلی به دو دسته، مطالعات معطوف به مقدار (مطالعات توصیفی) و مطالعات معطوف به تصمیم (مطالعات توصیفی-تحلیلی) تقسیم می‌شوند. یکی از راه‌های برآورده حجم نمونه در پژوهش‌های معطوف به مقدار، استفاده از روش فاصله اطمینان است. در این روش به دلیل اینکه در بسیاری از موارد از تقریب نرمال برای برآورده حداقل حجم نمونه مورد نیاز استفاده می‌شود، طبیعی است که حجم نمونه کافی بزرگ باشد. در مطالعات معطوف به تصمیم، علاوه بر توصیف، بدنبال استنتاج در مورد اهداف، فرضیه‌ها یا سوالات مطرح شده نیز هستیم. در این مطالعات حداقل حجم نمونه مورد نیاز، بایستی براساس آزمون فرضیه‌های مورد استفاده تعیین شود. در این صورت، چنانچه با توجه به مقادیر مفروض احتمال خطاهای نوع اول و دوم و سایر اطلاعات مربوط به جامعه و آزمون فرضیه‌ها، حداقل حجم نمونه مورد نیاز برآورده گردد، امکان بزرگ بودن احتمال رخدادن یکی از خطاهای تصمیم گیری نسبت به دیگری وجود نخواهد داشت. چنانچه حجم نمونه کمتر از میزان لازم در نظر گرفته شود، ممکن است نتایج استنباط شده از آن در مورد جامعه از دقت کافی برخوردار نباشد و امکان تعمیم نتایج حاصل از نمونه به جامعه وجود نداشته باشد. بسیاری از پژوهشگران در علوم مختلف جهت تعیین حجم نمونه مورد نیاز برای انجام تحقیق‌های خود از جدول مورگان استفاده می‌کنند. باید گفت استفاده از این جدول برای هر هدفی نادرست است. در واقع برای هر روش آماری تجزیه و تحلیل داده‌ها، یک روش تعیین حجم نمونه وجود دارد. کتاب تکنیک‌های نمونه‌گیری از ویلیام جی. کوکران دارای فرمول‌های زیادی جهت محاسبه حجم نمونه مورد نیاز و هر کدام مختص موضوع خاصی می‌باشد و فرمول کوکران که بسیار مشهور است و اکثراً استفاده می‌کنند تنها یکی از فرمول‌های این کتاب است، که این فرمول مربوط به برآورده حجم نمونه در برآورده نسبت می‌باشد. بنابراین نمی‌توان برای هر هدفی در مطالعات از فرمول کوکران استفاده نمود. توصیه می‌شود با توجه به هدف و روش آماری موردنیاز جهت پاسخ به فرضیه‌های تحقیق از نرم افزارهای R، NCSS & PASS یا G-Power برای تعیین حجم نمونه مناسب گردد.

استفاده مناسب

3- تناسب حجم نمونه با حجم جامعه

معمولًا تصور می‌شود که هر چه حجم جامعه بزرگتر باشد حجم نمونه نیز باید به همان نسبت بیشتر و مناسب با بزرگی جامعه باشد. ولی البته چنین نیست. یکی از تصورات شهودی نادرست این است که اگر (به عنوان مثال) از جامعه‌ای دارای 10,000 واحد، نمونه‌ای به حجم 50 لازم است، از جامعه‌ای دارای 100,000 واحد، نمونه‌ای به حجم 500 نیاز هست. در واقع جزء اولین دستاوردهای آمار استنباطی (آمار مبتنی بر حساب احتمالات) این بود که این درک شهودی نادرست است. به عنوان یک مثال ساده فرض کنید می‌خواهیم نسبت P را با دقت d و با ضریب اطمینان $(\alpha - 1)$ از جامعه‌ای که دارای N واحد است، برآورده کنیم.

حجم نمونه لازم از فرمول

$$n = \frac{z^2 P(1 - P)/d^2}{1 + (z^2 P(1 - P)/d^2)/N}$$

به دست می‌آید، که در آن Z عددی از جدول توزیع نرمال می‌باشد. جدول زیر مقادیر حجم نمونه لازم را به ازای $d=0.7$ ، $P=0.5$ و $1 - \alpha = 0.95$ ، مقادیر مختلف حجم جامعه به دست می‌دهد:

∞	1000000	100000	10000	1000	500	100	10	N
196	196	196	192	164	141	66	10	n
0	0.0196	0.196	1.92	16.4	28.2	66	100	$(n/N)*100$

با استفاده از نتایج بدست آمده مشاهده می‌شود که هر چه حجم جامعه بزرگتر می‌شود حجم نمونه مورد نیاز به همان نسبت بیشتر نخواهد شد و از یک مرحله‌ای به بعد حجم نمونه لازم به ازای بزرگتر شدن حجم جامعه، ثابت باقی می‌ماند. بنابراین عباراتی از قبیل "نمونه ای به حجم 10٪ جامعه برای دقت مورد نظر کافی است" بدون اینکه اطلاعی از حجم جامعه داشته باشیم بسیار گمراه کننده است و نباید بکار برد.

4- مشکل حجم نمونه بسیار زیاد

در میان پژوهشگرانی که از روش‌های آماری برای رفع نیازهای علمی حیطه تخصصی خود بهره می‌گیرند، این برداشت نادرست وجود دارد که هر چه حجم نمونه بزرگتر باشد (یا به حجم جامعه نزدیکتر باشد)، نتایج بهتری حاصل خواهد شد. اما باید توجه داشت که همواره در نمونه‌های با حجم بزرگ، تضمینی برای اتخاذ تصمیم منطبق بر واقعیت وجود ندارد .

اکثر کسانی که تجربه‌ای در تحلیل داده‌های آماری دارند می‌دانند یکی از مسائل اساسی که به دنبال انتخاب نمونه‌ای با حجم بزرگ، ممکن است حادث شود، این است که توان آزمون‌های آماری مورد استفاده در مطالعات تحلیلی، بسیار افزایش می‌یابد، که این موضوع در عمل باعث می‌شود فرضیه صفر به نادرستی رد شود و شанс این که نتایج "معنی‌دار" شوند بیشتر است. به عبارتی دیگر با توجه به اینکه معمولاً سطح معنی‌داری برابر 0.05 و گاهی هم (به ندرت) برابر 0.01 در نظر گرفته می‌شود، زمانی که حجم نمونه بسیار بزرگ است، احتمال پذیرش نادرست فرضیه H_0 به مراتب از 0.01 و امثال آن کمتر است. پس راه حل چیست؟ راه حل معقول این است که هیچگاه به بدست آوردن یک نتیجه معنی‌دار اکتفا نکنیم بلکه همراه با انجام آزمون، تفاوت میانگین دو جامعه (یا مقدار ضریب همبستگی) را نیز برآورد کنیم و در صورتی که آن را ناچیز تشخیص دهیم، نتیجه را غیر معنی‌دار اعلام نماییم. به عنوان مثال محققی، زمانی نمرات ریاضی دانش‌آموزان دو ناحیه آموزش و پرورش را مقایسه کرده و تفاوت میانگین نمرات ریاضی دو ناحیه را بسیار بسیار معنی‌دار یافته و از ناهماهنگی آموزش ریاضی در دو ناحیه به شدت نگران شده بود. اما بعد از اینکه متوجه شد تفاوت میانگین دو نمونه کمتر از 0.03 بوده است و معنی‌داری فوق العاده این تفاوت به خاطر وجود چند هزار دانش‌آموز در هر گروه بوده است، نگرانی‌اش برطرف شد.

زمانی که از p -مقدار جهت رد یا پذیرش فرضیه صفر استفاده می‌شود، باید گفت p -مقدار در واقع بیان می‌کند که آیا نتایج در اثر شناس رخ داده است یا خیر و به خودی خود نمی‌تواند ملاک خوبی برای نشان دادن معنی‌دار بودن یا نبودن رابطه باشد و همیشه باید با اندازه اثر (بزرگی و جهت تفاوت، زمانی که دو گروه مقایسه می‌شوند) و فاصله اطمینان 95٪ برای تفاوت (فاصله اطمینان حدودی را می‌دهد که ما انتظار داریم مقدار واقعی جامعه در آن محدوده قرار بگیرد) همراه باشد. بنابراین p -مقدار، اندازه اثر و فواصل اطمینان برای اندازه اثر باید در کنار هم، برای رسیدن به نتایج معنی‌دار، گزارش شوند. همچنین زمانی که حجم نمونه بسیار زیاد باشد، ممکن است سوالاتی از جمله اینکه 1- آیا وجود یا عدم وجود همبستگی بین دو متغیر یا وجود یا عدم وجود تفاوت بین دو میانگین جامعه به این بستگی دارد که ما چه حجم نمونه‌ای به کار ببریم؟ 2- آیا اگر حجم نمونه ما به اندازه کافی بزرگ باشد، می‌توانیم از معنی‌دار شدن نتایجمان اطمینان داشته باشیم؟ در ذهن محقق شکل بگیرد که در پاسخ به سوالات مطرح شده باید گفت گرچه جواب سوال اول مسلمان (و عقلاءً) منفی می‌باشد ولی جواب سوال دوم متأسفانه مثبت است. اما باید دید که

چرا چنین است؟ دلیل آن بسیار ساده است و آن این واقعیت است که در عمل (در مورد متغیرهای پیوسته) هیچگاه تفاوت دو میانگین جامعه (یا ضریب همبستگی دو متغیر تصادفی) صفر نیست. حجم نمونه کوچک قادر به کشف تفاوت‌های کوچک (ضریب همبستگی‌های نزدیک صفر) نیست، در صورتی که حجم نمونه بزرگ به راحتی کوچک‌ترین تفاوت‌ها (کمترین ضرایب همبستگی) را ظاهر می‌سازد. البته باید به این موضوع نیز اشاره نمود که یک اندازه نمونه کوچک لزوماً باعث یک مطالعه ضعیف نمی‌شود بلکه توانایی تعمیم دادن و کشف نتیجه درباره جمعیت مورد نظر دشوارتر می‌شود. همچنین، با اندازه نمونه‌های کوچک، باید همواره مراقب بود که در مورد قدرت شواهد موجود اغراق نشود یا نتیجه‌ای فراتر از آنچه که مشاهده شده، گرفته نشود.

5- جامعه هدف، جامعه در دسترس

جامعه هدف^۳ جامعه‌ای است که نتایج به دست آمده از نمونه بایستی به آن تعمیم داده شود. باید اذعان داشت دسترسی به جامعه هدف به صورت مستقیم امکان‌پذیر نخواهد بود. در واقع محقق زیرگروهی از جامعه هدف را که به آن جامعه مورد مطالعه^۴ یا جامعه منبع^۵ گفته می‌شود، در اختیار دارد و در این راستا بایستی نماینده^۶ گویایی از جامعه هدف باشد. نمونه‌ای^۷ که جهت انجام تحقیقات جمع‌آوری می‌شود، از جامعه مورد مطالعه حاصل می‌شود که این نمونه بایستی بگونه‌ای انتخاب شود که وضعیت جامعه مورد مطالعه را به خوبی به تصویر بکشد.

موردی که بایستی بدان اشاره نمود این هست که گرفتن نمونه از بخشی از جامعه و تعمیم آن به کل جامعه بسیار متدائل است که این کار اغلب نادرست بوده و اگر آگاهانه انجام شود، تقلب محسوب می‌شود. ولی تحت بعضی شرایط نه تنها اشکالی ندارد بلکه تنها راه انجام دادن پژوهش آماری موردنظر است. مثلاً گرفتن نمونه از شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران که در مشهد مستقرند، برای برآورد میزان سود کل شرکت‌های عضو بورس، کار درستی نیست. یا به عنوان مثالی دیگر گرفتن نمونه از کارمندان، برای برآورد میزان محبوبیت کاندیدایی که مدافعان حقوق زنان است، کار درستی نیست. ولی گرفتن نمونه از دانشآموزان ابتدایی شهر مشهد و بررسی تاثیر تنبیه بدنه بر پیشرفت تحصیلی و تعمیم نتایج به کل کلان شهرهای ایران ممکن است قابل توجیه باشد. در مثال فوق، نتایج را می‌توان به دانشآموزان ابتدایی کلان شهرها در سال‌های آینده نیز تعمیم داد. در واقع دلیل انجام چنین تحقیقی فقط می‌تواند امکان همین تعمیم اخیر باشد. تحت شرایط مناسب، حتی می‌توان نتایج پژوهشی را که در مورد کارگران یک کارخانه انجام شده است، به همه کارخانه‌های مشابه در کشور تعمیم داد. معمولاً پژوهش در مورد پارامترهای دو متغیری (نظیر استقلال، همبستگی و رگرسیون) از یک زیرجامعه به کل جامعه بلا اشکال است، چون، برخلاف پارامترهای یک متغیری، روابط بین متغیرها از زیرجامعه به جامعه خیلی تغییر نمی‌کنند. یکی از اشتباهات دیگری که خیلی رایج است وقتی پیش می‌آید که زیرجامعه (جامعه در دسترس) کاملاً سرشماری می‌شود. در این موارد پژوهشگر با استناد به این که تمام جامعه سرشماری شده است، از هر گونه تحلیل آماری سر باز می‌زند و به آمار توصیفی بسنده می‌کند. در صورتی که جامعه اصلی (جامعه هدف) بسیار بزرگ‌تر است و برای آن نیاز به استنباط داریم.

³ target population

⁴ study population

⁵ source population

⁶ representative

⁷ sample population

6- عدم استفاده از میانگین در شرایطی که صحیح نیست

میانگین راحت‌ترین معیاری است که مرکزیت داده‌ها را نشان می‌دهد ولی همواره مناسب‌ترین شاخص نیست. می‌دانیم که در جامعه‌های با توزیع شدیداً چوله، استفاده از میانگین برای نشان دادن متوسط جامعه مناسب نیست. برای مثال، جمله "میانگین پاداش آخر سال کارمندان دولت 290 هزار تومان است"، با توجه به این که تعداد کمی وزیر و مدیر ارشد با پاداش بسیار زیاد و تعداد زیادی کارمند جزء با پاداش بسیار کم وجود دارد، تصویر نادرستی از توزیع پاداش‌ها به خواننده می‌دهد. اشتباه دیگری که در پژوهش‌های آماری رایج‌تر است این است که بعضی اقدام به میانگین‌گیری از مقادیر متغیرهایی می‌کنند که در مقیاس ترتیبی یا اسمی هستند. می‌دانیم که در مقیاس ترتیبی، برخلاف مقیاس‌های فاصله‌ای و نسبتی، فواصل مساوی دارای ارزش یکسان نیستند و لذا میانگین‌گیری از مقادیر آن‌ها منطقی نیست. البته در این موارد (مقیاس رتبه‌ای) استفاده از میانه اشکال ندارد.

7- همبستگی‌های گمراه کننده

یکی دیگر از اشتباهات رایجی که در استفاده از مفاهیم آماری وجود دارد و بایستی بدان پرداخت، مبحث همبستگی است که در این قسمت به مفاهیمی چون همبستگی جعلی، استنباط علیت معکوس از همبستگی، همبستگی صوری و همبستگی بسیط یا تعدیل نشده پرداخته می‌شود.

7.1- همبستگی جعلی

به عنوان مثال فرض کنید در یک پژوهش، محققی همبستگی بین "عيار چغندرهای کرت A" و "عيار چغندرهای کرت B" را محاسبه و اعلان کرده است. برای کرت A از بذر (الف) و برای کرت B از بذر (ب) استفاده کرده است. تنها وجه تشابه (ارتباط) دو کرت در این است که هر یک دارای 36 بوته است.

مثال دیگر، محاسبه همبستگی بین "درآمد حسابداران مرد" و "درآمد حسابداران زن" در شرکت‌های خصوصی است. چنین همبستگی‌هایی نه تنها قابل محاسبه نیستند، بلکه غیرمنطقی بوده و قابل تعریف نیز نمی‌باشند.

برای تعریف (و محاسبه) همبستگی دو متغیر، لازم است آن دو متغیر روی واحدهای آماری یک مجموعه (نمونه آماری) تعریف شده باشند. مثل همبستگی بین وزن و عیار چغندرهای یک کرت. (واحد آماری = چغندر) یا مثل همبستگی بین درآمد حسابداران یک بیمارستان با درآمد همسرانشان. (واحد آماری = خانواده)

7.2- استنباط علیت معکوس از همبستگی

در بحث مباحث همبستگی، گرفتن نتیجه علیتی به راحتی امکان‌پذیر نیست به طور مثال، آیا از این که "رتیه گروه‌های آمار دانشگاه‌ها با رتبه دانشگاه‌ها همبستگی قابل توجهی دارد"، می‌توان نتیجه گرفت که کیفیت یک دانشگاه به کیفیت گروه آمار آن بستگی دارد؟ یا این که درست برعکس است و واقعیت این است که دانشگاه‌های خوب، گروه‌های آمار با کیفیت دارند؟ متأسفانه باید بیان نمود آمار هیچ راهی برای پاسخ دادن به این سوال ندارد. تنها زمانی که دو متغیر تقدم و تاخر زمانی داشته باشند تشخیص علت و معلول ساده است در غیر این صورت نمی‌توان از رابطه همبستگی بین دو متغیر استنباط علیتی داشت.

7.3- همبستگی صوری

بسیار پیش می‌آید که همبستگی نسبتاً بالایی بین دو متغیر مشاهده می‌شود، در حالی که هیچ یک بر دیگری تاثیر ندارد، بلکه وجود متغیر سومی در میان است. برای مثال همبستگی بالایی بین جویدن آدامس و ارتکاب تخلف‌های رانندگی مشاهده می‌شود، در صورتی که هیچ یک علت یا معلول دیگری نیست، بلکه بی‌قراری و عصبی مزاج بودن فرد، علتی برای هردو محسوب می‌شود و با هر دو همبستگی واقعی دارد.

7.4- همبستگی بسیط یا تعدیل نشده

گاهی تاثیر چندین متغیر مستقل بر متغیر وابسته، به خاطر وجود همبستگی بین خود آن‌ها، به سادگی قابل تفکیک نیست. مثال واقعی زیر از مجله Statistician چاپ آمریکا سال 1982 نقل می‌شود. بر اساس ارقام مربوط به

X =فروش سالانه مواد دخانی (سیگار، سیگار برگ و توتون پیپ) و Y =تعداد موارد گزارش شده از سرطان ریه در سال.

رشد پیوسته و یکنواخت فروش سالانه مواد دخانی از سال 1880 تا سال 1980 میلادی، و همراه با آن رشد پیوسته و یکنواخت تعداد سالانه موارد سرطان ریه و در نتیجه همبستگی بسیار بالایی (برابر با 0.98) بین X و Y مشاهده می‌شود. اما این یک همبستگی ساده (بسیط) است و همبستگی واقعی (خالص) زمانی به دست می‌آید که تاثیر سایر متغیرهایی که در همین بازه زمانی دارای رشد بوده، و بر Y نیز موثر بوده‌اند، را حذف کنیم. این کار به کمک محاسبه همبستگی جزئی (Partial Correlation) X و Y امکان‌پذیر است. در طول دوره فوق، متغیرهای دیگری نیز بر Y موثر بوده‌اند که همگی نیز در این مدت در حال رشد بوده‌اند، نظیر افزایش فاضلاب شیمیایی کارخانه‌جات، تشعشعات رادیویی، تلویزیونی، میکروویو، آزمایش‌های هسته‌ای، استرس حاصل از افزایش شهرنشینی، افزایش استفاده از مواد افزودنی و نگهدارنده و بسیاری از متغیرهای سرطان‌زای دیگر، که متأسفانه ارایه فهرست کاملی از آن‌ها، امکان‌پذیر نیست. پس از تعدیل ضریب همبستگی به وسیله حذف اثرات متغیرهای نام برده شده، ضریب همبستگی جزئی (واقعی، خالص) بین X و Y برابر 0.46 به دست می‌آید، که البته باز هم کم نیست ولی با مقدار اولیه 0.98 تفاوت زیادی دارد.

8- بروز اشتباه در اثر ادغام داده‌ها

این مطلب را با ذکر یک مثال شروع می‌کنیم، بدین صورت که در اواسط جنگ جهانی دوم، کارمندی در وزارت دفاع آمریکا متوجه شد که سهم زنان در تک تک صنایع آمریکا، نسبت به قبل از جنگ، افزایش یافته است، در حالی که سهم زنان در کل صنایع آمریکا، نسبت به قبل از جنگ، کاهش یافته است! مثال عددی زیر امکان بروز تعارض‌هایی از این دست را ثابت می‌کند:

سطح محافظه‌کاری

پایین			بالا			سطح کیفیت اطلاعات حسابداری
186	120	بالا	71	42	پایین	
257	162	جمع				
٪72	٪74					درصد با کیفیت بالا

گرچه تفاوت در صد با کیفیت بالا در دو گروه (سطح محافظه‌کاری) معنی‌دار نیست، ولی ارقام فوق با تصور قبلی در مورد اثر مثبت داشتن سطح کیفیت اطلاعات حسابداری بر سطح محافظه‌کاری مغایرت دارد. اما زمانی که اطلاعات به تفکیک نوع شرکت (خدماتی و غیرخدماتی) جمع‌آوری شود،

سطح محافظه‌کاری

پایین	بالا	سطح کیفیت اطلاعات
6	30	حسابداری
1	6	(شرکت خدماتی)
7	36	جمع
/.85	/.83	در صد با کیفیت بالا
180	90	سطح کیفیت اطلاعات
70	36	حسابداری
250	126	(شرکت غیرخدماتی)
/.72	/.71	در صد با کیفیت بالا

"تغییر مقدار یا تغییر جهت همبستگی در اثر ادغام داده‌ها" را تعارض سیمپسون (Simpson's paradox) می‌نامند. توالی همبستگی صوری، همبستگی تعديل نشده و تعارض سیمپسون در این متن تصادفی نبوده است، زیرا دو مورد قبلی حالات خاصی از مورد سوم هستند.

9- اطمینان از درستی نتیجه یک آزمون مبتنی بر p -مقدار

اگر آزمونی را در سطح (مثال) 0.05 انجام دهید و فرضیه H_0 رد نشود (پذیرفته شود)، نمی‌توانید ادعا کنید که 95٪ به نتیجه آزمون اطمینان دارید. ببینیم چرا؟ عدد 0.05 در بالا احتمال خطای نوع اول است، یعنی آزمون طوری انجام شده است که احتمال رد کردن H_0 ، در صورتی که H_0 درست باشد، 0.05 است. لذا اگر H_0 رد می‌شد، یک نوع "اطمینان" نسبی به درستی H_1 وجود داشت، ولی H_0 رد نشده (پذیرفته شده) است و میزان "اطمینان" به درستی H_0 به احتمال خطای نوع دوم بستگی دارد که معمولاً معلوم نیست و در مواردی که فرضیه H_1 مرکب است، مقدار مشخصی ندارد. لذا در این شرایط، اطمینانی به صحت پذیرش H_0 نمی‌توان داشت. حتی زمانی که H_0 رد می‌شود، گرچه یک "اطمینان" نسبی به درستی H_1 وجود دارد، اما نمی‌توان گفت که "احتمال درستی H_1 برابر 95٪ است" زیرا احتمال درستی H_1 روی فضای $\{H_0, H_1\}$ تعریف می‌شود، در صورتی که احتمال رد H_0 روی فضای متغیر تصادفی مشاهده شده (X) تعریف می‌شود. و این دو فضای احتمال، کاملاً متفاوت‌اند. در واقع اگر پژوهشگر به دنبال این باشد که ببیند داده‌ها چقدر از فرضیه H_0 یا از H_1 پشتیبانی می‌کنند، باید از روش‌های استنباط شواهدی استفاده نماید.

10- نداشتن مشاهده در تمام سطوح یک متغیر

محققی ضمن تحقیق در مدارس کودکان استثنایی اصفهان، مشاهده کرده بود که والدین 31٪ از کودکان عقب مانده ذهنی، خویشاوند هستند و نتیجه گرفته بود که این بررسی آماری آنچه را که علم ژنتیک به آن رسیده است (یعنی تاثیر سوء ازدواج خویشاوندی بر سلامت ذهنی کودکان) تایید می‌کند. این دامی است که خیلی‌ها در امثال آن گرفتار می‌شوند (غفلت از دیدن روی دیگر سکه). این محقق چون 31٪ به نظرش درصد بالایی از ازدواج خویشاوندی آمده است، چنین استنباطی کرده است. در صورتی که اگر می‌دانست درصد ازدواج خویشاوندی در کل جامعه نیز نزدیک به همین مقدار است (صرف نظر از این که چنین تاثیری واقعاً وجود دارد یا نه) چنین نتیجه‌هایی نمی‌گرفت. یا (مثلاً) اگر درمی‌یافتد که در کل جامعه این درصد برابر 35٪ است نتیجه‌های کاملاً بر عکس می‌گرفت. لذا برای نتیجه‌گیری‌هایی از این قبیل، باید به درصد ازدواج خویشاوندی در هر دو گروه (کسانی که ازدواج خویشاوندی داشته‌اند و کسانی که نداشته‌اند) دسترسی داشته باشیم. به طور کلی، استنباط‌هایی که بر اساس یک یا بخشی از سطوح یک متغیر انجام می‌شود ناقص بوده و می‌تواند گمراх کننده باشد.

11- استفاده نکردن از روش‌های چند متغیری

در بسیاری از مطالعات، چندین متغیر روی هر واحد نمونه، اندازه‌گیری می‌شود. این گونه داده‌ها را داده‌های چند متغیری می‌نامند. مقایسه میانگین‌های این متغیرها در دو (یا چند) نمونه را باید با روش‌های تحلیل چند متغیری انجام داد. گاهی داده‌های چند متغیری را یکی‌یکی و با استفاده از روش‌های یک متغیری بررسی می‌کنند. این کار به معنی دور ریختن بخش مهمی از اطلاعات، که در ساختار کوواریانس متغیرها نهفته است، می‌باشد، که گاه منجر به نتایج نادرست می‌شود. برای مثال فرض کنید اندازه‌گیری قد و وزن 64 نفر به نتایج زیر منجر شده است:

$$\bar{X} = (68, 172) \quad , \quad S = \begin{bmatrix} 112 & 55 \\ 55 & 114 \end{bmatrix} \quad , \quad \begin{cases} H_0: (\mu_1, \mu_2) = (\mu_{10}, \mu_{20}) = (66, 174) \\ H_1: (\mu_1, \mu_2) \neq (66, 174) \end{cases}$$

از آزمون چند متغیری T^2 هتلینگ استفاده می‌کنیم، که نتیجه می‌شود:

$$T^2 = n(\bar{X} - \mu_0)'S^{-1}(\bar{X} - \mu_0) = 10.84 > 10.09 * \frac{(n-1)p}{(n-p)} F_{0.01}(2, 64)$$

مشاهده می‌شود که فرضیه صفر آزمون رد خواهد شد. حال فرض کنید که کوواریانس مشاهده شده را ندیده گرفته و به جای آزمون چند متغیری، دو آزمون یک متغیری برای فرضیه‌های $H_0: \mu_1 = 66$ و $H_1: \mu_2 = 174$ انجام دهیم. نتیجه می‌شود:

$$|T_1| = \frac{|68 - 66|}{\sqrt{\frac{112}{64}}} = 1.42 < 2 \quad , \quad |T_2| = \frac{|172 - 174|}{\sqrt{\frac{114}{64}}} = 1.88 < 2$$

که بر اساس این محاسبات هیچ یک از دو فرضیه صفر رد نمی‌شوند. بنابراین نباید آزمون‌های تک متغیری را جایگزین آزمون‌های چند متغیره کرد و به جای هم به کار برد.

12- عدم تشخیص لانه‌ای بودن طرح مطالعه

طرح آزمایشی در مطالعات آزمایشی فوق العاده اهمیت دارد. برای اینکه بخواهیم در هر طرح اندازه‌گیری روش آماری مناسب را انتخاب کنیم بایستی به ماهیت آن طرح توجه زیادی داشته باشیم. چون در خیلی از موارد ممکن است یک طرح در واقع لانه‌ای باشد ولی با استفاده از روش‌های آنالیز واریانس بلوکی اثر تیمار بررسی شود، یا ممکن است طرح لانه‌ای باشد اما از طرح کرت خرد شده استفاده شود و یا بالعکس. به عنوان مثال برای مقایسه سود 4 محصول شیمیایی (a, b, c, d) از 3 کارخانه (A, B, C)، در هر کارخانه از هر یک از محصولات 3 نمونه تصادفی گرفته و در هر نمونه میزان سود اندازه‌گیری شده است. اگر دقت کافی لحاظ نشود، ممکن است برای تحلیل داده‌ها از یک طرح دو عاملی استفاده شد، که به جدول آنالیز واریانس زیر منجر می‌شود:

منبع تغییرات	SS	df	MS	F	sig
کارخانه‌ها	15.1	2	7.5	1.02	0.38
مواد	25.6	3	8.6	3.24	0.038
اثر متقابل	44.3	6	7.4	2.80	0.033
خطا	43.3	24	2.6		

ملحوظه می‌شود که سود مواد به طور معنی‌داری متفاوت است. به علاوه اثر متقابل (معنی‌دار) به عامل غیر معنی‌دار کارخانه‌ها ربط داده شده و می‌تواند تحلیل‌گر را به این نتیجه‌گیری سوق دهد که کارخانه‌ها واقعاً متفاوت‌اند ولی اثر آن‌ها توسط اثر متقابل پنهان شده است. اما با کمی دقت متوجه می‌شویم که باید داده‌ها را حاصل یک طرح لانه‌ای ببینیم. ماده شیمیایی تولید شده توسط کارخانه A همان ماده شیمیایی تولید شده توسط کارخانه B یا C نیست. بنابراین اثر متقابل بین کارخانه و مواد معنی ندارد. مواد شیمیایی هر کارخانه منسوب به همان کارخانه است و نه به کارخانه‌های دیگر. اصطلاحاً، عامل مواد در داخل عامل کارخانه لانه کرده است. با استفاده از طرح لانه‌ای به جدول آنالیز واریانس زیر می‌رسیم:

منبع تغییرات	SS	df	MS	F	sig
کارخانه‌ها	15.1	2	7.5	1.02	0.38
مواد	69.9	9	7.8	2.24	0.056
خطا	43.3	24	2.6		

ملحوظه می‌شود که با اعمال روش صحیح تفاوت بین سود مواد شیمیایی معنی‌دار نیست. بنابراین بایستی در انتخاب و تشخیص طرح اندازه‌گیری دقت شود و نتایج واقعیت وضعیت موجود را نشان دهنند.

13- عدم تشخیص طرح اندازه‌گیری‌های مکرر

به طور کلی می‌توان گفت هرگاه اثر K تیمار روی یک متغیر پاسخ، مورد تحقیق باشد، باید توجه داشت که آیا آن K تیمار در K نوبت متوالی روی یک مجموعه از افراد (یا اشیاء) اعمال می‌شود مثل اجرای یک طرح کلان اقتصادی (مثلاً هدفمندی یارانه‌ها) در کشور در 3 دوره زمانی مختلف و مجزا یا روی K مجموعه مختلف از افراد (یا اشیاء) اعمال می‌شود. مثل اعمال سه طرح کلان

اقتصادی برای 3 دسته روزتاهای شهرها و کلان شهرها. در مورد اول یک طرح اندازه‌گیری‌های مکرر داریم و در مورد دوم با طرح آنالیز واریانس یک طرفه مواجه هستیم. حالت خاص اندازه‌گیری‌های مکرر وقتی است که $K=2$ باشد، که در آن صورت نام طرح "مقایسه میانگین‌های زوج شده" است، که آن نیز گاهی با طرح "مقایسه میانگین‌ها بر اساس دو نمونه مستقل" اشتباہ می‌شود. زمانی که متغیر پاسخ نیز یک متغیر کیفی است، باید داده‌های طرح مقایسه‌های زوج شده را با "آزمون تقارن در جدول‌های توافقی (مک نمار، کاپا یا کاپا وزنی یعنی آزمون Bowker)" تحلیل کنیم. این طرح نیز اغلب با طرح آزمون استقلال در جدول‌های توافقی "اشتباه می‌شود. در این حالت وقتی $K=2$ است، باید از آزمون مک نمار استفاده شود.

14- استفاده از روش آماری نامنطبق بر طرح اندازه‌گیری

اغلب محققان در مطالعاتی که می‌خواهند تاثیرگذار بودن یا نبودن یک صفت مداخله را روی یک صفت بررسی نمایند و اندازه‌گیری‌ها در قبل و بعد از مداخله انجام می‌شود، از آنالیز کواریانس استفاده می‌کنند، بدین صورت که مقادیر صفت در قبل از مداخله را به عنوان کوواریوت و بعد از مداخله را به عنوان متغیر پاسخ در نظر می‌گیرند. باید گفت این مطالعات نوع خاصی از آنالیز واریانس اندازه‌های مکرر نوع II (هم اثر درون‌گروهی و هم اثر بین‌گروهی وجود داشته باشد) می‌باشد و استفاده از آنالیز کواریانس در چنین مسائلی صحیح نمی‌باشد. برای روشن شدن این موضوع و دلایل نادرست بودن استفاده از آنالیز کواریانس به جای طرح آنالیز واریانس اندازه‌های مکرر نوع II، سه طرح زیر را در نظر بگیرید، فرض کنید تعدادی از افراد را براساس حجم نمونه‌ای که تعیین شده است در اختیار دارید، این افراد را می‌توان به طور تصادفی به 2 گروه یکی گروه کنترل و دیگری گروه آزمایش تقسیم نمود. بر روی افراد در گروه کنترل هیچ مداخله‌ای انجام نمی‌شود اما در گروه آزمایش مداخله انجام می‌شود. سپس صفت موردنظر (که باید مستقیماً اندازه‌گیری شود) مثلاً میزان HDL افراد در هر دو گروه کنترل و آزمایش اندازه‌گیری می‌شود. این طرح را طرح شماره 1 می‌نامیم. این دو گروه جدا و مستقل از هم می‌باشند. اگر هدف بررسی تاثیر مداخله بر میزان HDL افراد باشد، باید آزمونی را که برای مقایسه میانگین‌های گروه‌های مستقل (دو یا چند گروه) استفاده می‌شود، بکار برد. گاهی شرایطی پیش می‌آید که نمی‌توان افراد را به دو گروه تقسیم نمود مثلاً محدودیتی که وجود دارد این است که تعداد افراد زیادی در دسترس نیست (مثلاً دو یا سه نفر در دسترس هستند)، طرح شماره 2 به این صورت طراحی می‌شود که قبل از اینکه افراد در یک دوره تمرینی شرکت کنند تست HDL از آن‌ها گرفته شود سپس مداخله روی افراد انجام شود و پس از انجام مداخله مجدد از آنها تست HDL گرفته شود. در این طرح باید از آزمون‌هایی برای مقایسه دو گروه زوجی استفاده شود، چون افراد در دو گروه یکی هستند. اگر دو طرح شماره 1 و 2 را با هم مقایسه کنیم خواهیم دید منطق حاکم بر مقایسه این دو طرح با هم فرق می‌کند. در طرح 1 باید همه افراد گروه کنترل را با همه افراد گروه آزمایش مقایسه شوند اما در طرح شماره 2 پس آزمون هر فرد با پیش آزمون خودش مقایسه می‌گردد. البته طرح شماره 1 حالت کلی تری نسبت به طرح شماره 2 می‌باشد. مشکل بزرگی که در طرح شماره 1 وجود دارد این است که ممکن است افرادی که به تصادف در گروه کنترل نسبت به گروه آزمایش قرار گرفتند وضعیت HDL بهتری را قبل از اعمال مداخله داشتند. پس یکی از دلایلی که ممکن است گفته شود مداخله بر HDL تاثیر داشته یا نداشته در واقع همگن نبودن دو گروه می‌باشد. در طرح شماره 2 هم گفته می‌شود که هر فردی با خودش مقایسه شده و ویژگی‌های فردی در این کار دخالت داشته است و ممکن است به این دلایل تاثیر مداخله معنی دار شده یا نشده است. بنابراین هریک از این دو طرح یک مشکل اساسی دارند. حال ترکیبی از این دو طرح را در نظر بگیرید و طرح شماره 3 بنامید. در طرح شماره 3 ابتدا افراد به طور تصادفی به دو گروه تقسیم می‌شوند و از هر دوی این گروه‌ها پیش آزمون گرفته می‌شود و سپس در گروه آزمایش مداخله موردنظر و در گروه

کنترل مداخله‌نما (طرح‌های یکسو کور و دوسو کور و سه‌سو کور هم در اینجا وارد بحث‌های کارآزمایی بالینی نمی‌شون) اعمال می‌گردد. سپس از هر دو گروه پس آزمون گرفته می‌شود، حال در اینجا از یک دیدگاه دو گروه زوجی داریم و از یک دیدگاه دو گروه مستقل داریم. در این طرح اگر بخواهیم اثر مداخله را بررسی نماییم، باید چکار کرد و چه آزمونی انجام داد؟ ۵ آزمون می‌توان انجام داد، پیش آزمون دو گروه را با هم و پس آزمون دو گروه با هم مقایسه شود (دو آزمون تی مستقل)، پیش آزمون و پس آزمون گروه کنترل با هم و پیش آزمون و پس آزمون گروه آزمایش با هم مقایسه شود (دو آزمون تی زوجی)، آزمون دیگر اینکه اختلاف پیش آزمون و پس آزمون در دو گروه با هم مقایسه شود (یک آزمون تی مستقل). واضح است که انجام این ۵ آزمون درست نمی‌باشد، یک دلیل این می‌باشد که سطح معنی‌داری (سطح خطای) زیاد می‌شود (البته این مورد قابل رفع می‌باشد یعنی به جای ۰.۰۵ با ۰.۰۱ مقایسه شود). اما ایراد اساسی که وجود دارد این است که در هر آزمون تی که انجام می‌شود اثر گروه‌های غایب در نظر گرفته نمی‌شود. زمانی که اختلاف پیش آزمون و پس آزمون با هم مقایسه می‌شود در واقع اثر پیش آزمون یا به عبارتی اثر مداخله از بین می‌رود. در طرح شماره ۳ اثرات باید همزمان در نظر گرفته شود.

$$(Y_{pre}, Y_{post}) = \mu + \tau_i + \varepsilon_{ij}$$

در این طرح وقتی اثر مداخله معنی‌دار است که

گروه کنترل و گروه آزمایشی در پیش آزمون با هم فرق نکنند اما در پس آزمون با هم فرق کنند.

پس آزمون نسبت به پیش آزمون در گروه کنترل فرق نکند اما در گروه آزمایشی فرق کند.

که البته اگر چند گروه و چند زمان وجود داشته باشد تفاسیر به مراتب متفاوت خواهد بود. اما در اینجا حالت ساده با دو گروه و دو زمان در نظر گرفته شده است. عده‌ای از افراد از آنالیز کواریانس برای تحلیل چنین مسائلی استفاده می‌کنند که باید اظهار نمود استفاده از این روش کاملاً نادرست می‌باشد. در اینجا لازم است قبل از بیان دلایل نامناسب بودن استفاده از آنالیز کواریانس، ابتدا به تعریف متغیر هم‌عامل پرداخته شود. هم‌عامل متغیری کمی است که متفاوت از متغیر وابسته و هم‌جنس نیستند و از طرفی با افزایش یا کاهش متغیر هم‌عامل متغیر وابسته نیز افزایش یا کاهش می‌یابد.

دلیل اول اینکه استفاده از آنالیز کواریانس بجای آنالیز واریانس اندازه‌های مکرر اشتباه می‌باشد این است که متغیر پیش آزمون و پس آزمون از یک جنس هستند و این با تعریف متغیر هم‌عامل در تنافض است.

دلیل دوم اینکه که در آنالیز واریانس واریانس متغیر وابسته به دو یا چند مولفه تجزیه می‌شود اما در آنالیز کواریانس، کواریانس بین متغیر وابسته و هم‌عامل تجزیه می‌شود یعنی در واقع ماتریس کواریانس به دو یا چند مولفه تجزیه می‌شود. در طرح شماره ۳ همبستگی بین پیش آزمون و پس آزمون اصلاً مهم نیست، بلکه اختلاف پیش آزمون و پس آزمون مهم است که باید در گروه آزمایشی و نه در گروه کنترل معنی‌دار باشد.

دلیل سوم اینکه طرح‌ها یا آینده‌نگر هستند یا گذشته‌نگر. طرح شماره ۳ یک طرح آینده‌نگر می‌باشد بنابراین باید به این نکته هم توجه کرد که وقتی طرح موردنظر آینده‌نگر باشد روش آماری نمی‌تواند گذشته‌نگر باشد در حالیکه آنالیز کواریانس برای یک مطالعه گذشته‌نگر می‌باشد و متغیر هم‌عامل و متغیر وابسته باید همزمان اندازه‌گیری شده باشند نه اینکه یکی در یک زمان و دیگری در یک زمان دیگر یا چند روز بعد اندازه‌گیری شده باشند. بنابراین وقتی طرح موردنظر یک طرح آینده‌نگر باشد روش

آماری هم باید قطعاً آینده‌نگر باشد و روش‌های آینده‌نگر هم مطالعات طولی هستند. آنالیز کواریانس هم طولی نیست بلکه مقطعی است. آنالیز واریانس اندازه‌های تکراری حالت خاص وساده‌ای از مطالعات طولی است.

نتیجه‌گیری

در نهایت چنین نتیجه می‌شود که وجود برخی اشتباهات رایج آماری در مقالات علمی حتی بهترین مجله‌ها به دلیل عدم تسلط کافی محققان بر مراحل انجام تحقیق و عدم آشنایی کامل آنها با فنون و مفاهیم مربوط به روش‌های آماری می‌باشد و باعث می‌گردد نتایج گزارش شده مورد اعتماد و قابل اتكا نباشد، از این رو به محققان توصیه می‌شود خطاهای و اشتباهاتی که ممکن است در طی فرآیند تحقیق و یا به هنگام تجزیه و تحلیل، تفسیر و ارائه نتایج رخ دهنده را تا حد ممکن در نظر گرفته و در جهت رفع آن‌ها بکوشند.

1. اولین و مهمترین بخش فرآیند علمی آمار، موضوع تعیین حجم نمونه می‌باشد که از اهمیت بسیار بالایی برخوردار می‌باشد. به تبع تعیین حجم نمونه، روش‌های آمارگیری سرشماری و انواع روش‌های نمونه‌گیری مطرح می‌شوند که در نتایج حاصل تاثیر بسزایی خواهند داشت. طرح آمارگیری نمونه‌ای باید با توجه به خصوصیات، چارچوب جامعه آماری و همچنین اهداف تحقیق تعیین شود و هر طرح آمارگیری نمونه‌ای را نمی‌توان برای هر جامعه آماری به کار برد. طبیعتاً باید دید که شرط‌های مربوط به طرح‌های آمارگیری نمونه‌ای چه هستند و اگر آن شرایط بر اساس هدف وجود داشته باشند از طرح مناسب استفاده کرد و در شرایطی که لازم هست سرشماری نیز صورت گیرد. پس از مشخص نمودن طرح آمارگیری، در صورت نیاز به نمونه‌گیری، بایستی حجم نمونه لازم براساس نوع مطالعه و اهداف تحقیق تعیین شود زیرا حجم نمونه خیلی بزرگ یا خیلی کوچک در صحت نتایج تاثیر گذاشته و ما را از واقعیت دور می‌کند.

2. مسئله دیگری که باید به آن توجه نمود این است که روش‌های آمار توصیفی و آمار استنباطی به صورت نادرست، به جای هم استفاده نشوند. در برخی شرایط در صورتی که جامعه در دسترس، کاملاً سرشماری می‌شود، تنها به ارائه آمار توصیفی بسته نشود و از روش‌های آمار استنباطی جهت پی بردن به اطلاعات و شناخت جامعه سرشماری شده استفاده شود، زیرا جامعه هدف بسیار بزرگتر می‌باشد و دسترسی به آن امکان‌پذیر نمی‌باشد. جامعه‌ای که کاملاً سرشماری شده است زیرگروهی از جامعه هدف می‌باشد که در دسترس محقق بوده است و در مرحله بعد برای جامعه هدف نیاز به استنباط داریم.

3. در تفسیر نتایج به دست آمده از روش‌های آمار استنباطی دقت لازم به عمل آید و از تفسیر نادرست آنها اجتناب شود یا به عبارتی دیگر تفاوت بین آزمون فرضیه‌های آماری و آزمون‌های معنی‌داری به خوبی درک و ارتباط آنها با استنباط شواهدی در نظر گرفته شود. به عنوان مثال در حجم نمونه‌های بالا همواره فرضیه صفر در مقابل فرضیه مقابله‌ردمی شود و از این رو بایستی p -مقدار همراه با اندازه اثر و فاصله اطمینان برای اندازه اثر جهت رد یا عدم رد فرضیه صفر گزارش شود. ضمن این که وقتی گفته می‌شود فرضیه صفر رد می‌شود یا رد نمی‌شود فقط یک اصطلاح است و هیچ آزمون آماری درستی یا نادرستی فرضیه صفر را تضمین نمی‌کند. در آمار کلاسیک از p -مقدار برای استنباط شواهدی استفاده می‌شود که برای این کار مناسب نیست و باید از روش‌های آمار شواهدی استفاده نمود زیرا هیچ یک از معایبی که p -مقدار به آن دچار است را ندارد، بلکه دارای مزیت‌هایی است که p -مقدار چنین مزایایی را ندارد.

4. در تشخیص طرح اندازه‌گیری و روش آماری مبتنی بر آن طرح، تلاش لازم صورت پذیرد و از به کار بردن روش آماری نامناسب یا نامنطبق بر طرح اندازه‌گیری اجتناب گردد. به عنوان مثال، دقت شود که طرح‌های لانه‌ای، آنالیز واریانس بلوکی، طرح‌های کرت خرد شده و یا ... به اشتباه به جای هم به کار برده نشوند. زیرا در نتایجی که به دست می‌آید به شدت تاثیرگذار و این نتایج فاقد اعتبار خواهند بود. موضوع دیگری که بسیار مهم است و اغلب محققین در استفاده از آن دچار اشتباه می‌شوند مربوط به استفاده از آنالیز کواریانس به جای طرح آنالیز واریانس اندازه‌های مکرر در مطالعاتی که مقادیر یک صفت به صورت قبل و بعد از مداخله جمع‌آوری می‌شود، می‌باشد. باید تعریف دقیق متغیر هم عامل را بشناسند و از پیش آزمون به عنوان متغیر هم عامل استفاده نکنند. خصوصاً اینکه مطالعات قبل و بعد حالت خاصی از آنالیز واریانس اندازه‌های مکرر (یک مطالعه طولی آینده‌نگر) می‌باشد و نباید به جای آن از آنالیز کواریانس که یک مطالعه مقطعی گذشته‌نگر می‌باشد، استفاده نمود.

منابع :

1. ارقامی، ن. ر.؛ 1374، بعضی اشتباهات رایج در تحقیقات علوم تربیتی و اجتماعی، گردهم آیی اعضاء شوراهای تحقیقات کشور، تبریز
2. برومیده، ع. ا.؛ 1384، با برخی از اشتباهات رایج در تحلیل‌های آماری آشنا شویم، اندیشه آماری، سال نهم شماره اول
3. ارقامی، ن. ر.؛ 1380، سنجیری فارسی پور، ن. و بزرگنیا، ا.، مقدمه‌ای بر بررسی‌های نمونه‌ای، انتشارات دانشگاه فردوسی
4. اسماعیلی، ح.؛ 1378، اشتباهات رایج آماری در پژوهه‌های تحقیقاتی، اولین سمینار دانشجویی آمار ایران، دانشگاه شهید بهشتی
5. خردمندی، م.؛ 1377، عواقب نادیده گرفتن کواریانس‌ها، اندیشه آماری، سال سوم شماره اول
6. مشکانی، م. ر.؛ 1370، آمار مقدماتی، جلد اول، چاپ دوم، مرکز نشر دانشگاهی
7. موسوی، س. ج.؛ سرمهد، م.؛ 1380، فریب‌های آماری، پژوهه کارشناسی آمار
8. ارقامی، ن. ر.؛ زمانزاده، ا.؛ راحتی، س.؛ 1388، شواهدآماری (1)، اندیشه آماری، سال 13 شماره اول
9. ارقامی، ن. ر.؛ دست برآورده، ع.؛ زمانزاده، ا.؛ 1389، شواهدآماری (2)، اندیشه آماری، سال 13، شماره دوم
10. جباری نوقابی، ه.؛ جباری نوقابی، م.، نکاتی در برآورد حجم نمونه و معرفی نرم افزار مربوطه، ندا، شماره دوم، سال چهارم، 21-13
11. عمیدی، ع.؛ 1381، روش‌های نمونه گیری 1، چاپ چهارم، انتشارات دانشگاه پیام نور
12. ذوق دار مقدم، م.؛ 1389، نقش استنباط شواهدی در تشخیص توزیعهای نرمال آمیخته، پایان نامه کارشناسی ارشد، رشته آمار ریاضی
13. ارقامی، ن. ر.؛ زمانزاده، ا.؛ 1389، آزمون فرضیه، آزمون معنی داری و اندازه گیری شواهد آماری، دهمین کنفرانس آمار ایران
14. Zinsmeister AR, Connor JT. Ten common statistical errors and how to avoid them. *The American journal of gastroenterology*. 2008;103(2):262-6.
15. Lang, Tom. "Twenty statistical errors even you can find in biomedical research articles." *Croatian medical journal* 45 4 (2004): 361-70 .
16. Lang T. Common statistical errors even you can find, I errors in descriptive statistics and in interpreting probability values2012.