

مبانی نظری و فلسفی آزمون فرضیه آماری به زبان ساده

دکتر علی محمد امیر تاش

استاد واحد علوم و تحقیقات دانشگاه آزاد اسلامی

تاریخ دریافت: ۹۱/۰۵/۲۳

تاریخ پذیرش: ۹۱/۰۸/۱۶

چکیده

هدف: هدف از پژوهش حاضر، جستجو در مبانی نظری آزمون فرضیه‌های آماری و نگاهی به مفاهیم منطقی و فلسفی و همچنین نتایج عملی این رویکرد علمی در تولید دانش نو بوده است.

روش‌شناسی: بدین منظور، بدنه ادبیات علمی موجود در مبانی، راهکارها و کاربردها، و فنون محاسباتی آزمون فرضیه برای دسترسی به رابطه‌های علت و معلولی و همبستگی بررسی شده‌اند. در پایان، استنتاج‌های منطقی و فلسفی که عمدتاً برداشت‌های نویسندگان از آن‌ها به کمک حدود چهار سال تدریس و تجربه در این وادی بوده، بیان شده است.

یافته‌ها: نتایج نشان داد که استنباط‌های آماری همیشه سه نوع اشتباه را دربردارد: اشتباه نوع اول، اشتباه نوع دوم و سایر اشتباهاتی که در فرایند مطالعه ممکن است برای محقق اتفاق بیفتد (انحراف استاندارد یا معیار). یافته‌های آمار استنباطی هنگامی ارزشمندند که کمترین میزان اشتباه و خطا را داشته باشند. برای رسیدن به این هدف، میزان احتمال اشتباه نوع اول را پژوهشگر کنترل می‌کند، همچنین اشتباه نوع دوم، با افزایش شدت متغیر مستقل، افزایش تعداد آزمودنی و استفاده از روش‌ها یا آمار پارامتری به جای غیرپارامتری کنترل می‌شود. نهایتاً، برای کاهش اشتباه یا انحراف استاندارد راه‌های مختلفی در علوم آمار و روش‌های پژوهش وجود دارند. قابل توجه است که خطاهای تصمیم‌گیری‌های آماری نوع اول و دوم با یکدیگر رابطه معکوسی دارند، از این رو هر کدام به بهای افزایش دیگری کاهش پیدا می‌کند. ضمناً از دیدگاه کاربردی، اشتباه نوع اول یافته‌های بدون فایده و اثر را به جامعه توصیه می‌کند، درحالی‌که اشتباه نوع دوم جامعه را از اطلاعات مفید و اثربخش محروم می‌سازد.

نتیجه‌گیری: قطعاً با اشتباهات پژوهشگر، جامعه از هر دو سو دچار خسارت می‌شود. انتخاب پژوهشگر بین دو نوع اشتباه اول را میزان اهمیت کاربرد نتیجه تعیین می‌کند.

واژه‌های کلیدی: استنباط آماری، توانایی آزمون، انواع اشتباه تصمیم‌گیری آماری، اشتباه استاندارد، انحراف استاندارد.

مقدمه

منطق و فلسفه وجودی امور مادی جهان از دو حالت خارج نیست: یا وجود دارند و یا وجود ندارند. کار پژوهش در آزمون فرضیه نیز تابع این دو حالت است. یعنی، اثر یا همبستگی بین متغیرها یا تأیید می‌شود و یا نمی‌شود. به عنوان نمونه، می‌توان قبول کرد و به تجربه هم ثابت می‌شود که بین قد و وزن رابطه وجود دارد، ولی عملکرد ورزشی با ساختار کف اقیانوس به احتمال زیاد بی‌رابطه است، و یا اینکه نوعی تمرینات ورزشی آمادگی جسمانی را تغییر می‌دهد یا تأثیری بر آن ندارد. بر مبنای این منطق و واقعیت، اثر و رابطه یا وجود دارد و یا وجود ندارد. نبودن اثر یا رابطه را در محث آمار با فرضیهٔ پوچ و بودن آن را با فرضیهٔ یک یا مقابل نشان می‌دهند. در شرایطی که فرضیهٔ پوچ صحیح هم باشد، قابل اثبات است که در تکرارهای بسیار زیاد اندازه‌های اثر یا رابطه دقیقاً مساوی و مشابه به دست نمی‌آیند. از حاصل این تفاوت‌ها توزیعی ساخته می‌شود که به آن توزیع "فرضیهٔ پوچ در شرایطی که صحیح است" می‌گویند.

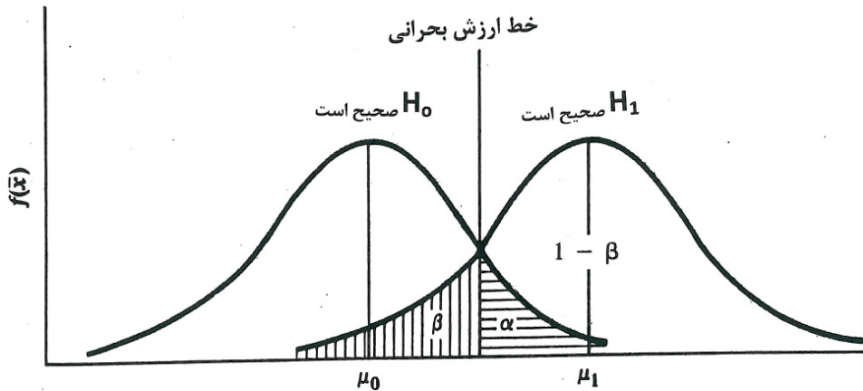
از آنجا که این توزیع از تفاوت‌هایی که همگی تصادفی بوده‌اند ساخته شده است، به آن توزیع احتمالات نیز گفته می‌شود. تمام اعداد بحرانی که ملاک رد و یا قبول اثر یا رابطه (همبستگی) هستند، از چنین توزیعی که به این صورت برای آن‌ها ساخته شده است ارزش‌های t , F , r و سایر شاخص‌های آماری مورد نظر استخراج می‌شوند.

در شرایطی که فرضیهٔ مقابل صحیح است (یعنی فرضیهٔ پوچ غلط است)، مجدداً از همهٔ مواردی که اثر یا رابطهٔ واقعی وجود دارد، بر حسب مورد، توزیعی ساخته می‌شود که به آن "توزیع فرضیهٔ مقابل در شرایطی که صحیح است" می‌گویند.

به این ترتیب، در طبیعت و به صورتی که خداوند آفریده است برای همهٔ اثرها و روابط فقط دو نوع توزیع وجود دارد: مواردی که فرضیهٔ پوچ صحیح است و مواردی که فرضیهٔ مقابل صحیح است. هر دوی این توزیع‌ها از نوع "توزیع احتمالات" هستند و به همین منظور مورد استفاده قرار می‌گیرند (شکل ۱).

می‌دانیم هر توزیع احتمالات که از تعداد بسیار زیاد آمارهٔ معین به دست می‌آید، دارای شکل طبیعی است و در این توزیع می‌توان بین درصدهای زیر منحنی، که جمع آن‌ها صد در صد یا یک است، با ارزش‌های آمارهٔ سازندهٔ محور ایکس‌ها رابطه برقرار کرد، یعنی، با در دست داشتن یکی از آن‌ها دیگری را به دست آورد. مثلاً چنانچه آمارهٔ مورد نظر عدد استاندارد "Z" که روی محور ایکس‌ها قرار دارد باشد، برای هر عدد "Z" می‌توان حجم زیرمنحنی را به دو بخش بیشتر و کمتر تقسیم کرد، یا اینکه مشخص کرد بین دو عدد "Z" چند درصد از حجم زیرمنحنی احتمالات قرار می‌گیرد، و یا درصدهای بالاتر و یا پایین‌تر از آن‌ها چه میزان هستند. مثلاً می‌دانیم که شصت و هشت درصد از فضای زیرمنحنی طبیعی بین مثبت و منفی یک انحراف استاندارد (معیار) "Z" قرار

دارد، شانزده درصد از آن بیشتر و شانزده درصد هم از آن کمترند. برای سایر آماره‌ها نیز وضعیت دقیقاً به همین صورت است و آماره "Z" فقط به عنوان نمونه آورده شده است.



شکل ۱- خطاهای نوع اول و دوم و توانایی آزمون در یافتن اثر یا ارتباط

از سوی دیگر می‌دانیم که هر منحنی احتمالات از منهای بی‌نهایت تا مثبت بی‌نهایت ادامه (کشیدگی) دارد، یعنی دامنه‌های منحنی هیچ‌گاه محور ایکس‌ها را قطع نمی‌کنند (به صفر نمی‌رسد)؛ به این صورت است که در توزیع احتمالات مربوط به صحیح بودن فرضیهٔ پوچ، هیچ موردی، حتی اگر بسیار استثنایی (بزرگ) هم باشد، نمی‌تواند از آن خارج و به توزیع "فرضیهٔ مقابل صحیح است" منتقل شود. معنای آن این است که در هیچ شرایطی هیچ‌گونه اثر یا همبستگی پیدا نخواهد شد. برای حل این مشکل، پژوهشگر خطر پنج یا یک درصد اشتباه را می‌پذیرد و تصمیم می‌گیرد چنانچه احتمال وقوع این مورد استثنایی در توزیع "فرض پوچ صحیح است" فقط پنج یا یک درصد (بر حسب مورد) بود، آن را به عنوان یک مورد نادر و کمیاب از این توزیع خارج و نتیجه‌گیری کند که فرضیهٔ پوچ غلط است (اثر یا رابطه، بر حسب مورد وجود دارد). معنای "پذیرفتن خطر اشتباه به میزان پنج یا یک درصد" این است که اگر همین تجربه صد مرتبه تکرار شود، در پنج بار یا یکبار، تصمیم پژوهشگر در تأیید اثر یا رابطه اشتباه خواهد شد. بدین صورت تصمیمی که پس از آزمایش یک فرضیهٔ آماری روی نتایج آن گرفته می‌شود، می‌تواند به یکی از دو صورت زیر با اشتباه (خطا) همراه شود:

الف) ممکن است H_0 (فرضیهٔ پوچ) صحیح باشد، ولی تصمیم‌گیری محقق روی نتایج آزمون به اشتباه این باشد که باید آن را رد کرد (یعنی، $P \leq \alpha$ به دست آید). این را خطای نوع اول^۱ می‌نامند و با علامت α نشان می‌دهند. توضیح اینکه P میزان احتمال وقوع عدد مورد محاسبه است که رایانه به دست می‌دهد و از همان مقیاس و جنس آلفا (α) است.

ب) ممکن است H_0 غلط باشد ولی، تصمیم‌گیری محقق روی نتایج آزمون به اشتباه این باشد که باید آن را قبول کرد (یعنی، $P > \alpha$ به دست آید). این را خطای نوع دوم^۲ می‌نامند و با علامت بتا (β) نشان می‌دهند.

توضیح مهم: تصمیم‌گیری و تفسیر نتایج باید فقط بر مبنای H_0 (فرضیهٔ پوچ) صورت پذیرد. زیرا هیچ‌گونه معیار ریاضی یا آماری برای ارزشیابی فرضیهٔ مقابل (H_1) وجود ندارد از این‌رو، این فرضیه مقابل اصولاً قابل آزمایش و تفسیر کردن نیست. البته، فرضیهٔ مقابل همیشه بر اساس نتیجه‌ای که از H_0 به دست می‌آید "استنباط" می‌شود. یکی از دلایل مهمی که به روش‌های آزمون فرضیه "استنباطی" می‌گویند نیز همین است. دلیل دیگر استنباطی بودن آزمون فرضیه نیز این است که اطلاعات جامعه از روی یافته‌های نمونهٔ آن استنباط می‌شود (امیرتاش، ۱۳۸۸).

نقطهٔ احتمال α ، در توزیع احتمالات مواردی که H_0 صحیح بوده است، نقطه‌ای از مرز اعداد بحرانی روی محور X ها را نشان می‌دهد که چنانچه عدد مشاهده شده (تجربی) با آن مساوی و یا بزرگ‌تر از آن شود (یا عدد پی آن مساوی یا کوچک‌تر از آلفای تعیین شدهٔ پژوهشگر شود)، فرض پوچ (H_0) رد خواهد شد (یعنی رابطه یا تفاوت وجود دارد).

نقطهٔ احتمال β ، در توزیع احتمالات مواردی که H_0 غلط بوده است، نقطه‌ای از مرز اعداد بحرانی را روی محور X ها نشان می‌دهد که چنانچه عدد مشاهده شده (تجربی) کوچک‌تر از آن به دست آید، فرضیهٔ پوچ (H_0) قبول خواهد شد (یعنی رابطه یا تفاوت وجود ندارد).

واضح است که در آزمایش یک فرضیهٔ آماری به دنبال تصمیمی هستیم که با کمترین خطا همراه باشد. قابل اثبات است که خطاهای نوع اول و دوم با یکدیگر رابطهٔ معکوس دارند (امیرتاش، ۱۳۸۸). اینک، با توجه به رابطهٔ معکوس بین دو نوع خطاهای تصمیم‌گیری یاد شده، این سؤال مطرح می‌شود که تأکید روی کدام یک از این دو نوع خطا مناسب‌تر است. به این صورت، چنانچه تأکید را روی α بگذاریم (یعنی آن را کوچک بگیریم)، مقدار β افزایش خواهد یافت. این موضوع نشان می‌دهد که احتمال بروز یکی از انواع خطاهای تصمیم‌گیری را فقط به قیمت افزایش احتمال خطای نوع دیگر می‌توان کاهش داد.

1. Type one error
2. Type two error

یک آزمون آماری خیلی خوب، باید کوچکترین مقادیر ممکن α و β را دارا باشد ولی، با توجه به مطالب گفته شده، دست کم تا زمانی که حجم نمونه ثابت بماند، به چنین آزمونی نمی‌توان دست یافت. چون نتیجه بروز خطای نوع اول مهم‌تر از خطای نوع دوم است، از این رو روش عملی مناسب به این ترتیب است که پژوهشگر میزان خطای نوع اول را از قبل در "سطح قابل اغماض" تعیین و سپس، سعی می‌کند برای چنین مقداری از خطای نوع اول آزمونی را انتخاب کند که در آن میزان خطای نوع دوم کوچک‌ترین مقدار ممکن را داشته باشد. تعیین "سطح قابل اغماض" برای احتمال خطای نوع اول، یک مسئله آماری نیست و مقدار آن با توجه به اهمیتی که رد کردن فرضیه پوچ دارد معین می‌شود. معمولاً سطح احتمالات کوچکی مانند $0/05$ و یا $0/01$ را در انجام آزمون‌های آماری برای این نوع خطا به کار می‌گیرند. با توجه به تعبیری که از "احتمال" به دست می‌آوریم، اینکه مقدار خطای نوع اول برابر با $0/05$ باشد، معنایش این است که در هر 100 آزمون مستقل، 5 بار فرضیه پوچ را به غلط رد می‌کنیم (خطای تصمیم‌گیری نوع اول). یعنی، تأثیر یا رابطه‌ای را به اشتباه تأیید می‌کنیم (یافته‌ای را به کار می‌گیریم که در اصل غلط است).

در آخر، این نکته قابل توجه است که یکی از راه‌های کاهش هم‌زمان α ، و به‌ویژه β در یک آزمون آماری، برای کاهش میزان خطای نوع اول در تصمیم‌گیری پژوهشگر، و برای کاهش میزان خطای نوع دوم افزایش حجم نمونه (N) است. واضح است که افزایش حجم نمونه با صرف هزینه بالا و وقت و انرژی زیادتر همراه است که مقرون به صرفه نیست ولی، راه مؤثری برای کاهش میزان خطای نوع دوم است. ضمناً، یکی از معیارهای رد کردن H_0 ، همان‌گونه که قبلاً اشاره شد، مقایسه P است که رایانه از محاسبات خود درباره احتمال وقوع عدد مشاهده به دست می‌دهد و در توزیع احتمالات α قرار می‌گیرد و از همان جنس و مقیاس است. یعنی اگر P مساوی و کوچک‌تر از α تعیین شده در همان توزیع احتمالات "فرضیه پوچ صحیح است" به دست آمد، H_0 رد خواهد شد.

قابل توجه است که P و α هر کدام، یک عدد بحرانی مشخص را روی محور X ها در توزیع احتمالات α (توزیعی که فرضیه پوچ در آن صحیح است) مشخص می‌کنند که α را محقق تعیین می‌کند ولی، P را مقدار اختلاف میانگین‌ها، یا رابطه همبستگی (عدد مشاهده)، مشخص می‌سازد. این برداشت ممکن است این سوء تفاهم را به وجود آورد که هر چه P کوچک‌تر به دست آید، شدت تأثیر متغیر مستقل (فاصله میانگین گروه تجربی از میانگین گروه کنترل)، یا میزان رابطه، بیشتر بوده است. به نظر کپل (۱۹۶۲) مقدار α یا P فقط میزان احتمال خطای نوع اول را مشخص می‌کنند و ارتباطی به میزان تأثیر یا شدت متغیر مستقل یا میزان همبستگی ندارند. میزان رابطه همبستگی، یا شدت تأثیر متغیر مستقل، با شاخص‌های آماری دیگری مانند t^2 ، ω^2 ، و ES ... قابل بررسی هستند (تامسس و نلسون، ۲۰۰۱). ضمناً، قابل توجه است که تنها در پژوهش‌های علت و معلولی و همبستگی فرضیه‌های آماری تدوین و به آزمایش گذارده می‌شوند.

رابطه بین α ، β ، N ، تفاوت μ_1 و μ_2 ، و یا همبستگی بین متغیرها

- α خطای نوع اول: H_0 صحیح است (اختلاف بین میانگین‌ها یا همبستگی بین متغیرها وجود ندارد)، ما آن را رد می‌کنیم. احتمال دیدن یک تفاوت یا رابطه غیر واقعی، یعنی احتمال اینکه تفاوت یا رابطه واقعی وجود ندارد ولی، ما آن را می‌بینیم و تأیید می‌کنیم.

- β خطای نوع دوم: H_0 غلط است (تفاوت میانگین‌ها یا رابطه بین توزیع‌ها حقیقی است)، ولی ما آن را قبول می‌کنیم. احتمال ندیدن یک تفاوت واقعی، یعنی احتمال اینکه یک تفاوت واقعی وجود دارد ولی ما آن را نمی‌بینیم.

- α را خود ما تعیین می‌کنیم و ملاک آن اهمیتی است که کاربرد نتایج دارد (حیاتی یا مهم بودن آن)، هر چه α کوچک‌تر شود، عدد بحرانی بزرگ‌تر می‌شود (مشاهده یک جدول اعداد بحرانی این را نشان خواهد داد)، از این رو رد کردن H_0 مشکل‌تر می‌شود، یعنی تفاوت بین میانگین‌ها باید خیلی بزرگ باشد تا به عنوان مثال، t مشاهده شده ارزش بزرگ‌تر از t بحرانی به دست آورد (مقدار عددی پی از آلفا کوچک‌تر شود).

- رد کردن H_0 علاوه بر α به N نیز بستگی دارد: هر چه N بزرگ‌تر باشد، t بحرانی کوچک‌تر می‌شود، پس رد کردن H_0 آسان‌تر خواهد شد. به عبارت دیگر، تفاوت کمتری بین میانگین‌ها لازم است تا H_0 رد شود. قابل توجه است که تصمیم درباره میزان احتمال α بدون توجه به احتمال اشتباه نوع دوم (β) گرفته می‌شود.

- عمل و اثر β بحث طولانی و مفصلی دارد که طرح همه جزئیات آن در اینجا خارج از موضوع است. در یک آلفای معین ($\alpha=0.05$)، مقدار عددی β را N و تفاوت واقعی بین μ_1 و μ_2 (میانگین‌ها) معین می‌کنند.

- در یک α و N معین، مقدار عددی β با افزایش تفاوت واقعی بین میانگین‌ها کاهش پیدا می‌کند ($1-\beta$ افزایش می‌یابد).

- توانایی آزمون ($1-\beta$) مقدار احتمال یک آزمون آماری را در یافتن اثر یا همبستگی، هنگامی که آن اثر یا همبستگی واقعاً وجود دارد، نشان می‌دهد. به این احتمال اصطلاحاً "توانایی آزمون آماری" گفته می‌شود. بعضی منابع به آن "توان آزمون" هم گفته‌اند. مک کال (۱۹۷۵) نظر می‌دهد که برای افزایش آن لازم است:

۱- تعداد اعضای نمونه افزایش داده شود.

۲- مقدار احتمال α افزایش داده شود.

- ۳- شدت متغیر مستقل زیاد شود.
- ۴- هر کجا که ممکن است، به جای آمارهای غیر پارامتریک از انواع مشابه پارامتریک استفاده شود.
- اگر حکم یک طرفه به دو طرفه (تقسیم مقدار α در دو دامنه راست و چپ توزیع) تبدیل شود، α عملاً نصف می شود (کوچک تر می شود)، پس (β بزرگ می شود) و بنابراین احتمال ندیدن تفاوت یا رابطه واقعی زیاد می شود ($1-\beta$ کاهش می یابد).
- هر چه تفاوت بین میانگین ها زیادتر شود، احتمال رد H_0 ، وقتی غلط است، افزایش پیدا می کند. (β کوچک تر می شود).
- در یک تفاوت معین بین میانگین ها و N معین، ارزش β با کاهش α افزایش خواهد یافت. بنابراین، در صورتی که α خیلی سخت گیرانه باشد (کوچک باشد)، احتمال ندیدن یک تفاوت یا رابطه واقعی (β)، حتی اگر مهم هم باشد، رد کردن H_1 مشکل تر می شود چون عدد بحرانی افزایش می یابد و ضمناً β هم افزایش پیدا می کند.
- با افزایش N ، که موجب کاهش میزان اشتباهات تجربی (انحراف استاندارد)، - که در همه محاسبات آماری وجود دارد - می شود، توانایی آزمون ($1-\beta$) در رد کردن فرضیه پوچ غیر صحیح (احتمال پیدا کردن تأثیر یا رابطه هنگامی که واقعاً وجود دارد)، افزایش می یابد.
- در هر تفاوت بین میانگین ها، یا همبستگی بین متغیرها در حد α معین، میزان اشتباه نوع دوم (β) را فقط N تعیین می کند یعنی، هر چه N کوچک تر شود، ارزش β بزرگ تر می شود (احتمال ندیدن تفاوت یا رابطه واقعی افزایش می یابد).

بحث و نتیجه گیری

- ۱- قابل توجه است که احتمال دیدن یک تفاوت واقعی (رد کردن H_0) مفهومی این نیست که H_0 به طور مطلق غلط است. با وجود این، اکثر محققان حتی برای نمونه های واقعاً کوچک، هنگامی که به یک حد معنی دار بودن مورد نظر نمی رسند، تمایل دارند نتیجه گیری کنند که تفاوت، یا رابطه موجود نیست، یا قابل توجه نیست. با عنایت به بحثی که درباره β گذشت، مشاهده می شود این نتیجه گیری ضمانت کافی را ندارد. البته، می توان یک ضابطه دو معیاری را برای قبول H_0 در نظر گرفت: چنانچه α (احتمال اتفاق تفاوت یا رابطه) بیشتر از مثلاً 0.05 و β (احتمال عدم اتفاق تفاوت) کمتر از 0.05 بود، H_0 قبول شود.
- ۲- وجود α در قوانین علمی موجب می شود تا جامعه از موارد یا اطلاعاتی استفاده کند که اصولاً مناسب یا صحیح نیستند، یعنی تأثیر روش، دارو یا تمرینات ورزشی ای توصیه شده اند که عملاً بی اثر یا غیر قابل استفاده هستند.

۳- وجود β در قوانین علمی موجب می‌شود تا جامعه از موارد یا اطلاعاتی که عملاً برای مفید یا قابل استفاده‌اند محروم بماند، یعنی تأثیر روش، دارو یا تمرینات ورزشی‌ای که می‌توانسته‌اند برای مفید یا قابل استفاده باشند، رد شده و توصیه نشده‌اند.

۴- با این ترتیب، وجود α و β در یافته‌ها و قوانین علمی جامعه را از هر دو سو متضرر می‌کند، وقت، پول و انرژی‌ای که صرف یافتن آن‌ها شده است عملاً از بین رفته و از سوی دیگر، دامنه رشد و توسعه علم در آن زمینه‌ها قطع شده است.

شاید ریشه بخشی از این اشتباهات در اطلاعات ناکافی بعضی از پژوهشگران از نظام، ماهیت، و روش‌های آزمایش فرضیه‌های آماری نهفته است. به سبب اینکه درک بهتر هدف و پایه‌های علمی آزمایش فرضیه به نسبت پیچیده است و اطلاعات داده شده در بعضی منابع موجود در این باره نیز تا حدودی محدودند یا روشنی کافی را ندارند، و به دلیل اهمیت بسیار زیادی که حذف اشتباهات تفسیری و نتیجه‌گیری از آزمایش فرضیه‌های آماری در توسعه و گسترش علم برای خودکفایی و کاهش وابستگی‌های علمی در یک جامعه دارند، توصیه می‌شود که استادان فن نقش بیشتری را در شکافتن و ارائه آن به زبان ساده‌تر بر عهده گیرند.

منابع

۱. امیرتاش، علی محمد (۱۳۸۸). *جزوه درسی مقدمات آمار استنباطی*، واحد علوم و تحقیقات دانشگاه آزاد اسلامی تهران.
۲. امیرتاش، علی محمد (۱۳۹۱). *انواع روش‌های پژوهش برای علوم انسانی با تأکید بر تربیت بدنی و علوم ورزشی*، (در دست چاپ) واحد علوم و تحقیقات دانشگاه آزاد اسلامی تهران.
۳. توماس، جری و نلسون، جک (۲۰۰۱). *روش تحقیق در فعالیت بدنی*، جلد اول، ترجمه علی محمد امیرتاش و فرشاد تجاری (۱۳۸۸)، تهران، انتشارات واحد علوم و تحقیقات دانشگاه آزاد اسلامی تهران.
۴. توماس، جری و نلسون، جک (۲۰۰۱). *روش تحقیق در فعالیت بدنی*، جلد دوم، ترجمه علی محمد امیرتاش و فرشاد تجاری (۱۳۹۰)، تهران، انتشارات واحد علوم و تحقیقات دانشگاه آزاد اسلامی تهران.
5. Anderson, B.F (1971). *The Psychology Experiment, an Introduction to the Scientific Method (2ndEd)*, Brooks/Cole Publishing Co. Belmont California.
6. Campbell, D.T., & Stanley, J.C (1963). *Experimental and Quasi-Experimental Designs for Research*, Rand McNally College Publishing Co Chicago.
7. Gay, L.R (1981). *Educational Research, Competencies for Analysis and Application, (2nd.ed.)*, Charles E.Merrill Publishing Co, A Bell and Howell Company, Columbus.
8. Jerry, R.T., & Jack, K.N (2001). *Research Methods in Physical Activity*.
9. John, W.B., & James, V.K (1997). *Research in Education (8th Edition)*.

10. Keppel, G (1973). *Design And Analysis, A Researcher's Handbook*, Prentice-Hall. INC. New Jersey.
11. Kerlinger, F.N (1973). *Foundations of Behavioral Research*. Hol T, Rinehart and Winston Inc.
12. McCall, R.B (1975). *Fudamental Statistics for Psychology (2nd.)*, Harcourt Brace Jovanovich, San Fransisco.
13. Minium, E.W (1970). *Statistical Reasoning in Psychology & Education*, John Wiley & Sons, Inc.
14. Stephen, I., & William, B.M (1997). *Handbook in research and Evaluation (third Ed.)*, EDITS Publishers, San Diego USA.
15. Thomas, J.R., Nelson, J.K., & Silverman, S.J (2005). *Research Methods in Physical Activity (5th Ed)*, HumanKinetics, Champaign IL. USA.
16. Udinsky, B.F., Steven, J.O., & Samuel, W.L (1981). *Evaluation Resource Handbook, Gathering, Analyzing. Reporting Data* Edits Publishers San Diego USA.

Theoretical and philosophical foundations of hypothesis testing presented in a simple language

Amirtash A.M.*

Professor in Science and Research Branch- Islamic Azad University

Abstract

Objective: The purpose of the study was to investigate the theoretical foundations of hypothesis testing and to look into its rational and philosophical concepts and practical implications of this scientific process in generating new knowledge.

Methodology: To this end, the existing literature and leading references on the scientific bases, procedures, and statistical methods in relation to hypothesis testing in order to determine cause and effect relationship or the correlation between the variables were purposefully studied. The rational, philosophical, and practical considerations and inferences which are presented in the present article are the result of the author's nearly forty years of experience and teaching in these domains.

Results: The results of the studies indicated that almost all the statistical inferences encompass three kinds of errors: Type I, Type II, and other types of errors that researchers may make during the process of their studies (Standard Deviation). The inferential statistics results are valuable when they have least amount of mistakes and errors. To achieve this goal, the Type I error can be controlled by minimizing the probability of its occurrence by the researcher. Type II error is decreased by undertaking provisions like: adding the intensity of independent variable, increasing the number of the participants, choosing a higher degree of probability for Type I error, and using parametric statistical methods instead of nonparametric ones. Ultimately, there is a considerable number of statistical and research method techniques for lowering the indices of variability and deviations of the scores. Interestingly enough, there is a negative relationship between Type I and Type II error. In other words, lowering the probability of one of them will have an incremental effect on the other. From a practical standpoint, Type I error recommends inefficient and unproductive knowledge to the target population; whereas, Type II error may deprive people from useful and effective information.

Conclusion: Researchers' mistakes may absolutely create double-sided damages to society. The researcher's choice between the first two types of errors relies on the significance of the application of the results.

Key Words: Statistical inference, Power of the test, Types of error, Standard deviation, Standard error.
