

## اعتبار درونی و اعتبار نتیجه‌گیری آماری پژوهشهای آزمایشی و شبه آزمایشی

دکتر حسین شکرکن \*

موضوع نوشتار حاضر بحث و بررسی در خصوص اعتبار درونی و اعتبار نتیجه‌گیری آماری طرحهای پژوهشی آزمایشی و شبه آزمایشی است. آزمایش بخشی از پژوهش است که در آن متغیرهای مستقل در ارتباط با فرضیه‌های تحقیق دستکاری می‌شوند و تاثیراتشان بر متغیرهای وابسته مورد مشاهده قرار می‌گیرد. اعتبار به معنای حصول اطمینان از صحت یک رابطه علی است. مسلم دانستن یک رابطه علی و مشخص ساختن جهت علیت مستلزم سه شرط است: تقدم زمانی علت بر معلول، تغییر متقارن علت و معلول و فقدان هرگونه تبیین موجه دیگر. مراد از اعتبار درونی آن است که تغییر حاصل در متغیر وابسته باید منحصرآ "ناشی از تغییر در متغیر مستقل باشد. چهارده متغیر ناخواسته اعتبار درونی طرحهای پژوهشی آزمایشی و شبه آزمایشی را مورد تهدید قرار می‌دهد. این متغیرها در دودسته، تحت عنوان عوامل تهدیدکننده مهارپذیر، و عوامل تهدیدکننده مهارناپذیر، مورد بحث قرار گرفته‌اند. اعتبار نتیجه‌گیری آماری عبارت از این است که اثر یا رابطه کشف شده در یک پژوهش آزمایشی باید واقعی و ناشی از مداخله آزمایشی باشد نه حاصل عوامل تصادفی یا خطا. هفت متغیر ناخواسته اعتبار نتیجه‌گیری آماری را به خطر می‌اندازد. این متغیرها نیز به ایجاز تشریح شده‌اند و چگونگی مهار آنها مورد بحث قرار گرفته است.

آزمایش به بخشی از پژوهش اطلاق می‌شود که در آن در یک یا چند متغیر مستقل دستکاری به عمل می‌آید و تأثیرات آنها بر یک یا چند متغیر وابسته مورد بررسی قرار می‌گیرد (کمپیل و استانلی<sup>۱</sup>، ۱۹۶۳). لازم به تذکر است که روی سخن این نوشتار با پژوهشهایی است که به روش آزمایشی و شبه‌آزمایشی انجام می‌گیرند و پژوهشهایی را که به دیگر روشها، مثلا "روش همبستگی ساده، انجام می‌گیرند مورد توجه قرار نمی‌دهد. پژوهش آزمایشی حقیقی پژوهشی است که در آن مهار و کنترل بسیاری از سوی آزماینده بر متغیرهای مختلف مربوط به آزمایش اعمال می‌شود. در این پژوهش گزینش آزمودنیها از جامعه‌های آماری و گماردن آنان به گروههای آزمایشی و گواه به صورت کاملا "تصادفی انجام می‌گیرد.

پژوهش شبه‌آزمایشی به پژوهشی اطلاق می‌شود که در آن یا انتخاب آزمودنیها، به صورت کاملا "تصادفی نیست و در نتیجه گروههای آزمایشی و گواه معادل و هم‌ارز نیستند، و یا میزان مهار متغیرهای نامربوط در سطحی دلخواه و مطلوب نیست. اعتبار در پژوهشهای آزمایشی و شبه‌آزمایشی به بهترین تقریب موجود در جهت درستی قضیه‌ها، از جمله قضیه‌های مربوط به علیت، اطلاق می‌شود. (کوک و کمپیل<sup>۲</sup>، ۱۹۷۹) آزمایشها وسایل بررسی فرضیه‌های علی هستند. کوک و کمپیل (۱۹۶۶-۱۹۷۹) علی بودن رابطه، میان دو متغیر و مشخص ساختن جهت علیت از متغیر مستقل یا علت به متغیر وابسته یا معلول را مستلزم سه شرط زیر می‌دانند:

- ۱- تقدم زمانی علت بر معلول. شرط نخست مربوط به تقدم زمانی علت بر معلول است. در آزمایشها این شرط به صورت تقدم زمانی مداخله‌آزمایشی<sup>۳</sup> یا دستکاری متغیر مستقل بر تغییر متغیر وابسته که در پس‌آزمون<sup>۴</sup> منعکس است متجلی می‌شود.
- ۲- تغییر متقارن علت و معلول. شرط لازم دوم برای استنباط قابل اعتماد رابطه، علی از متغیر مستقل به متغیر وابسته این است که مداخله یا مداخله‌های

۱-Cook and Stanley

۲-Cook and Campbell

۳-experimental treatment

۴-Posttest

آزمایشی و معلول یا متغیر وابسته باهم تغییر متقارن داشته باشند. اگر علت و معلول بالقوه به هم مربوط نباشند، نمی‌توان گفت که یکی علت دیگری است.

۳- فقدان هر گونه تبیین موجه دیگر. شرط لازم سوم برای استنباط رابطه علی این است که هیچ گونه علت یا تبیین موجه دیگر غیر از متغیر مستقل نباید برای متغیر وابسته وجود داشته باشد. به بیان دیگر، باید اطمینان حاصل شود که فقط متغیر مستقل الف علت متغیر وابسته ب است و لا غیر.

اعتبار آزمایش چهار وجه مختلف دارد که به نامهای ویژه اعتبار درونی<sup>۱</sup>، اعتبار نتیجه گیری آماری<sup>۲</sup>، اعتبار بیرونی<sup>۳</sup>، و اعتبار سازه<sup>۴</sup> موسوم شده اند. میان اعتبار درونی و اعتبار نتیجه گیری آماری از یک سو و اعتبار بیرونی و اعتبار سازه از سوی دیگر گونه ای شباهت وجود دارد که در جای خود در خصوص آنها توضیح لازم داده خواهد شد. هر یک از این اعتبارهای چهارگانه را عوامل خاصی به خطر می‌اندازند. این عوامل به نامهای مختلفی از جمله عوامل مزاحم، متغیرهای ناخواسته، متغیرهای نامربوط، و متغیرهای تهدید کننده اعتبار پژوهش آزمایشی خوانده شده‌اند. اگر این عوامل مزاحم مهار نشوند می‌توانند به صورت تبیینهای رقیبی برای مداخله آزمایشی درآیند، بدین معنا که می‌توان تغییرات حاصل در متغیر وابسته را به آنها منسوب کرد تا به مداخله آزمایشی. اطمینان کامل از صحت آزمایشهای علمی مستلزم این است که اثرهای متغیرهای ناخواسته مربوط به هر یک از انواع چهارگانه اعتبار پژوهش آزمایشی زائل یا مهار شوند. نوشتار حاضر برآن است تا اعتبار درونی و اعتبار نتیجه گیری آماری پژوهشهای آزمایشی و شبه آزمایشی را به لحاظ شباهتی که با هم دارند در یک مقاله مورد بررسی قرار دهد. عوامل تهدید کننده هر یک از این دو را به اختصار تشریح کند. و شیوه‌های مهار آنها را مطرح سازد. بحث و بررسی در خصوص اعتبار بیرونی و اعتبار سازه پژوهشهای آزمایشی و مسائل مربوط به آنها به فرصت دیگری موکول می‌شود.

۱- internal validity

۲- statistical conclusion validity

۳- external validity

۴- construct validity

## اعتبار درونی

مراد از اعتبار درونی آن است که تغییر حاصل در متغیر وابسته باید منحصرآ ناشی از تغییر در متغیر مستقل باشد. اگر، در یک آزمایش، بتوان مسجل ساخت که تنها عاملی که بر متغیر وابسته تأثیر بخشیده و سبب ایجاد تغییراتی در آن شده است منحصرآ "مداخله آزمایشی یا دستکاری متغیر مستقل است گوئیم که آن آزمایش از اعتبار درونی برخوردار است، اعتبار درونی شرط اصلی هر آزمایش است و بدون آن هیچ آزمایشی معتبر نیست (کمپیل و استانلی<sup>۱</sup>، ۱۹۶۶). دسته‌ای از عوامل مزاحم یا متغیرهای ناخواسته اعتبار درونی پژوهش‌های آزمایشی را به مخاطره می‌افکنند. برای حصول اطمینان از اعتبار درونی یک پژوهش آزمایشی باید بتوان نشان داد که هیچ کدام از این عوامل مزاحم علت پیدایش تغییر در متغیر وابسته نیست. عوامل تهدیدکننده اعتبار درونی را می‌توان به دو دسته تقسیم کرد: ۱- دسته اول عوامل مزاحمی که با انتخاب تصادفی آزمودنیها می‌توان تأثیرات مخل آنها را مهار کرد. به اینها عوامل تهدید کننده مهارپذیر اطلاق می‌کنیم. ۲- دسته دوم عوامل مزاحمی هستند که انتخاب تصادفی آزمودنیها نمی‌تواند تأثیرات مخل آنها را مهار کند، بلکه برای این‌کار به تمهیدات دیگری نیاز هست. اینها را عوامل تهدید کننده مهارناپذیر نامیده ایم.

## عوامل تهدید کننده مهار پذیر اعتبار درونی

هشت عامل تهدید کننده اعتبار درونی وجود دارد که تأثیرات مخل آنها را در طرحهای پژوهشی آزمایشی می‌توان از طریق انتخاب تصادفی آزمودنیهای گروههای آزمایشی و گواه مهار کرد. این عوامل تهدید کننده به شرح زیرند:

## ۱- تاریخ

در فاصله زمانی میان پیش‌آزمون<sup>۲</sup> و پس‌آزمون رویدادهای تغییرزای بسیاری به غیر از مداخله آزمایشی ممکن است رخ بدهند که می‌توانند بر متغیر وابسته تأثیر

۱ - Campbell and Stanley

۲ - Pretest

بگذارند ، و سبب ایجاد تغییراتی در آن بشوند . همه این رویدادها را در مجموع اصطلاحاً "عامل" تاریخ<sup>۱</sup> می‌نامند . برای این‌که متغیر تاریخ به‌صورت رقیب موجهی برای مداخله آزمایشی درآید ، باید اکثریت آزمودنیهای مورد مطالعه را تحت تأثیر قرار دهد ، مثلاً "از طریق یک کلاس درس یا از رهگذر یک برنامه خبری که پخش‌گسترده‌ای داشته است . هر چه مدت زمان میان پیش‌آزمون و پس‌آزمون بیشتر باشد امکان اعمال نفوذ متغیر تاریخ بیشتر می‌شود . در آزمایشی که یکی دو ساعت به‌طول انجامد ، تاریخ ممکن است مسأله‌ای جزئی باشد ، گرچه حتی در این‌جا نیز باید مراقب رویدادهای بیرونی از قبیل خنده آزمودنیها ، عوامل مخل ، و جز اینها بود . پاره‌ای از اثرهای ممکن مربوط به فصول سال یا رویدادهایی که طبق برنامه در یک مؤسسه به وقوع می‌پیوندند ، تحت عنوان متغیر تاریخ طبقه بندی می‌شوند . مثلاً ، "خوشبینی ممکن است با فصول سال و اضطراب با برنامه امتحانات نیمسال تغییر کنند (کروک<sup>۲</sup> ۱۹۳۷ ، ویندل<sup>۳</sup> ۱۹۵۴) . چنین اثرهایی ممکن است در فاصله زمانی میان پیش‌آزمون و پس‌آزمون در متغیر وابسته تغییری ایجاد کنند که با اثر مداخله آزمایشی اشتباه گرفته شود . بدین ترتیب ، در پاره‌ای از موارد تغییر حاصل در متغیر وابسته می‌تواند ناشی از عامل تاریخ باشد ، مگر این‌که اثر این عامل مزاحم را بتوان مهار یا زائل کرد .

## ۲- رشد و نمو

در فاصله زمانی میان پیش‌آزمون و پس‌آزمون تغییراتی ممکن است در خود آزمودنیها روی دهد که از مداخله آزمایشی مستقل باشند . این تغییرات را اصطلاحاً "متغیر ناخواسته" رشد و نمو<sup>۴</sup> نامیده‌اند . اصطلاح رشد و نمو در اینجا دلالت بر همه فرایندهای زیستی و روانشناختی دارد که با گذشت زمان به صورتی منظم در آزمودنیها روی می‌دهند و از رویدادهای بیرونی ویژه مستقلند . این متغیر ناخواسته می‌تواند بر متغیر وابسته تأثیر بگذارد و به صورت رقیبی برای متغیر مستقل درآید ، مگر این‌که بتوان اثر آن را مهار یا زائل کرد . بدین ترتیب ، در فاصله زمانی میان پیش‌آزمون و پس‌آزمون آزمودنیهای شرکت کننده در یک آزمایش ممکن است مسنتر ، گرسنه تر ، خسته تر ، کسلتر وغیره بشوند و تفاوت مشهود در نمره‌های پس‌آزمون آنان

۱ - history

۲ - Crook

۳ - Windle

۴ - maturation

می‌تواند منعکس کننده، این عوامل باشد نه مداخله آزمایشی.

### ۳ - متغیر آزمودن

در جریان اجرای یک پژوهش آزمایشی پاسخ به پیش آزمون توسط آزمودنیها می‌تواند سبب آشنائی آنان با شیوه و زمینه، مربوط به آن آزمون گردد و در نتیجه نمره های پس آزمون آنان را افزایش دهد. این متغیر ناخواسته، یعنی تأثیر پاسخ به یک آزمون بر نمره های حاصل از پاسخ به آزمون دیگر، "متغیر آزمودن"<sup>۱</sup> خوانده می‌شود. چنین تغییری صرفاً "مربوط به متغیر آزمودن است و ارتباطی به مداخله آزمایشی ندارد. مثلاً"، در آزمونهای پیشرفت تحصیلی و هوش، دانش آموزانی که برای بار دوم به آزمونی پاسخ می‌گویند، یا صورت همتائی از همان آزمون را می‌گیرند، معمولاً "نمره های بالاتری به دست می‌آورند تا کسانی که برای نخستین بار آن آزمون را می‌گیرند (انستازی<sup>۲</sup>، ۱۹۵۸، ص ۱۹۱ - ۱۹۰، کین و هایم<sup>۳</sup>، ۱۹۵۰). در رابطه با آزمونهای شخصیت نیز اثر مشابهی مشاهده شده است (ویندل<sup>۴</sup>، ۱۹۵۴).

آگاهی از وجود متغیر آزمودن سبب شده است که آزمونها از جهت ایجاد واکنش در آزمودنیها با هم متفاوت باشند. این بدان معناست که برخی از آزمونها در آزمودنیها واکنش ایجاد می‌کنند و بعضی نمی‌کنند. این موضوع که خود فرایند سنجش ممکن است آنچه را سنجیده می‌شود تغییر دهد به اثرواکنشی معروف است. مدتهای مدید است که این اثر در علوم اجتماعی شناخته شده است. هر وقت که فرایند آزمایش صرفاً "منحصر و محدود به ثبت و ضبط رفتار نباشد، بلکه محرکی برای تغییر رفتار نیز باشد، می‌توان چنین اثر واکنشی را انتظار داشت. مثلاً"، گماردن مشاهده‌گران در یک کلاس درس به منظور مشاهده، مهارتهای روابط انسانی یک معلم در محل تعامل یک گروه متعامل می‌تواند نحوه، تعامل آن رادگرگون سازد، و جزاینها. به طور کلی، چنین انتظار می‌رود که هر قدر ابزار آزمون بدیعتر و توان انگیزشی آن بیشتر باشد، واکنش بیشتری را موجب شود.

۱ - Testing

۲ - Anastasi

۳ - Cane and Heim

۴ - Windle

متغیر ناخواسته "تغییر در ابزار سنجش" به تغییراتی اطلاق می‌شود که در فاصله زمانی میان پیش‌آزمون و پس‌آزمون در ابزار سنجش روی می‌دهد. تغییر در درجه بندی یک مقیاس یا در مشاهده کران و یا در نمره‌های به کار گرفته شده در جریان آزمایش، نمونه‌ای از این متغیر است. در نوشته‌های روش شناختی مربوط از این متغیر اصطلاحاً "تحت عنوان" به کارگیری یا فرسودگی ابزار سنجش<sup>۱</sup> یاد شده است (کمپبل ۱۹۵۷). این تغییرات، مستقل از مداخله آزمایشی، می‌توانند در نتایج سنجشهای بعدی آزمودنیها دگرگونی ایجاد کنند. مثلاً، "فرسودگی قطعات ابزار یادگیری، یا یکنواختی و خستگی برای مشاهده‌گران انسانی ممکن است سبب تغییراتی در نمره‌های پس‌آزمون آزمودنیها شود. پس این متغیر به تغییراتی در ابزار سنجش اشاره دارد که ممکن است تفاوت میان نمره‌های پیش‌آزمون و پس‌آزمون را تبیین سازند. اگر به امتحانات تشریحی نمره داده می‌شود، معیارهای نمره‌گذاری ممکن است در فاصله میان پیش‌آزمون و پس‌آزمون تغییر کنند. اگر شرکت در کلاس درس مورد مشاهده است، مشاهده‌گران در نوبت دوم ممکن است ماهرتر یا کم توجه تر شوند. اگر از والدین دانش‌آموزان مصاحبه به عمل می‌آید، آشنائی مصاحبه‌گر با برنامه مصاحبه و یا با آن والدین خاص ممکن است در نوبت دوم تغییراتی را موجب شود. بنابراین، تفاوت‌های میان نمره‌های پیش‌آزمون و پس‌آزمون آزمودنیها ممکن است ناشی از این متغیر باشد و نه مداخله آزمایشی.

#### ۵ - گرایش به سوی میانگین

فهم پدیده گرایش به سوی میانگین که به "بازگشت آماری"<sup>۲</sup> نیز معروف است. نیاز به دقت نظر دارد. هرگاه میان دو متغیر همبستگی نا کامل وجود داشته باشد و آزمودنیهای مورد نظر تنها از میان کسانی انتخاب شوند که در متغیر اول نمره‌های

\* مراد از ابزار سنجش هر وسیله‌ای است که توسط آن آزمودنیها در رابطه با متغیرهای مورد نظر سنجیده می‌شوند و می‌تواند مشتمل بر اشیاء و اشخاص باشند.

- ۱ - Instrumentation or instrument decay      ۲ - Campbell  
۳ - Statistical regression

انتهاائی، یعنی نمره های خیلی بالا یا خیلی پایین داشته باشند، نمره های آنان در متغیر دوم به سوی میانگین جامعه آماری مربوط بازگشت می کند. این بدان معناست که کسانی که در متغیر اول نمره های خیلی بالا دارند، نمره های آنان در متغیر دوم افت می کند. برعکس آنانی که در متغیر اول نمره های خیلی پایین دارند، نمره هایشان در متغیر دوم افزایش می یابد. پدیده "بازگشت آماری" در طرح های پژوهشی آزمایشی می تواند به شرح زیر به صورت عاملی مزاحم برای اعتبار درونی درآید. فرض کنید در یک آزمایش آن آزمودنیها را برای بررسی بیشتر انتخاب کنیم که در پیش آزمون نمره های خیلی پایینی کسب کنند و دیگر آزمودنیهای راکه در پیش آزمون شرکت جسته بودند کنار بگذاریم. آن گاه آزمودنیهای برگزیده را در معرض مداخله ای آزمایشی قرار می دهیم که برای افزایش نمره های آنان در متغیر وابسته تهیه شده است. سپس از این آزمودنیها پس آزمون به عمل می آوریم و معلوم می شود که نمره های آنان در پس آزمون در سطح معنی دار آماری بالا رفته است. آیا می توان این افزایش در نمره های پس آزمون آزمودنیها را منحصرآ "ناشی از مداخله" آزمایشی دانست؟ عامل مزاحم بازگشت آماری به این پرسش پاسخ منفی می دهد و مدعی است که اگر میان پیش آزمون و پس آزمون همبستگی ناکامل وجود داشته باشد، افزایش در نمره های پس آزمون آزمودنیها در آزمایش فرضی بالا می تواند صرفآ از پدیده "بازگشت آماری" حاصل شود و هیچ گونه ارتباطی به مداخله آزمایشی نداشته باشد.

مطلب فوق در قالب مثالی عینی بازگشت آماری را روشنتر می سازد. فرض کنید در آموزشگاهی بخواهند برای بهبود وضع آموزشی دانش آموزان ضعیف در پایه تحصیلی خاصی برنامه آموزشی ویژه ای اجرا کنند. به منظور تعیین دانش آموزان ضعیف یک آزمون مناسب پیشرفت تحصیلی تهیه می شود و در مورد همه دانش آموزان آن پایه تحصیلی اعم از ضعیف، متوسط، و قوی اجرا می گردد. آن گاه دانش آموزانی که نمره آنان در آن آزمون از حد معینی پایینتر است به عنوان دانش آموزان ضعیف برگزیده می شوند و برنامه آموزشی ویژه در مورد آنان اجرا می شود. پس از گذراندن آن دوره آموزشی بار دیگر این دانش آموزان با همان آزمون پیشرفت تحصیلی یا آزمونی همتای آن آزمایش می شوند و ملاحظه می گردد که میانگین آنان در این آزمون نسبت به میانگین آنها در آزمون قبلی افزایش یافته است. چنین نتیجه می گیرند که بهبود حاصل ناشی از تأثیر مثبت آن برنامه آموزشی است. توجه داشته باشید که در آزمون پیشرفت تحصیلی که پس از برگزاری برنامه آموزشی ویژه اجرا شد تنها



دانش‌آموزانی شرکت کردند که نمره<sup>۴</sup> آنان در آزمون پیشرفت تحصیلی اولیه از حد معینی پایینتر بود، در حالی که در آزمون پیشرفت تحصیلی اولیه همه<sup>۵</sup> دانش‌آموزان آن پایه تحصیلی شرکت جسته بودند.

جنبه‌های مهم مثال فوق را در قالب اصطلاحات فنی مربوط به یک آزمایش می‌توان چنین بیان داشت. آزمون پیشرفت تحصیلی که برای تعیین دانش‌آموزان ضعیف در مورد همه دانش‌آموزان آن پایه تحصیلی خاص به کار رفت در حکم پیش‌آزمون است. دانش‌آموزان ضعیف برگزیده شده آزمودنیهای این آزمایش هستند. برنامه<sup>۶</sup> آموزش ویژه به منزله<sup>۷</sup> مداخله<sup>۸</sup> آزمایشی است. سرانجام، آزمون پیشرفت تحصیلی که پس از پایان آن دوره<sup>۹</sup> آموزشی، و صرفاً<sup>۱۰</sup> در مورد دانش‌آموزانی که در آن دوره شرکت کرده بودند، اجرا شد، پس‌آزمون محسوب می‌شود.

آیا نتیجه‌گیری فوق که افزایش در میانگین پس‌آزمون آزمودنیها را ناشی از مداخله<sup>۱۱</sup> آزمایشی می‌داند صحت دارد؟ پدیده<sup>۱۲</sup> بازگشت آماری بر صحت این نتیجه‌گیری سایه تردید می‌افکند و مدعی است که افزایش در میانگین پس‌آزمون آزمودنیها می‌تواند ناشی از متغیر<sup>۱۳</sup> "بازگشت آماری" باشد نه مداخله<sup>۱۴</sup> آزمایشی، متغیر<sup>۱۵</sup> آزمودن و جز اینها. بدین معنا که حتی اگر دانش‌آموزان مزبور اصولاً<sup>۱۶</sup> از برنامه<sup>۱۷</sup> آموزشی ویژه استفاده نمی‌کردند، باز هم میانگین پس‌آزمون آنان از میانگین پیش‌آزمونشان بیشتر می‌شد. این امر صرفاً<sup>۱۸</sup> ناشی از همبستگی ناکامل میان پیش‌آزمون و پس‌آزمون است. هر چه همبستگی میان دو متغیر (در این جا، پیش‌آزمون و پس‌آزمون) کمتر باشد، و هر چه میانگین گروه برگزیده از میانگین جامعه آماری مربوط دورتر باشد، بازگشت به سوی میانگین بیشتر است. بدین سان، تفاوت‌های مشهود میان نمره‌های پیش‌آزمون و پس‌آزمون می‌تواند ناشی از متغیر<sup>۱۹</sup> بازگشت آماری باشد و نه مداخله<sup>۲۰</sup> آزمایشی.

پدیده<sup>۲۱</sup> بازگشت آماری را می‌توان چنین توجیه کرد. هر چه نمره‌ای از میانگین انحراف بیشتری داشته باشد، احتمالاً<sup>۲۲</sup> خطای اندازه‌گیری بیشتری را دربردارد. به یک معنا کسی که نمره<sup>۲۳</sup> خیلی بالایی گرفته است به گونه‌ای غیر معمول از بخت و اقبال نیکوئی بهره مند شده است (خطای مثبت زیاد). از سوی دیگر، کسی که نمره<sup>۲۴</sup> خیلی پایینی کسب کرده است دچار بدشانسی زیادی شده است (خطای منفی زیاد). از آنجا که بخت و اقبال و یا بدشانسی ثبات ندارند و همواره به افراد یکسانی رو نمی‌آورند، چنین انتظار می‌رود که کسانی که در پیش‌آزمون نمره‌های بالا گرفته بودند در پس‌آزمون افت کنند و آنانی که نمره<sup>۲۵</sup> پایینی گرفته بودند وضع نسبی خود را بهبود بخشند.

بازگشت آماری پدیده ای گسترده است و منحصر به پیش آزمودن و پس آزمودن با یک آزمون واحد یا با صورتهای همتای یک آزمون نیست. در یک مدرسه اغلب مشاهده می شود که در یک آزمون پیشرفت تحصیلی باهوشترین دانش آموزان بالاترین نمره ها را کسب نمی کنند و کودنترین آنان نیز پائینترین نمره ها را نمی گیرند. اگر مدیر این مدرسه اعلام کند که مدرسه اش باهوشترین دانش آموزان را به اندازه کافی بر نمی انگیزد و کودنترین دانش آموزان را سخت به کار می کشد مرتکب سفسطه بازگشت آماری شده است. چه اگر دانش آموزانی را انتخاب کنیم که بالاترین و پائینترین نمره ها را در آزمون پیشرفت تحصیلی داشته باشند و آن گاه بد نمره های هوشبهر آنان نظر افکنیم مجبور خواهیم شد که به همان صورت غیراصولی نتیجه گیری متضادی به عمل آوریم.

از سوی دیگر، هرگاه نمونه ای معرف از جامعه آماری معینی، مثلا "دانش آموزان ضعیف یا قوی، انتخاب شود و در یک پیش آزمون میانگین خیلی بالا یا خیلی پائین به دست آورد و همان گروه در پس آزمون هم شرکت کند انتظار نمی رود میانگین آن گروه در پس آزمون بازگشت کند. علت این امر این است که منابع تصادفی پراش آزاد بوده اند که نمره های پیش آزمون آنان را در هر دو جهت تحت تأثیر قرار دهند. بنابراین، در مثال فوق اگر دانش آموزان ضعیف مثلا "، بر مینای میانگین تجمع پائین خود در دوران تحصیلی انتخاب می شدند، میانگین آنان در پس آزمون دچار افت بازگشت آماری نمی شد. در این حالت برخلاف قبل همه دانش آموزان منتخب برای شرکت در برنامه آموزشی ویژه، در پیش آزمون دچار بدشانسی نمی گردند، بلکه برخی از آنان از بخت و اقبال نیکویی برخوردار می شوند و نمره ای بالاتر از حد استحقاق خود می گیرند، در حالی که عده ای دچار بدشانسی می گردند و نمره ای کمتر از آنچه که باید بگیرند می گیرند. بدین ترتیب، عوامل تصادفی آزادند که نمره های پیش آزمون آنان را هم در جهت بالا و هم در جهت پائین تحت تأثیر قرار دهند. همچنین در این حالت همه دانش آموزانی که پیش آزمون را گرفته بودند به پس آزمون نیز پاسخ می گویند. در چنین موقعیتی عوامل تصادفی پراش در پس آزمون نیز تأثیری مشابه با تأثیر خود در پیش آزمون دارند، یعنی نمره عده ای را بالاتر می برند و نمره برخی را پائینتر می آورند. بدین سان، اثر عوامل تصادفی در

پیش‌آزمون و پس‌آزمون همدیگر را خنثی می‌کنند و جایی برای عمل پدیده<sup>۱</sup> بازگشت آماری در پس‌آزمون باقی نمی‌ماند.

#### ۶- گزینش

متغیر ناخواسته "گزینش"<sup>۱</sup> به اعمال سوگیریها و غرض‌ورزیهای<sup>۲</sup> ناآگاهانه و در گزینش آزمودنیها و گماردن آنان به گروههای آزمایشی و گواه دلالت دارد. انتخاب آزمودنیها برای شرکت در آزمایش به علت دسترسی به آنان، آشنائی با آنان، برخورداری آنان از هوش بالا، و یا به هر دلیل دیگر مواردی از اعمال نفوذ متغیر گزینش هستند. اعمال چنین سوگیریهای سبب می‌شود که گروههای مقایسه در آزمایش از آغاز با هم معادل نباشند. بدین معناکه، میانگینهای گروههای مقایسه در متغیر وابسته می‌توانند حتی پیش از اعمال مداخله آزمایشی باهم متفاوت باشند. در نتیجه، اگر این عامل مزاحم مهار نشود می‌تواند به‌صورت رقیبی برای مداخله آزمایشی در تبیین تغییرات حاصل در متغیر وابسته درآید. به بیان دیگر، تفاوتهای مشهود در نمره‌های پس‌آزمون گروههای آزمایشی و گواه می‌توانند ناشی از متغیر "گزینش" باشد و نه مداخله آزمایشی. تنها روشی که اعمال سوگیریها و غرض‌ورزیها را در گزینش آزمودنیها و گماردن آنان به گروههای مقایسه مانع می‌شود کاربرد نمونه‌گیری تصادفی از یک جامعه مشترک است.

#### ۷- افت آزمایشی

مراد از افت آزمایشی<sup>۳</sup> از دست دادن تعدادی از آزمودنیهای گروههای مقایسه، آزمایشی و گواه به‌گونه‌ای نابرابر و یا ناهم‌اند در فاصله زمانی میان پیش‌آزمون و پس‌آزمون است. این امر سبب می‌شود که ترکیب گروههای مقایسه به‌هنگام پس‌آزمون با هم متفاوت باشد. تفاوت در ترکیب گروههای مقایسه نیز به نوبه خود هیجان‌انگیزهای پس‌آزمون این گروهها تفاوت ایجاد می‌کند. بدین ترتیب، حتی اگر گروههای مقایسه در آغاز کار باهم معادل باشند، ممکن است به‌هنگام اجرای پس‌آزمون به دلیل افت آزمایشی باهم تفاوت پیدا کنند. این تفاوت به علت هیچ‌گونه تغییری

که در اثر مداخله آزمایشی در آزمودنیها ایجاد شده باشد نیست، بلکه صرفاً به این دلیل است که تعداد آزمودنیها در دو آزمایش نامساوی و ترکیب آنان ناهمبند بوده است.

در پژوهشهای آموزشی این مشکل در مطالعاتی پیش می‌آید که هدف آنها تعیین تأثیر آموزش بر دانشجویان است. این تأثیر بر اساس مقایسه دانشجویان سال اول آزمایشی را دریافت کرده‌اند (تعیین می‌شود. لیکن، پرواضح است که در فاصله میان سالهای اول و چهارم تعدادی از دانشجویان به‌علل مختلف تحصیلاتی دانشگاهی را رها می‌کنند. بنابراین، دانشجویان سال اول و سال چهارم، صرفنظر از تفاوت آموزش آنان، با هم معادل نیستند. بنابراین، در چنین مطالعاتی تلفات آزمایشی ممکن است علت تفاوت در میانگینها باشد.

#### ۸- تعامل با گزینش<sup>۱</sup>

اغلب عوامل تهدید کننده، اعتبار درونی، می‌توانند با متغیر ناخواسته گزینش تعامل کنند و نیروهای پدید آورند که به‌طور ساختگی به‌عنوان اثرهای مداخله، آزمایش قلمداد شوند. معمولترین این نیروها تعامل گزینش و رشد و نمو است. این تعامل وقتی حاصل می‌شود که گروههای مقایسه مرکب از افرادی باشند که با سرعتهای متفاوت رشد می‌کنند. مثلاً، هرگاه گروه آزمایشی مرکب از بیماران روانی و گروه گواه متشکل از افراد بهنجار باشد، تغییر در وضع گروه آزمایشی به هنگام پس‌آزمون می‌تواند در اثر فرایند بهبود خود به خود که خاص چنین گروهی است پیش‌آید و ارتباطی به مداخله آزمایشی، یعنی روان‌درمانی، نداشته باشد. پس چنین تغییری در آزمودنیهای گروه آزمایشی می‌توان بدون هیچ‌گونه روان‌درمانی و صرفاً "به دلیل تغییری درونی در آزمودنیها (رشد و نمو) حاصل شود. تعامل گزینش و رشد و نمو، یا تعامل گزینش و تاریخ، و یا تعامل گزینش و آزمودن می‌توانند با اثر مداخله، آزمایشی اشتباه گرفته شوند. تعامل گزینش و رشد و نمو حتی در جایی که میانگینهای پیش‌آزمون گروههای مقایسه با هم یکسان باشند نیز می‌تواند رخ بدهد. معمولترین این موارد جایی است که یک گروه از گروه دیگر رشد و نمو یا تغییر خود به خود بیشتری داشته باشد. در پژوهشهای آزمایشی حقیقی کاربرد روش تصادفی در گزینش آزمودنیها و

گماردن آنان به گروه‌های آزمایشی و گواه حفاظ مناسبی علیه‌هشتم عامل تهدید کننده فوق فراهم می‌کند. در این آزمایش‌ها گروه‌های آزمایشی و گواه از طریق انتخاب تصادفی با هم معادل شده‌اند. در چنین شرایطی عوامل مزاحم تاریخ، رشد و نمو، گزینش، افت آزمایشی، و تعامل گزینش با رشد و نمو مجالی برای اعمال نفوذ افتراقی در دو گروه مقایسه ندارند. همه آزمایش‌ها با شرایط آزمایشی و با ابزار سنجش یکسانی سر و کار دارند. این امر سبب مهار متغیرهای ناخواسته آزمایشی، و تغییر در ابزار سنجش می‌گردد. گزینش عمدی کسانی که نمره‌های بالا یا پایین دارند، جز در مواردی که آزمایش‌ها نخست بر مبنای یک متغیر مثلاً "نمره‌های پیش‌آزمون به صورت زوج درآمده و سپس به شرایط آزمایشی گمارده شده باشند، به عمل نمی‌آید. این موقعیت متغیر ناخواسته بازگشت آماری را مهار می‌کند.

#### عوامل تهدید کننده مهار ناپذیر اعتبار درونی

دسته دوم عوامل تهدید کننده اعتبار درونی عواملی را در بر می‌گیرد که با انتخاب تصادفی آزمودنی‌ها نمی‌توان آنها را مهار کرد. مهار این عوامل مستلزم در نظر گرفتن تمهیدات دیگری به غیر از انتخاب تصادفی در طرح پژوهشی است. این عوامل به شرح زیرند:

#### ۱- ابهام در مورد جهت رابطه علی

متغیر ناخواسته "ابهام در مورد جهت رابطه علی"<sup>۱</sup> به موقعیتی اشاره دارد که در آن به هنگام تبیین رابطه میان دو متغیر در آزمایش، کلیه تبیین‌های مربوط به متغیرهای موجه ناخواسته دیگر رد شده باشند و با وجود این، روشن نباشد که کدام متغیر علت است و کدام معلول. مثلاً، در آزمایش‌هایی که میزان سرپرستی سرپرستان را با بهره‌دهی زیردستان همبسته کرده‌اند به همبستگی منفی معنی داری دست یافته‌اند. لیکن، این همبستگی معلوم نمی‌سازد که آیا سرپرستی کمتر موجب بهره‌دهی بیشتر می‌شود یا بهره‌دهی بیشتر سبب سرپرستی کمتری گردد. این تهدید در بیشتر آزمایش‌ها از برجستگی برخوردار نیست، زیرا که ترتیب زمانی رویدادها در آنها مشخص است. همچنین این تهدید در آن مطالعات همبستگی که یکی از دو متغیر مربوط به عنوان

۱- ambiguity about the direction of causal influence

علت نسبتاً "غیرموجه است، نیز مشکلی به حساب نمی‌آید. مثلاً"، می‌دانیم که میان کاهش دمای محیط و افزایش در مصرف مواد سوختی رابطه وجود دارد؛ با وجود این، در چنین موقعیتی این استنباط موجه‌تر است که کاهش در دمای محیط موجب افزایش در مصرف مواد سوختی است و نه برعکس. لیکن، این تهدید می‌تواند در دیگر مطالعات همبستگی مشکلی به شمار آید.

## ۲- اشاعه مداخله آزمایشی یا تقلید از آن

متغیر ناخواسته "اشاعه مداخله آزمایشی یا تقلید از آن"<sup>۱</sup> به موقعیتی اشاره دارد که در آن یا مداخله‌های آزمایشی متضمن برنامه‌هایی باشند که در سطحی گسترده اشاعه می‌یابند، و یا آزمودنیهای گروههای آزمایشی و گواه می‌توانند یا یکدیگر ارتباط برقرار سازند. در این صورت، آزمودنیهای گروه گواه می‌توانند به یادگیری محتوای آن برنامه‌ها بپردازند و در واقع به گونه‌ای نامنظم و غیر رسمی به دریافت مداخله آزمایشی نائل آیند. در چنین مواردی، آزمایش بی‌اعتبار می‌شود، زیرا عملاً "تفاوتی میان گروههای آزمایشی و گواه، از لحاظ دریافت مداخله آزمایشی، وجود ندارد. در بسیاری از شبه آزمایشها مطلوب این است که گروههایی به کار ببریم که حتی المقدور از همه جهات جز دریافت مداخله آزمایشی با هم مشابه باشند. برای این منظور، معمولاً از گروههایی استفاده می‌شود که از بخشهای سازمانی مجاور انتخاب شده‌اند. لیکن نفس مجاورت واحدهای سازمانی که گروههای مقایسه از آنها انتخاب شده‌اند سبب می‌شود که همه این گروهها بتوانند به گونه‌ای رسمی یا غیررسمی در معرض مداخله آزمایشی قرار بگیرند. بدین ترتیب، اثر مداخله آزمایشی زائل می‌شود و آزمایش بی‌اعتبار می‌گردد.

## ۳- معادل سازی جبرانی مداخله آزمایشی

متغیر ناخواسته "معادل سازی جبرانی مداخله آزمایشی"<sup>۲</sup> را می‌توان به شرح زیر بیان داشت. گاهی مداخله آزمایشی برای گروه آزمایشی منافعی در بردارد. در چنین موقعیتی این امکان وجود دارد که گروه گواه احساس نا برابری کند و

۱ - diffusion or imitation of the treatment

۲ - Compensatory coualization of treatment

در نتیجه نسبت به آزمایش بی‌میل و دلسرد شود. آنچه که اغلب برای اجتناب از وقوع این پدیده انجام می‌گیرد، این است که منافع، مشابهی نیز در اختیار گروه گواه قرار می‌گیرد. لیکن، خود این عمل مشکل دیگری را به وجود می‌آورد. توضیح این که گروه گواه ممکن است در نتیجه دریافت منافع، در پس‌آزمون عملکردی نظیر عملکرد گروه آزمایشی داشته باشد. در صورتی که این اتفاق روی دهد، تفاوت میان گروه‌های آزمایشی و گواه از میان می‌رود، و مداخله آزمایشی نمی‌تواند آن چنان که باید تأثیر خود را نمایان سازد. در واقع، مانند این است که گروه‌های آزمایشی و گواه هر دو در معرض مداخله آزمایشی قرار بگیرند و در نتیجه تفاوت معنی‌داری در عملکردهای آنها پدید نیاید. لیکن، عدم دستیابی به تفاوتی معنی‌دار میان عملکردهای گروه‌های مقایسه‌بدین معنا نیست که مداخله آزمایشی تأثیری نداشته است، بلکه بدین معناست که هر دو گروه کمابیش مداخله آزمایشی را دریافت کرده‌اند.

#### ۴- رقابت جبرانی

متغیر ناخواسته "رقابت جبرانی"<sup>۱</sup> به موقعیتی اشاره دارد که در نتیجه کم‌آوردن علنی افراد یا واحدهائی (بخشهای) از سازمانها به شرایط آزمایشی و گواه، رقابت اجتماعی پیش‌آید. در چنین موقعیتی، گروه گواه به‌تلاشی ناشی از هم‌چشمی و رقابت برانگیخته می‌شود، این امر سبب می‌گردد که تفاوت مورد انتظار میان عملکردهای گروه‌های آزمایشی و گواه کاهش یابد و یا حتی معکوس شود. این متغیر به‌ویژه درجائی احتمال وقوع دارد که یا واحدهای یکپارچه و کامل سازمانی (از قبیل، بخشها، کارخانه‌ها، گروه‌های کار، و جزاینها) به شرایط مختلف آزمایشی و گواه گمارده شوند، و در صورت موفقیت آمیز بودن مداخله آزمایشی به گروه گواه زیان و خسارت وارد آید. نمونه‌ای از اعمال نفوذ این متغیر را می‌توان در مطالعاتی ملاحظه کرد که توسط سارتسکی<sup>۲</sup> (۱۹۷۲) بازنگری شده‌است. در این مطالعات یادگیری کودکانی که توسط معلمان معمولی تدریس می‌شدند با یادگیری کودکانی که توسط معلمان قراردادی آموزش می‌دیدند، مقایسه شده‌است. در این مطالعات گروه گواه معلمان معمولی هستند که به صورت ثابت و ماهیانه حقوق دریافت می‌کنند. گروه آزمایشی متشکل از معلمانی است که به صورت قراردادی استخدام شده‌اند و برحسب پیشرفت آموزشی کودکان به آنان حقوق پرداخت

می‌شود. مداخله<sup>۱</sup> آزمایشی نحوه، تدریس معلمان قراردادی است. در ضمن این آزمایشها معلوم گردید که عملکرد آموزشی کودکانی که توسط معلمان معمولی تدریس می‌شدند در اثنای دوره‌ای که آزمایش اجرا می‌شد، از سالهای گذشته بهتر بود. این بدان معناست که تفاوت یادگیری کودکانی که توسط معلمان معمولی آموزش دیده بودند با یادگیری کودکانی که توسط معلمان قراردادی تدریس شده بودند، کاهش یافته‌است.

۵- دلسردی ناشی از آزردهی پاسخگویانی که مداخله‌ای با مطلوبیت کمتر دریافت کرده‌اند.<sup>۱</sup>

هنگامی که آزمایش علنی است، واکنش گروهی که مداخله‌ای با مطلوبیت کمتر دریافت کرده است به جای توأم بودن با رقابت جبرانی، می‌تواند با رنجش و دلسردی همراه باشد. اگر چنین آزمایشی در محیطی صنعتی انجام شود، گروه گواه می‌تواند با کاهش بهره‌دهی خود، از کارفرما انتقام بگیرد. این موقعیت احتمالاً به تفاوت معنی‌داری میان میانگینهای پس‌آزمون گروههای آزمایش و گواه منجر می‌شود. انتساب این تفاوت به مداخله<sup>۲</sup> آزمایشی خطاست. در واقع، تفاوت حاصل ناشی از پائین آمدن عملکرد پس‌آزمون گروه گواه در واکنش به نامطلوبیت موقعیت خویش است و نتیجه<sup>۳</sup> بهبود عملکرد گروه آزمایشی به دنبال اعمال مداخله<sup>۴</sup> آزمایشی نیست. شاید مناسبتر باشد که گروه گواه را گروه مداخله<sup>۵</sup> آزردهی بنامیم. لازم به تذکر است که این پدیده منحصر به گروههای گواه که در معرض مداخله<sup>۶</sup> آزمایشی قرار نگرفته‌اند نیست، بلکه می‌تواند هر جا که مداخله‌های مختلف از لحاظ مطلوبیت تفاوت داشته‌اند و آزمودنیها از آن آگاه باشند، روی دهد.

پرتال جامع علوم انسانی

ع- تاریخ موضعی

عامل مزاحم " تاریخ موضعی"<sup>۲</sup> هنگامی عمل می‌کند که یا مداخله<sup>۱</sup> آزمایشی در یک جلسه اعمال می‌گردد و یا داده‌های لازم از هر یک از گروههای آزمایشی و گواه همزمان و در یک جلسه، گروهی واحد جمع‌آوری شود. این موقعیت ممکن است سبب

۱- Resentful demoralization of respondents receiving less desirable treatment

۲- Local history



شود که رویدادهای ویژه‌ای منحصرًا " در یکی از دو گروه آزمایشی و گواه اتفاق افتد و همه اعضای آن گروه راتحت تأثیر قرار دهد. چنین تأثیری می‌تواند با اثر مداخله آزمایشی اشتباه گرفته شود. بدین سان، درنمره‌های پس‌آزمون گروههای مقایسه ممکن است تفاوتی پدید آید که ناشی از آن رویدادهای ویژه باشد و نه مداخله آزمایشی. گماردن تصادفی آزمودنیها به گروههای آزمایشی و گواه این عامل مزاحم را مهار نمی‌کند، زیرا این عامل پس از گماردن آزمودنیها به این گروهها و از رهگذر رویدادهایی که منحصرًا " در رابطه با یکی از آنها پیش می‌آید، اعمال نفوذ می‌کند. مثلاً، اگر هنگام پاسخگویی گروه آزمایشی به پس‌آزمون حادثه غیر مترقبه‌ای، از قبیل از حال رفتن یکی از آزمودنیها، روی دهد، این حادثه می‌تواند بر عملکرد آن گروه در پس‌آزمون تأثیر ببخشد، درحالی که گروه گواه در معرض چنین حادثه‌ای قرار نمی‌گیرد و از تأثیر سوء آن نیز مصون است. بدین ترتیب، تفاوت میان عملکردهای گروههای آزمایشی و گواه می‌تواند ناشی از این حادثه باشد و نه مداخله آزمایشی، به منظور جلوگیری از اعمال نفوذ این عامل مزاحم باید هر جا که میسر است مداخله را نسبت به تک تک آزمودنیها و یا نسبت به بسیاری از گروههای کوچک در جلسات متعدد اعمال کرد. همچنین بهتر است داده‌های مربوط به پیش‌آزمون و پس‌آزمون از تک تک افراد و یا در جلسات متعدد از گروههای کوچک بسیار جمع‌آوری شود. افزون بر آن، جلسات متعدد اعمال مداخله آزمایشی و جمع‌آوری داده‌ها باید در صورت امکان از لحاظ زمانی و مکانی طبق یک برنامه تصادفی برگزار شوند. این کار باعث می‌شود که هیچ رویداد ویژه‌ای نتواند تأثیر تعیین‌کننده‌ای بر داده‌های حاصل از همه اعضای یک گروه باقی‌گذارد و بدین گونه، عامل " تاریخ موضعی " مهار می‌گردد. در صورتی که آزمایش بدین نحو انجام گیرد، درجات آزادی<sup>۱</sup> مناسب برای تجزیه و تحلیل‌های آماری براساس تعداد گروهها تعیین می‌شود نه افراد. احتمال اعمال نفوذ " تاریخ موضعی " در شبه‌آزمایشها اغلب بیشتر از آزمایشهای حقیقی است. علت این امر این است که

---

۱- مراد از درجات آزادی (degrees of freedom) تعداد اطلاعات مستقلی است که شاخص آماری نمونه براساس آنها به وجود می‌آید. آماره‌های مختلف با توجه به درجات آزادی متفاوت محاسبه می‌شوند. به طور کلی، مقدار درجات آزادی یک شاخص آماری بستگی دارد به حجم نمونه و تعداد آماره‌هایی که در محاسبه آن به کار رفته‌اند.

در اغلب شبه آزمایشها گروههای مقایسه با هم معادل نیستند و مداخله با تفاوتهای موجود در این گروهها آمیخته می‌شود.

### ارزیابی اعتبار درونی

برآورد اعتبار درونی هر آزمایش فرآیندی مبتنی بر استدلال قیاسی<sup>۱</sup> است که در آن محقق باید منتقد سختگیر خود باشد. وی باید به گونه‌ای اصولی بیندیشد که چگونه هر یک از عوامل مزاحم می‌تواند بر داده‌های او تأثیر بخشیده باشد. آن گاه باید داده‌ها را بررسی کند تا دریابد که کدام یک از این تهدیدهای بالقوه در داده‌های حاصل از آزمایش او تأثیری نداشته‌اند. اگر بتواند امکان اعمال نفوذ همه آنها را رد کند، ارائه نتیجه گیریهایی مطمئن در خصوص وجود یک رابطه علی میسر است. لیکن، اگر رد همه تهدیدها، به دلیل در اختیار نداشتن داده‌های مناسب و یا به سبب اعمال نفوذ برخی از متغیرهای ناخواسته ممکن نباشد، محقق باید چنین نتیجه‌گیری کند که رابطه یافته شده میان دو متغیر ممکن است علی باشد یا نباشد.

همان طور که قبلاً اشاره شد روش تصادفی در پژوهشهای آزمایشی حقیقی اثر هشت عامل نخست تهدید کننده اعتبار درونی را مهار می‌سازد. لیکن، در رابطه "با پنج تهدید آخر اعتبار درونی برای آزمایشهای حقیقی، و نیز در رابطه با همه تهدیدهای اعتبار درونی برای شبه آزمایشها موقعیت به گونه‌ای دیگر است. محقق باید نخست عوامل مزاحم را آشکارا مطرح سازد و آن گاه در صورت امکان و با اتکا به تمهیداتی که در آزمایش برای مهار آنها اندیشیده است یکی یکی رد کند. بدین سان، در شبه آزمایشها وظیفه محقق سنگین تر است. با وجود این، در این آزمایشها استنباط علی نهایی به استحکام موقعی که آزمایشی واقعی اجرا می‌گردد، نیست.

### اعتبار نتیجه گیری آماری

اعتبار نتیجه گیری آماری به این معنا اشاره دارد که اثر یا رابطه یافته شده در آزمایش باید واقعی باشد و نه حاصل عوامل تصادفی یا خطا. به بیان دیگر، اگر در پی اعمال یک مداخله آزمایشی، تفاوتی میان میانگینهای پس از آزمون گروههای آزمایشی و گواه مشاهده شود، باید اطمینان حاصل کرد که این تفاوت واقعی است و از اعمال

نفوذ عوامل تصادفی و خطا حاصل نشده است. اعتبار نتیجه‌گیری آماری مبتنی است بر کاربرد روشهای مناسب آماری برای تجزیه و تحلیل داده‌های آزمایش. این کار مشخص می‌سازد که اثر یا رابطه یافته شده در آزمایش با چه احتمالی واقعیت دارد. به احتمال واقعی بودن یافته‌های آزمایشها سطح معنی داری<sup>۱</sup> گویند. معمولاً، آزمایشها به منظور تصمیم‌گیری در رابطه با مسائل مورد آزمایش، انجام می‌گیرند. به همین جهت محقق مایل است بداند که آیا مداخله آزمایش او اثری داشته است یا نه، یا بر مبنای اطلاعات موجود نمی‌تواند چنین استنباطی به عمل آورد. گاهی به دلایل عقلی تنها اثرهایی به حساب می‌آیند که مقدار آنها از حد معینی کمتر نباشد. مثلاً، در آزمایشی که هدف آن تعیین اثر یک متغیر مستقل بر بهره دهی و غیبت کارکنان است، مداخله آزمایشی تنها هنگامی مؤثر محسوب می‌شود که بهره دهی ۱۰٪ افزایش یابد، یا غیبت کارکنان از ۲٪ همه روزهای کار آنان تجاوز نکند. این مقادیر گویای آنند که علاقه ما به مقدار یک اثر است نه به خود وجود آن.

هرچند، معمولاً "علاقه" ما به خود وجود یک اثر است. اغلب می‌خواهیم بدانیم از تفاوت معنی دار آماری میان آماره‌های (شاخصهای آماری حاصل از نمونه‌ها، مثلاً) میانگین، ضریب همبستگی، و جز اینها) همانند متعلق به گروههای آزمایشی و گواه می‌توان یک اثر را استنباط کرد؟ در اینجا با تهدیدهای سر و کار داریم که مانع نتیجه‌گیریهای معتبر در خصوص نفس وجود یک اثر در پی اعمال مداخله آزمایشی می‌شوند. البته، این به معنای بی‌اهمیت تلقی کردن برآورد مقدار اثرها نیست، بلکه بدان معناست که این امر هنگامی از اهمیت برخوردار می‌شود که نخست بتوان نفس وجود یک اثر را به گونه‌ای مستدل استنباط کرد.

سننا<sup>۲</sup>، در رابطه با نتیجه‌گیری و استنباط یک اثر از داده‌های گروههای مقایسه، سطوح معنی دار دلخواهی متداول شده است. مشهورترین آنها سطح معنی دار ۰/۰۵ است. معنی دار بودن یک یافته در سطح ۰/۰۵ بدین معناست که تنها در ۵ مورد از ۱۰۰ مورد احتمال می‌رود که آن یافته نادرست باشد و در ۹۵ مورد دیگر درست است. هنگامی که سطح معنی دار ۰/۰۵ انتخاب شود، روابط و تفاوت‌هایی که احتمال مشاهده آنها از ۰/۰۵ بیشتر باشد غیرواقعی پنداشته می‌شوند. با وجود این، می‌توانیم در این نتیجه‌گیری که میانگینهای جامعه‌های مربوط باهم متفاوتند بر خطا باشیم، حتی، اگر

سطح معنی داری از ۰/۰۵ کمتر باشد. همچنین می‌توانیم در این نتیجه‌گیری که میانگینها با هم تفاوتی ندارند بر خطا باشیم، حتی اگر سطح احتمال بالاتر از ۰/۰۵ باشد.

اعتبار نتیجه‌گیری آماری، همان‌طور که پیش از این متذکر گردید، با اعتبار درونی شباهت دارد و حتی از لحاظی مورد خاصی از اعتبار درونی تلقی می‌شود. درک شباهت میان اعتبار درونی و اعتبار نتیجه‌گیری آماری مستلزم دریافت این موضوع است که تغییر در متغیر وابسته می‌تواند به سه علت روی دهد. (۱) تغییر می‌تواند واقعی باشد و منحصرًا "از دستکاری متغیر مستقل حاصل شده باشد. این موقعیت مطلوب هر آزمایش است. (۲) تغییر می‌تواند ناشی از سوگیری باشد. سوگیری در این جا به معنی تأثیر متغیرهای دیگری به غیر از مداخله آزمایشی است که به‌گونه‌ای منظم و اصولی میانگینهای گروههای مقایسه را در متغیر وابسته تحت تأثیر قرار می‌دهند. اینها همان عوامل تهدید کننده، مهار پذیر و مهار ناپذیرند که زائل ساختن تأثیر آنها به اعتبار درونی پژوهشهای آزمایشی می‌انجامد. (۳) تغییر می‌تواند ناشی از خطا باشد. خطا در این جا به معنی متغیرهای دیگری به غیر از مداخله آزمایشی است که به نحوی نامنظم و غیراصولی بر متغیر وابسته تأثیر می‌گذارند. حذف تأثیر این متغیرها به اعتبار نتیجه‌گیری آماری پژوهشهای آزمایشی منجر می‌شود. بدین سان، پژوهشی آزمایشی هنگامی از اعتبار درونی برخوردار است که تغییر در متغیر وابسته ناشی از سوگیری نباشد و زمانی واجد اعتبار نتیجه‌گیری آماری است که تغییر در متغیر وابسته حاصل خطا نباشد. بنابراین، این نتیجه‌گیری غلط از یک شبه آزمایش " که متغیر "الف" علت متغیر "ب" است"، می‌تواند به دو دلیل روی دهد. عوامل تهدید کننده، اعتبار درونی ممکن است میانگینهای مربوط را تحت تأثیر قرار داده و یا عوامل خطا ممکن است تفاوت میان میانگینها را به وجود آورده باشند.

همچنین، این نتیجه‌گیری غلط که متغیر "الف" اثری بر متغیر "ب" ندارد، نیز می‌تواند به دو علت پیش آید. عوامل تهدید کننده، اعتبار درونی ممکن است با اعمال نفوذ بر میانگینهای مربوطه تفاوت واقعی میان آنها را پنهان کنند، یا عوامل خطا ممکن است سبب کاهش تفاوت واقعی میان میانگینها شوند. کوتاه سخن، اعتبار درونی با منابع سوگیری منظم سروکار دارد، در حالی که سروکار اعتبار نتیجه‌گیری آماری با منابع خطاست. این خطاها بر اش نمره‌ها را افزایش و احتمال حصول نتایج معنی‌دار آماری را کاهش می‌دهند.

عوامل تهدید کننده اعتبار نتیجه‌گیری آماری به شرح زیرند:

۱- توان آماری اندک<sup>۱</sup>

به احتمال تصمیم صحیح در مورد رد فرضیه صفر<sup>۲</sup> نادرست توان آزمون آماری گفته می‌شود. به دیگر سخن، توان آزمون آماری بیانگر میزان احتمالی است که در صورت تفاوتی واقعی تحقیق آن را نشان می‌دهد. توان آزمون آماری برابر است با  $1 - \beta$  و  $\beta$  احتمال خطای نوع دوم است. در تلاش همزمان جهت کاهش مقدار  $\alpha$  (احتمال رد یک فرضیه صفر صحیح - خطای نوع اول) و کاهش  $\beta$  (احتمال عدم رد یک فرضیه صفر غیر صحیح - خطای نوع دوم) محقق با مشکل رو به رو می‌شود. از یک سو،  $\alpha$  و  $\beta$  با هم رابطه معکوس دارند. در نتیجه با کاهش میزان احتمال یکی میزان احتمال دیگری افزایش می‌یابد. از سوی دیگر، میان  $\beta$  و توان آماری نیز رابطه معکوس وجود دارد، در نتیجه وقتی از مقدار  $\alpha$  کاسته می‌شود، مقدار  $\beta$  افزایش پیدا می‌کند، لیکن در توان آزمون آماری  $(1 - \beta)$  نقصان روی می‌دهد. چون  $\alpha$  و  $\beta$  با هم رابطه معکوس دارند، محقق باید یکی را فدای دیگری سازد، یعنی برای کاهش خطای نوع دوم ( $\beta$ ) باید احتمال خطای نوع اول ( $\alpha$ ) را افزایش دهد. در موارد زیر خطای نوع دوم افزایش و توان آماری کاهش می‌یابد:

(الف)  $\alpha$  یا سطح معنی داری پائین تعیین شود (مثلاً "،  $P \leq 0/001$  ، (ب) حجم نمونه‌ها کوچک باشد، (ج) فرضیه‌های یک سویه<sup>۳</sup> به گونه‌ای نادرست برگزیده و آزمایش شوند و (د) انواع آماره‌های واجد توزیع آزاد<sup>۴</sup> یا بی پارامتر، برای آزمایش فرضیه به کار روند.

۱ - Low statistical power

۲ - Null hypothesis

۳- فرضیه یک سویه (One-tailed hypothesis) فرضیه‌ای است که جهت تأثیر متغیر مستقل را مشخص کرده باشد. مثلاً "، این فرضیه که (عدم وجود تفاوت میان والدین سبب بزهکاری فرزندان آنان می‌شود، فرضیه‌ای یک سویه است.

۴- مراد از شاخصهای آماری واجد توزیع آزاد (distribution-free) آن روشهای تجزیه و تحلیل داده‌هاست که هیچ گونه فرضی درباره توزیع حقیقی داده‌ها در جامعه به عمل نمی‌آورند، یا در معنایی محدودتر برای این مسئله که توزیع حقیقی در جامعه یک توزیع بهنجار است فرضی قائل نیست.

## ۲- تخلف از فرضها یا شرایط اصلی آزمون آماری<sup>۱</sup>

کاربرد اغلب آزمونهای آماری مربوط به آزمایش فرضیه صفر مستلزم برقراری فرضهای خاصی است تا بتوان نتایج حاصل از تجزیه و تحلیل داده‌ها را به گونه‌ای با معنا تفسیر کرد. بدین سان، باید از فرضهای ویژه یک آزمون آماری خاص اطلا داشته، و هر جا ممکن باشد، باید برقراری این فرضها را در رابطه با داده‌های یک آزمون مورد بررسی قرار داد.

مثلاً، آزمون  $t$  برای تفاوت بین میانگینهای دو گروه مستقل مبتنی بر سه فرض است: ۱. آزمودنیهای گروههای مقایسه به صورت تصادفی از جامعه‌های مربوطه انتخاب شده باشند. ۲. شکل توزیع پراکندگی در جامعه‌های مورد نظر بهنجار باشد. ۳. همگنی واریانس برقرار باشد یعنی پراشهای نمره‌های دو جامعه با هم مساوی باشند. پیش از استفاده از هر آزمون آماری باید نسبت به برقراری فرضهای آن بررسیهای لازم به عمل آید. بعضی از آزمونهای آماری در شرایطی خاص نسبت به پاره‌ای از فرضهای خود حساس نیستند و تخلف از آن فرضها نتیجه گیری آماری را دچار اشکال نمی‌سازد. لیکن در موارد دیگر تخلف از فرضهای آزمون آماری ممکن است به نتیجه گیری نادرست بینجامد. از آنجا که اهمیت فرضهای ویژه بستگی به نوع آزمونی دارد که به کار می‌رود، برای حصول اطمینان از شناخت و فهم فرضهای یک آزمون آماری خاص و به منظور چیره شدن بر مشکلات ناشی از آنها، بهتر است به یک مآخذ مناسب آماری رجوع شود.

## ۳- نتیجه گیری و مشکل ضریب خطا

عامل مزاحم "نتیجه‌گیری و مشکل ضریب خطا"<sup>۲</sup> در شرایطی اعمال نفوذ می‌کند که چند مقایسه میان تفاوت‌های میانگینها مبر باشد و این تشخیص داده نشود که تفاوت در برخی از این مقایسه‌ها ممکن است صرفاً "به دلیل عوامل تصادفی به سطح معنی داری برسد. مثلاً"، اگر در یک آزمایش ۵ گروه مختلف داشته باشیم، می‌توان ۱۰ مقایسه مختلف میان میانگینهای آنها به عمل آورد. در این آزمایش اگر سطح معنی داری

<sup>۱</sup>-Violated assumptions of statistical tests

<sup>۲</sup>-Fishing and the error rate problem

راه ۰/۰۵ انتخاب کنیم . ممکن است یک یا چند تفاوت تصادفاً به این سطح معنی‌داری برسد و ما آنها را واقعی تلفی کنیم ، لیکن در واقع ، نتیجه گیری ما نادرست باشد . این امر بدان جهت است که سطح معنی‌داری آماری منتخب ما تنها از عهده انجام یک مقایسه بر می‌آید و برای یک مقایسه در نظر گرفته شده است ، در حالی که ما آن را در رابطه با چند مقایسه به کار برده ایم . این موقعیت موردی از خطای نوع اول است .

برای چیره شدن بر این مشکل رایان<sup>۱</sup> (۱۹۵۹) از جمله دیگران ، روشی را پیشنهاد کرده است . این روش مستلزم محاسبه یک سطح جدید معنی‌داری بر اساس سطح معنی‌داری اولیه و تعداد مقایسه‌هاست . این سطح معنی‌داری جدید از تقسیم سطح معنی‌داری اولیه بر تعداد مقایسه‌ها به دست می‌آید . مثلاً ، اگر محقق سطح معنی‌داری ۰/۰۵ را انتخاب کند و بخواهد ۱۰ مقایسه به عمل آورد ، از تقسیم ۰/۰۵ به ۱۰ ، سطح معنی‌داری جدیدی که مساوی با ۰/۰۵۵ است به دست می‌آید . آن گاه باید مقدار متناسب با این سطح معنی‌داری جدید را در جدولهای مربوط پیدا کرد . بدیهی است که این مقدار از مقدار متناسب با سطح معنی‌داری ۰/۰۵ بیشتر است . اگر مقادیر مشاهده شده مربوط به تفاوت‌های میان میانگینها در آزمایش اوبرابریا بیشتر از این مقدار باشد ، می‌توان گفت که همه آنها در سطح ۰/۰۵ معنی‌دار هستند . این عمل در حکم یک محکم کاری است که به هنگام داشتن چند مقایسه به منظور حصول یک سطح معنی‌داری آماری حقیقی به عمل می‌آید .

روش دیگر برای رفع مشکل ضریب خطا مستلزم کاربرد آزمونهای محافظه کارانه مقایسه چندگانه توکی<sup>۲</sup> آیا<sup>۳</sup> است (ر.ک. به هیز<sup>۴</sup> ، ۱۹۷۳) . هنگامی که در یک آزمایش عاملی چند متغیر وابسته وجود داشته باشد ، از یک راهبرد تحلیل واریانس چند متغیری می‌توان استفاده کرد تا معلوم شود آیا معنی‌داری هر کدام از آزمونهای تک متغیری در رابطه با یک اثر اصلی یا تعاملی ناشی از تصادف است یا از دستکاریهای آزمایشی .

۱ - Rayan

۲ - Tukey

۳ - Scheffé

۴ - Hays





## ۶- عوامل نامربوط تصادفی در محیط آزمایش

عامل تهدید کننده<sup>۱</sup> عوامل نامربوط تصادفی در محیط آزمایش<sup>۱</sup> به متغیرهایی اشاره دارد که تصادفاً در محیط آزمایش وجود دارند و ممکن است بر نمره‌های متغیر وابسته تأثیر بگذارند و پراش خطا را افزایش دهند. این عامل تهدید کننده را می‌توان به طرق زیر مهار کرد: الف) انتخاب آن محیط‌های آزمایشی که از عوامل نامربوط تغییر زای خارجی عاری باشند، ب) انتخاب روش‌های آزمایشی که توجه آزمودنی‌ها را به مداخله آزمایشی معطوف می‌سازند و از برجستگی متغیرهای محیطی می‌کاهند. ج) تعیین آن عوامل موجود در محیط پیرامون آزمایش که برای همه گروه‌های شرکت کننده در آزمایش مشترکند و ایجاد پراش می‌کنند و به کاربردن آنها در تحلیل آماری.

## ۷- ناهمگنی تصادفی پاسخگویان

عامل مزاحم " ناهمگنی تصادفی پاسخگویان"<sup>۲</sup> عبارت از این است که آزمودنی‌های آزمایش می‌توانند از لحاظ عوامل عمده‌ای که با متغیر وابسته همبستگی دارند، با هم متفاوت باشند. هرچه پاسخگویان، از لحاظ متغیرهای همبسته با متغیر وابسته، ناهمگنتر باشند، پراش خطا، چه در آزمون اثرهای مداخله و چه در آزمون اثرهای تعاملی افزایش بیشتری خواهد یافت. این عامل نامربوط را می‌توان با انتخاب نمونه‌هایی از جامعه‌های همگن مهار کرد، لیکن این کار به ضرر اعتبار بیرونی آزمایش خواهد بود. راه دیگر مهار این عامل مزاحم این است که ویژگی‌های مربوط پاسخگویان را اندازه‌گیری کرد و تحت شرایط مناسب (ر.ک. به الاشف<sup>۳</sup>، ۱۹۶۹) به منظور طبقه بندی<sup>۴</sup> آزمودنی‌ها یا به عنوان متغیرهای همبسته<sup>۵</sup> به کاربرد.

۱ - Random irrelevancies in the experimental setting

۲ - Random heterogeneity of respondents

۳ - Elashoff

۴ - Blocking

۵ - Covariates

## منابع

Anastasi, Anne, *Differential Psychology*, (3rd ed.) New York; Macmillan, 1958. Boruch, R.F., & Gomez, H. Boruch, R.F., & Gomez, H. "Sensitivity, bias and theory in impact evaluation *Professional Psychology*, 1977, 8, 411-34.

Campbell, D.T. "Factors relevant to the validity of experiments in social settings". *Psychol. Bull.*, 1957, 54, 297-312.

Campbell, D.T., & Stanley, J.C. "Experimental and quasiexperimental designs for research on teaching". In: N.L. Gage (Ed.), *Handbook of research on teaching*. Chicago; Rand McNally, 1963. (Also published as experimental and quasi-experimental designs for research. Chicago; Rand McNally, 1966).

Cane V.R. & Heim, A.W. "The effects of repeated testing: Further experiments and general conclusions". *Quart. J. Exp. Psychol.* 1950, 2, 182-195.

Cook, T.D. & Campbell, D.T. *Quasi-Experimentation; Design & Analysis issues for field settings*. Boston; Houghton Mifflin Company, 1979.

Cook, T.D., & Campbell, D.T. "The design and conduct of quasi-experiments and true experiments in field settings". In: M.D. Dunnette (Ed) *Handbook of industrial and organizational psychology*. Chicago; Rand, McNally, 1966 (pp. 223-326).

Crook, M.N. "The constancy of neuroticism scores

and self-judgments of constancy". *J. Psychol.* 1937,4, 27-34.

Elashoff, J.D. "Analysis of Covariance: A delicate instrument", *Educational research Journal*, 1969,6,383-401.

Hays, W.L. *Statistics*, New York, Holt, Rinehart and Winston, 1963.

Ryan, T.A. "Multiple Comparisons in psychological research". *Psychological Bulletin*, 1959,56,26-47.

Saretsky G. "The OEO P.C. experiment and the John Henry effect". *Phi delta-Kappan*, 1972, 53,579-581.

Windle, C. "Test-retest effect on personality questionnaires", *Educ. Psychol. Measmt.* 1954, 14,617-633.



پروپش گاد علوم انسانی و مطالعات فرہنگی

پرتال جامع علوم انسانی