

بررسی پاره‌ای از خطاهای رایج آماری در پژوهش‌های رفتاری

دکتر حجت‌الله فراهانی*
دکتر حمیدرضا عریضی**
دکتر عباس رحیمی‌نژاد***
امیر قمرانی****

چکیده

هدف از پژوهش حاضر بررسی پاره‌ای از خطاهای رایج آماری، در پژوهش‌های رفتاری است. بر این اساس، چهار نوع خطای آماری رایج در پژوهش‌های رفتاری مشاهده شده است. خطاهایی نظیر خطای نوع اول (α) و تورم آن، خطای کاربرد نادرست فرضیه‌ی صفر، خطای گزارش نادرست مقدار P و خطای عدم گزارش حجم اثر که در این مقاله، شرح و بسط مفهومی می‌شود. افزون بر این، ملاحظات روش‌شناختی و کاربردی توجه به خطاهای مذکور، به تفصیل مورد بحث قرار می‌گیرد.

واژه‌های کلیدی: خطاهای آماری، پژوهش‌های رفتاری، ملاحظات روش‌شناختی.

مقدمه

پیشینه‌ی پژوهش‌های علوم رفتاری، انباشته از مباحثی پیرامون رویکردهای مناسب در استفاده از طرح‌ها، آزمون‌ها و روش‌های آماری است. برای افزایش دقت نتایج به دست آمده در پژوهش‌های علوم رفتاری، روش‌های آماری بازننگری شده‌اند و هدف از این بازننگری‌ها، کمک به بهبود و افزایش دقت نتایج پژوهش‌ها بوده است. در این بخش، خطاهای آماری رایج در این پژوهش‌ها بررسی می‌شود.

۱. خطای نوع اول (α) و میزان تورم آن

میزان ارتکاب به خطای نوع اول^۱ را سطح معناداری^۲ می‌نامند. خطای نوع اول، ردّ فرض صفر، در صورت درست بودن آن است (دلاور، ۱۳۷۶). اغلب، محقق یک سطح خاص معناداری را که تا اندازه‌ی زیادی قراردادی است، انتخاب می‌کند. انتخاب سطوح معناداری (۰/۰۵ یا ۰/۰۱) اغلب، در بین پژوهشگران، مرسوم و متداول است (فرگوسن و تاکانه^۳، ۱۳۷۷). این سطح، احتمال بروز تصادفی تفاوت مشاهده شده در جامعه، در صورت درست بودن فرضیه‌ی صفر را نشان می‌دهد. اگر این احتمال کمتر از (۰/۰۵) باشد، گفته می‌شود که این تفاوت در سطح (۰/۰۵) یا ۵ درصد و کمتر معنادار است. سطوح (۰/۰۵) و (۰/۰۱) مبین این هستند که تفاوت مشاهده شده، یک تفاوت واقعی و معنادار است که ناشی از شانسی، تصادف و خطای نمونه‌گیری نیست. از آنجایی که فرض صفر در صورتی رد می‌شود که آمار نمونه در ناحیه ردّ قرار گرفته باشد، به تبع احتمال ردّ فرض صفر در جایی که حقیقی است، دقیقاً معادل سطح معناداری است (مولر^۴ و همکاران، ۱۳۸۶).

تصمیم در مورد رد یا پذیرش فرض صفر در سطح معناداری α ، معمولاً بدون توجه به خطای نوع دوم اتخاذ می‌شود (فرگوسن و تاکانه، ۱۳۷۷). خطای نوع دوم هنگامی روی می‌دهد که فرض صفر واقعاً غلط باشد؛ اما رد نشود. احتمال ارتکاب به خطای نوع دوم را معمولاً با β نشان می‌دهند (هاراوی^۵، ۱۳۷۹). احتمال ارتکاب به خطای نوع دوم (β) انتخاب نمی‌شود. عملاً در پژوهش‌ها، از قبل α و n (حجم

1. error Type I
2. significant level
3. Fergusan and Takane
4. Mueller
5. Haraway

نمونه) معلوم می‌گردد. با در نظر گرفتن α و n می‌توان β را معلوم کرد، به دلیل وخامت خطای نوع اول باید از بزرگ‌ترین مقداری که برای α جایز است، استفاده کرد (تریولا^۱، ۱۳۷۴). هر قدر تفاوت بین μ_1 و μ_2 (میانگین‌های مورد مقایسه) بیشتر باشد، احتمال کم‌تری برای پذیرش فرض صفر وجود خواهد داشت؛ بنابراین، با توجه به تفاوت خاص بین μ_1 و μ_2 و مقدار مشخص N ، با کاهش α ، مقدار β افزایش خواهد یافت. فرگوسن و تاکانه معتقدند که هرگاه سطح معناداری دقیق انتخاب شود، در صورتی که واقعاً تفاوت نسبتاً اندکی بین μ_1 و μ_2 وجود داشته باشد، نمی‌توان فرض صفر را رد کرد. با توجه به هر تفاوت مشخص بین μ_1 و μ_2 ، α ، احتمال خطای نوع دوم، یعنی β تابعی از حجم نمونه (n) است. هر قدر که حجم نمونه کوچک‌تر باشد، ارزش β بزرگ‌تر خواهد بود؛ بنابراین، با وجود تفاوت زیاد بین میانگین‌ها، اثبات آن، در نمونه‌های اندک دشوار است.

بنابراین دقت در میزان خطای نوع اول (α) در تحلیل‌های آماری منجر به نتیجه‌گیری دقیق‌تری خواهد شد. در بسیاری از پژوهش‌ها، بی‌توجهی به این نکته سبب افزایش میزان خطای نوع اول (تورم خطای نوع اول) شده است؛ در این پژوهش‌ها به جای تحلیل هم‌زمان چندین متغیر وابسته، یعنی استفاده از تحلیل واریانس چندمتغیری^۲ (MANOVA) از تحلیل‌های جداگانه‌ی آزمون t یا تحلیل واریانس معمولی (AVONA) استفاده شده است، در این حالت، دو اشکال اساسی پدیدار خواهد شد، اول، تورم بالای میزان خطای نوع I و II، و دوم، عدم دستیابی به همبستگی‌های درونی بین متغیرهای وابسته (هومن، ۱۳۸۰). میزان خطای نوع اول، وقتی که یک متغیر وابسته بررسی می‌شود، بنا به معمول (۰/۰۱ یا ۰/۰۵) است، اما با افزایش تعداد متغیرهای وابسته، این خطا افزایش می‌یابد. بنابراین، با ۲ متغیر وابسته و استفاده از آزمون t یا تحلیل واریانس (ANOVA) جداگانه برای هر یک از آنها، میزان خطا افزایش یافته و مقدار آن از (۰/۰۵ به ۰/۰۹۷۵) افزایش می‌یابد و این میزان با زیاد شدن تعداد متغیرهای وابسته همچنان رو به افزایش است؛ افزون بر این، خطای نوع دوم (β) نیز تحت تأثیر قرار گرفته و در نتیجه، مقدار توان آزمون کاهش می‌یابد (گایلس^۳، ۲۰۰۲، به نقل از فراهانی

1. Triola
2. multivariate analysis of variance (MANOVA)
3. Giles

و عریضی، ۱۳۸۴). بنابراین، هنگام تحلیل‌های آماری با چند متغیر وابسته، استفاده از تحلیل واریانس چند متغیری (MANOVA) این امکان را فراهم می‌آورد تا از تورم خطای نوع اول و کاهش توان آزمون جلوگیری و از سوی دیگر، همبستگی درونی متغیرهای وابسته مشخص شود. در تحلیل واریانس چندمتغیری، فرضیه‌ی اصلی مورد بررسی این است که آیا بردار میانگین‌ها در جامعه یکسان است؟ به عبارت دیگر، سنتروئید^۱ گروه‌ها بررسی می‌شود. برخلاف تحلیل واریانس که اسکالر (مقدار عددی ثابت) گروه‌ها مورد مقایسه قرار می‌گیرد، در اینجا، سنتروئید که در حقیقت، بردار میانگین‌های گروه‌هاست، مقایسه می‌شود (همان).

در تحلیل واریانس چند متغیری (MANOVA) چندین آماره آزمون برای بررسی معناداری محاسبه می‌گردند که عبارتند از: لامبدای ویلکز^۲، اثر پیلای^۳، اثر هتلینگ^۴ و بزرگ‌ترین ریشه‌ی اختصاصی رُی^۵. این چهار آماره را نرم‌افزارهای آماری و از جمله SPSS می‌توانند محاسبه کنند. اثر پیلایی محافظه‌کارانه‌ترین آزمون است و اگر این آماره معنادار شود، می‌توان به گزارش نتایج خود اطمینان داشت. بزرگ‌ترین ریشه‌ی رُی یک آماره‌ی سودار است که دارای حد اکثر سوگیری^۶ است، اما رایج‌ترین آنها لامبدای ویلکز است. در تحلیل واریانس چند متغیری (MANOVA)، از ANOVA به عنوان آزمون پسین (تعقیبی) برای مقایسه‌ی گروه‌ها از نظر متغیرهای وابسته در نرم‌افزار آماری SPSS به‌طور خودکار متعاقب MANOVA محاسبه و در برون‌داد رایانه‌ای مشخص می‌شود. اما باید دانست بسیاری از کارشناسان آمار با استفاده از آزمون‌های پسین تک‌متغیری که به دنبال MANOVA انجام می‌شود، موافق نیستند. یک دلیل قدرتمند برای این استدلال این است که MANOVA و ANOVA برای سؤالات پژوهشی متفاوتی طراحی شده‌اند (فراهانی و عریضی، ۱۳۸۴).

بنابراین، می‌توان گفت، هرگاه مقایسه‌ی جداگانه سبب تورم خطای نوع اول گردد، باید از تحلیل واریانس چند متغیری (MANOVA) استفاده نمود. با این حال، در این حالت جایگزین‌های آماری دیگری مانند رگرسیون چندمتغیری نیز کاربرد دارد. اما برای هر یک از آنها سؤالات خاصی مطرح است که پاسخ‌گویی به آنها لزوم استفاده از هر یک از

1. Centroid
3. Pillai's Trace
5. Roy's greatest characteristics root criterion

2. Wilk's Lambda
4. Hottelling Trace
6. Bias

این تحلیل‌ها را مشخص می‌کند. افزون بر این، تحلیل واریانس چندمتغیری پیش‌فرض‌هایی دارد که باید هنگام استفاده از آن مورد توجه قرار گیرد. از جمله‌ی آنها، عدم وجود هم‌خطی چندگانه^۱ بین متغیرهای وابسته، همگنی^۲ واریانس‌ها و طبیعی بودن توزیع است.

۲. فرض صفر و کاربرد نادرست آن

فرض‌های آماری، اغلب به شکل فرض صفر بیان می‌شوند، یعنی ادعایی که مبین فرض وجود عدم تفاوت است. این ادعا فرضیه‌ی صفر نامیده می‌شود. در حقیقت، تبیین تصادفی بودن تفاوت، به عنوان فرض صفر در نظر گرفته می‌شود. فرض صفر حکمی است در مورد اینکه هیچ رابطه واقعی بین متغیرها وجود ندارد و هر رابطه‌ی مشاهده شده، تنها تابعی از یک تصادف است (آری^۳، جیکوبز^۴ و رضویه، ۱۳۸۰).

آنچه اهمیت دارد، این است که به هر حال، فرض صفر اثبات نمی‌شود، بلکه وقتی گفته می‌شود که فرض صفر رد نشد یا پذیرفته شد، نشان می‌دهد که شواهد حاصل از نمونه‌ی مورد بررسی، آن‌قدر محکم نیست که با اطمینان بتوان فرض صفر را رد کرد و هیچ‌گاه به معنای قطعیت نیست (تریولا، ۱۳۷۴). در پاره‌ای از پژوهش‌ها فرض صفر برای احراز معادل بودن اولیه‌ی گروه‌ها استفاده شده است. آزمون فرض صفر یک روش متداول مورد استفاده در پژوهش‌های تجربی بوده است (لویت^۵، ۲۰۰۱). این استفاده از فرض صفر، کاربرد نادرستی است و به آن انتقادهای جدی وارد شده است (دنیس^۶، ۲۰۰۳). این مشکل در پژوهش‌های علوم رفتاری هرچند به ندرت اما دیده می‌شود (فراهانی و همکاران، ۱۳۸۶).

بنابراین، پذیرش فرض صفر حاکی از آن است که تفاوت و رابطه‌ی مشاهده شده بر اساس شواهد موجود حقیقی نیست و هرگز به این معنا نیست که این عدم تفاوت یا رابطه امری قابل اثبات و همیشگی است، زیرا آزمون فرض صفر چیزی جز به کارگیری احتمالات نیست. پس نمی‌توان به صرف تأیید فرض صفر در متغیرهای مورد نظر، لزوماً حکم به تساوی اولیه‌ی گروه‌ها در آن متغیرها داد. همان‌گونه که ردّ فرض صفر

1. Multicollinearity
3. Ary
5. Leavitt

2. homogeneity
4. Jacobs
6. Denis

به این معناست که تفاوت مشاهده شده را نمی‌توان به‌طور منطقی، بر اساس خطای نمونه‌گیری تبیین کرد و احتمالاً می‌توان آن را به متغیر آزمایشی مورد نظر نسبت داد و حکم به معناداری نتیجه کرد (فرگوسن و تاکانه، ۱۳۷۷).

۳. مقدار P و کاربرد نادرست آن

همان‌گونه که اشاره شد، در فرض آزمایشی، میزان خطای نوع اول (α) از قبل بر اساس یک قاعده‌ی قراردادی به صورت ۰/۰۵ یا ۰/۰۱ انتخاب می‌شود، اما می‌توان به جای این کار از روش دیگری برای آزمودن فرض صفر (H_0) استفاده کرد، به این ترتیب که با فرض درست بودن فرض صفر، احتمال مشاهده‌ی مقداری مساوی یا بزرگ‌تر از آماره‌ی به دست آمده از گروه نمونه محاسبه شود، این احتمال را مقدار P می‌نامند. بسته‌های نرم‌افزاری از جمله SPSS مقدار P هر آماره را به دست می‌آورند (هاراوی، ۱۳۷۹). اگر این احتمال کم‌تر از ۰/۰۵ باشد، فرض صفر رد و فرض مقابل (H_1) پذیرفته می‌شود، یعنی احتمال بروز چنین تفاوتی با وجود درست بودن فرض صفر، بسیار اندک خواهد بود.

پژوهشگران اغلب، مقدار P از قبل تعیین شده را ذکر نمی‌کنند و از آن برای تعیین قدرت نتایج، به طور نادرست استفاده می‌کنند (اسکاتز^۱، و همکاران، ۲۰۰۵). مثلاً از واژه‌های نامناسب (مانند نزدیک به معنادار، معناداری مرزی و تقریباً معنادار) استفاده می‌شود (دار و همکاران، ۱۹۹۴، دنیس، ۲۰۰۳). هر چند فرگوسن و تاکانه بر این باورند که نیازی به گزارش مقدار دقیق P نیست و می‌توان آن را به صورت $P < 0/05$ و $P < 0/01$ گزارش کرد، اما به این دیدگاه انتقاد شده است و باید در گزارش‌ها مقدار دقیق P گزارش شود (مثلاً، $P = 0/04$) اما اغلب مقدار P به صورت عدم تساوی کوچکتر یا بزرگتر مثلاً، $P < 0/05$ گزارش می‌شود (فراهانی و همکاران، ۱۳۸۶). این نکته را بسیاری از پژوهشگران نادیده می‌گیرند و در نتیجه، انجام پژوهش‌های فراتحلیل و نتیجه‌گیری دقیق آماری را دشوار می‌سازد.

۴. حجم اثر و عدم گزارش آن

حجم اثر به قدرت یا مقدار رابطه یا میزان فاصله با فرض صفر اشاره دارد (روزنتال و روزنوو، ۱۹۹۱). حجم اثر در واقع، تفاوت μ_1 و μ_2 را بر حسب انحراف معیار جامعه به ما می‌دهد؛

1. Sckatz

به عبارت دیگر، تعداد واحدهای انحراف معیاری را که دو میانگین جامعه با هم تفاوت دارند، نشان می‌دهد (سرمد، ۱۳۸۴). ای. پی. ای^۱ به پژوهشگران توصیه می‌کند که آنها نباید فقط به خوانندگان، اطلاعاتی در مورد معناداری آماری ارائه دهند، بلکه باید علاوه بر آن، اطلاعات کافی برای درک میزان اثر مشاهده شده یا رابطه فراهم آورند (ای. پی. ای، ۲۰۰۱)؛ زیرا حجم اثر نشان‌دهنده‌ی میزان یا درجه‌ی حضور پدیده در جامعه و با فرض صفر در ارتباط است و هرگاه فرضیه‌ی صفر تأیید نشود، یعنی مقدار حجم اثر در جامعه غیر صفر است و هر چه حجم اثر بزرگ‌تر باشد، درجه‌ی حضور پدیده بیشتر است (جعفری، ۱۳۷۸، به نقل از فراهانی و عریضی، ۱۳۸۴). از آنجایی که آزمون‌های معناداری در علوم رفتاری با درصد بالایی از خطای نوع دوم همراه است، به منظور کاهش خطای نوع دوم یا ارتقای توان آزمون، علاوه بر راهبردهایی که روش‌شناختی پژوهش در اختیار پژوهشگران قرار می‌دهد، به برآورد حجم اثر و استفاده از آن در تصمیم‌گیری در مورد رد یا قبول صفر پرداخته می‌شود (فراهانی و عریضی، ۱۳۸۴). اغلب، پژوهشگران بر مقادیر P برای نشان دادن قدرت اثر دستکاری آزمایشی تکیه دارند (کوهن^۲، ۱۹۶۵، ۱۹۹۰). روزنتال (۱۹۹۱) معتقد است که اگر حجم اثر محاسبه نشود، احتمال بروز خطای نوع اول افزایش می‌یابد. سطح معناداری و حجم اثر اگر چه در مورد رد یا قبول فرض صفر استفاده می‌شوند، هر کدام اطلاعات جداگانه‌ای می‌دهند. با استفاده از آزمون معناداری تعیین می‌گردد که آیا نتیجه‌ی خاصی به علت شانسی رخ داده یا خیر؛ در حالی که از طریق حجم اثر متوجه می‌شویم که متغیر مستقل تا چه اندازه بر متغیر وابسته اثر گذاشته است (فراهانی و عریضی، ۱۳۸۴).

این نکته در تبیین نتایج پژوهش‌های علوم رفتاری و تسهیل انجام دقیق فراتحلیل اهمیت ویژه دارد. چنانچه در پژوهش‌های مورد بررسی برای فراتحلیل، حجم اثر از سوی پژوهشگران گزارش نشده باشد، می‌توان از جدول خطوط راهنمای کوهن استفاده نمود (همان).

بحث و نتیجه‌گیری

هدف از مقاله‌ی حاضر، بررسی پاره‌ای از خطاهای آماری رایج در پژوهش‌های علوم رفتاری بود. توجه به خطاهای آماری، سبب افزایش قدرت روایی نتیجه‌گیری‌های آماری

1. APA
2. Cohen

می‌گردد و در نتیجه، استنباط دقیق‌تری از داده‌های خام فراهم می‌آورد؛ زیرا آمار کور است و این پژوهشگر است که با بینش خود به استفاده و نتیجه‌گیری درست می‌پردازد، به همین دلیل، پژوهش‌های مختلفی بر این خطاها برای بهبود نتیجه‌گیری‌های آماری متمرکز بوده‌اند. مثلاً در پژوهشی اسکاتز و همکاران (۲۰۰۵) کاربرد نادرست آزمون‌های آماری در پژوهش‌های روان‌شناسی و نوروسایکولوژی را مورد بررسی قرار دادند، با مروری بر پژوهش‌های چند دهه‌ی گذشته در حیطه‌ی نوروسایکولوژی می‌توان استفاده‌ی نادرست از آزمون فرض صفر در پژوهش‌های تجربی (دار، سرلین و اوامر، ۱۹۹۴، لویت، ۲۰۰۰)، گزارش نادرست P-value (کوهن، ۱۹۹۰، دنیس، ۲۰۰۲)، افزایش خطای نوع اول و نیز عدم گزارش و توجه به حجم اثر (روزنتال و روزنو) را ملاحظه نمود. در پژوهشی که فراهانی و همکاران (۱۳۸۶) به منظور بررسی خطاهای آماری مذکور در پژوهش‌های مرتبط با نوروسایکولوژی، در برخی از مجلات ایرانی انجام دادند، مشخص گردید که نود درصد مقالات، بدون گزارش حجم اثر، چهل درصد با افزایش خطای نوع اول، هشتاد درصد با گزارش غیر دقیق P-value و ده درصد با کاربرد نادرست فرض صفر انتشار یافته بودند. بنابراین توجه هر چه بیشتر به این خطاهای آماری که در پژوهش‌های رفتاری صورت می‌گیرد، اهمیت دارد، آرایه‌ی حجم اثر به افزایش روایی نتیجه‌گیری‌های آماری در پژوهش‌های علوم رفتاری می‌انجامد و به خصوص، زمانی که حجم نمونه اندک باشد، حجم اثر شاخص دقیق‌تری از نتایج دستکاری آزمایشی خواهد بود (کوهن، ۱۹۹۲). هم‌چنین گزارش حجم اثر امکان انجام پژوهش‌های فراتحلیل را در پژوهش‌های آتی فراهم می‌آورد. گزارش مقدار گردشده‌ی P-value به جای گزارش مقدار دقیق آن نتیجه‌گیری‌های آماری را نامعلوم کرده، دقت آن را کاهش می‌دهد و انجام پژوهش‌های فراتحلیل را دشوار می‌سازد. هم‌چنین استفاده‌ی نادرست از فرض صفر، به ویژه هنگامی که با تورم خطای نوع اول همراه شود، به تفاوت‌های نامعتبر در گروه‌ها منجر می‌شود. بنابراین، در پژوهش‌های علوم رفتاری، گزارش حجم اثر، مقدار دقیق P-value، کاربرد صحیح آزمون‌های آماری (عدم تورم خطای نوع اول) و استفاده از تحلیل کوواریانس برای کنترل تفاوت‌های اولیه به جای استفاده از فرض صفر برای احراز معادل بودن اولیه‌ی گروه‌ها، به دستیابی نتایج دقیق و افزایش روایی نتیجه‌گیری‌های آماری منجر خواهد شد.

فهرست منابع و مآخذ

۱. آری، دونالد، جیکوبز، لوسی چسر و رضویه، اصغر؛ روش تحقیق در تعلیم و تربیت، ترجمه وازگن سرکیسیان، مینونیکو و ایمان سعیدیان، تهران، انتشارات مرکز تحقیقات و مطالعات و بخش برنامه‌های صدا و سیمای جمهوری اسلامی ایران، ۱۳۸۰.
۲. تریولا، ماریو؛ آمار کاربردی، ترجمه محمدصادق تهرانیان و ابوالقاسم بزرگ‌نیا، مشهد، انتشارات جهاد دانشگاهی مشهد، ۱۳۷۴.
۳. دلاور، علی؛ احتمالات و آمار کاربردی در روان‌شناسی و علوم تربیتی، تهران، انتشارات رشد، ۱۳۷۶.
۴. سرمد، زهره؛ آمار استنباطی: گزیده‌ای از تحلیل‌های آماری تک‌متغیری، تهران، انتشارات سمت، ۱۳۸۴.
۵. فراهانی، حجت‌الله و عریضی، حمیدرضا؛ روش‌های پیشرفته پژوهش در علوم انسانی، اصفهان، انتشارات جهاد دانشگاهی، ۱۳۸۴.
۶. فراهانی، حجت‌الله، عریضی، حمیدرضا، رحیمی‌نژاد، عباس و قمرانی، امیر؛ بررسی اعتبار نتیجه‌گیری‌های آماری در پژوهش نوروسیکولوژی. مقاله‌ی ارائه شده در دومین کنگره نوروسیکولوژی، تهران، جهاد دانشگاهی دانشگاه تربیت معلم، خرداد ماه ۱۳۸۶.
۷. فرگوسن، جرج، ا. و تاکانه، یوشیو؛ تحلیل آماری در روان‌شناسی و علوم تربیتی، ترجمه علی دلاور و سیامک نقش‌بندی، تهران، نشر ارسباران، ۱۳۷۷.
۸. مولر، جی. اچ.، شوسلر، کی. اف. و کاستنر، اچ. ال؛ استدلال آماری در جامعه‌شناسی، ترجمه هوشنگ نایی، تهران، نشر نی، ۱۳۸۶.
۹. هاراوی، جان؛ درآمدی بر روش‌های آماری، ترجمه لینا ملکیان و برومندزاده، تهران، دفتر پژوهش‌های فرهنگی، ۱۳۷۹.
۱۰. هومن، حیدرعلی؛ تحلیل داده‌های چندمتغیری در پژوهش رفتاری، تهران، انتشارات پارسا، ۱۳۸۰.

11. American Psychological Association. *Publication Manual of the American psychological Association (5 th ed)*. Washington, D. C: American Psychological Association. Boston, MA: MC Graw Hill Inc, 2001.
12. Cohen, J. Some statistical Issues in Psychological research. Inc B. B. *Wolman hand book of clinical psychology*. Mearawtill: Network, 1965.
13. Cohen, J. Things I have learned (so far) *American psychologists* 45 (12), 1304-1312, 1990.
14. Dar, R., Serlin, R. C. & omer, H. Misuse of statistical test in three decades of psychotherapy Research, *Journal of Counselling and clinical psychology*, 52 (1), 75-82, 1994.
15. Denis, D. J. Alternative to null hypothesis significance testing. *Theory science*, 4 (1). 321-329, 2003.
16. Leavitt, F. Evaluating scientific research: separating fact from fiction. NJ: Prentice Hull, 2001.
17. Rosenthal, R. L. & Rosnow, R. Essentials of behavioral research: Methods and data. *Research Methods*. 5 (2). 125-132, 1991.
18. Schatz, P. Kristin, A. J., Jason, M. Ja son, R. M. Misuse of statistical test in Archivies of clinical Neuropsychology publications. Archives of clinical Neuropsychology Publications. *Archives of clinical Neuropsychological*. 20, 1053-1059, 2005.