ارائه الگوی استفاده بهینه از آزمون‌های آماری با کاربرد پلیسی

منیرالسادات صمدیان[1]

تاریخ وصول: ۸۸/۷/۱۲ تاریخ پذیرش: ۸۸/۱۲/۲۵

چکیده

قرن بیستم شاهد افزایش روزافزون کاربرد استدلال آماری در تمامی رشته‌های علوم و مخصوصاً در علوم انسانی و اجتماعی بوده است. لذا امروزه آشنایی با اصول و تکنیکهای پایه‌ای تحلیل آماری در اکثر دوره‌های علمی و حرفه‌ای اهمیت خاصی دارد.

ارتباط تنگاتنگ پژوهش‌های علمی با روش‌های آماری برای تأیید یا رد فرضیه‌ها همواره مسائل و مشکلاتی را از لحاظ متدولوژی برای پژوهشگران ایجاد کرده است. از این جهت ارائه یک الگو برای رفع این کمبودها ضروری به نظر می‌رسد.

در این مقاله می‌کوشیم تا پس از بیان آزمون‌های آماری (آزمون پارامتری (میانگین یک جمعیت، تفاوت بین میانگین‌ها، مقایسه زوج‌ها، آزمون F، همبستگی پیرسون) آزمون ناپارامتری (کای اسکوئر، علامت، میانه، من‌ویتنی، کروسکال والیس)) با ذکر مثال بعد از هر آزمون والیس در حوزه وظایف و مأموریت‌های پلیس چگونگی استفاده از این مدل را مشخص کنیم.

واژگان کلیدی: آزمون‌های آماری، پارامتری، ناپارامتری، میانگین، واریانس، همبستگی، کاربرد پلیسی.

۱- عضو هیأت علمی پژوهشی سازمان تحقیقات و مطالعات ناجا

مقدمه

امروزه آمار واژه‌ای برسر زبانهاست، معنای علمی این کلمه چیست؟ چه توصیه‌هایی را برای محقّقین و مدیران تحقیقاتی که همه روزه باید با آخرین تحولات و توسعه‌های علمی آشنا باشند، دارد؟

نیروی انتظامی ج. ا. ا به علت تعدّد مأموریت‌ها و وظائف، ارتباط مردمی و گستردگی فعالیت‌ها با مسائل و موضوعات متعددی روبروست لذا برای اجرای موضوعات متنوع تحقیقاتی از کلیه روش‌های رایج آماری استفاده می‌کند. با پیچیده‌تر شدن پژوهش‌های اجتماعی، انتظامی علم آمار نیز در مسیر تکاملی، روش‌های بیشتر و بیشتری را متناسب با موقعیتهای بخصوصی عرضه می‌کند. سرانجام به نقطه‌ای می‌رسیم که مدیر تحقیقاتی به آسانی نمی‌تواند با همه روش‌های آماری مناسب با کار خود آشنا باشد. شاید منصفانه نیز نباشد که انتظار داشته باشیم یک مدیر تحقیقاتی شایسته و متخصّص لزوماً یک آماردان شایسته نیز باشد. اما واقعیت این است که به همان اندازه که مدیر تحقیقاتی باید در تفکر و حرفه پژوهشی جدی باشد لازم است نسبت به تشخیص و درک ابزارهای پژوهشی استفاده شده در طرح‌های تحقیقاتی خود، از جمله فنون آماری که بدون تردید از تدابیر با ارزش پژوهش است، نیز تا حد ممکن علاقه‌مند و کوشا باشد، تا بتواند صحت و سقم روش‌ها و دستاوردهای پیشنهادی توسط مجری تحقیق را مشخص کند و کنترل لازم بر روند صحیح تحقیق به عمل آورد.

آمار نقش یک ابزار قدرتمند را در تمامی شاخه‌های علوم پایه، تجربی و انسانی به عهده دارد. از سوی دیگر نیاز به وقوف از باطن قضایا و لزوم پیشگویی و پیش‌بینی در مسائلی نظیر علوم اقتصادی، اجتماعی و امنیتی برای علم آمار جایگاهی خاص پدید آورده است. به‌علاوه علم آمار این ویژگی را دارد که بنا به ماهیتش، روش و میزان دقت خود را در برخورد با هر مسئله واقعی شرح دهد و با توجه به ملموس بودن قریب به اتفاق مسائل پیش رو، این اجبار را دارد که بنا به اقتضا به قویترین ابزارهای نظری، بویژه علوم ریاضی توسّل جوید. بنابراین الزاماً مباحث مطروح در آن پیچیدگیهای زیادی را طلب می‌کنند. در نتیجه آیا می‌توان پژوهشی را انجام داد ولی از کاربردهای آمار در پیچیدگیهای آن اجتناب کرد؟

مقصود اصلی از تألیف این مقاله نیل به هدف بالا و آشنایی کلی مدیران تحقیقاتی و محقّقین با آزمون‌های آماری با تأکید بر مثال‌های کاربردی در ناجاست.

ضرورت و بیان مسئله

۱- علوم انتظامی یکی ازدانش‌های بین‌رشته‌ای است لذا همانند رشته‌های مشابه از همه روش‌های تحقیق استفاده می‌کند.

۲- با نگرش به مفاد بند اول ضرورتاً در تحقیقات علوم انتظامی از کلیه روش‌ها و آزمون‌های آماری بهره‌برداری می‌شود.

۳- با نگرش به تخصّصی شدن موضوعات علمی و تهیه بسته‌های نزم‌افزاری متعدد آگاهی از همه آزمون‌های آماری و چگونگی استفاده از آن برای محققین ضرورت دارد تا بتوانند صحت و سقم روش‌ها و آزمایش‌های آماری پیشنهادی توسط کارشناسان آمار را مشخص کرده و کنترل لازم بر روند صحیح تحقیق به عمل آورند.

آزمون‌های آماری و کاربرد آن

آزمون پارامتری:

آزمونی آماری است مبتنی بر چند پیش فرض درباره پارامترهای جمعیتی که نمونه از آن انتخاب شده است. دو تا از مهمترین این پیش فرض‌ها عبارتند:

۱) مشاهدات باید از جمعیتی با توزیع نرمال باشد.

۲) متغیرها حداقل برحسب مقیاس فاصله‌ای اندازه‌گیری شوند. نتایج آزمون پارامتری فقط در صورتی معنادار است که پیش فرض‌ها صادق باشد.

آزمون پارامتری معناداری میانگین یک جمعیت

در این بخش فرضیه آزمون آماری را در مورد میانگین جمعیت تحت شرایط زیر بررسی خواهیم کرد:

$$H_0 : \mu = c$$
$$H_1 : \mu \neq c$$

۱- وقتی که نمونه‌برداری از جمعیت برخوردار از توزیع نرمال و واریانس معلوم صورت گیرد.

۲- وقتی که نمونه‌برداری از جمعیت برخوردار از توزیع نرمال و واریانس نامعلوم صورت گیرد.

۳- وقتی که نمونه‌برداری از جمعیتی برخوردار از توزیع غیرنرمال صورت گیرد.

گرچه نظریه مربوط به شرایط ۱و ۲ تنها به جمعیتهای برخوردار از توزیع نرمال بستگی دارد، در صورتی که جمعیت‌های مربوط تنها به طور تقریبی از توزیع نرمال برخوردار باشند، استفاده از آن معقول و معمول می‌باشد. انجام این مهم تا زمانی که انحراف از نرمال بودن زیاد نباشد رضایت‌بخش خواهد بود.

نمونه‌برداری از جمعیت‌های برخوردار از توزیع نرمال، وقتی که واریانس جمعیت معلوم باشد:

شاخص آماری آزمون: چون آزمون فرضیه را حول میانگین انجام می‌دهیم و جمعیت تحت مطالعه از توزیع نرمال با واریانس معلوم برخوردار است، شاخص آماری آزمون به صورت زیر خواهد بود:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

با آگاهی از توزیع نمونه‌برداری و توزیع نرمال می‌دانیم:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

در صورتی که H_0 درست باشد، از توزیع نرمال با میانگین صفر و واریانس یک برخوردار است. شاخص آماری آزمون ممکن است مقادیر مختلفی را در نمونه‌گیری‌های گوناگونی به خود اختصاص دهد.

اگر مقدار محاسبه شده شاخص آماری آزمون در ناحیه رد فرضیه قرار گیرد، H_0 را رد می‌کنیم و اگر این مقدار در ناحیه قبول فرضیه واقع شود H_0 را قبول می‌کنیم. در اینجا باید نواحی رد و قبول فرضیه را مشخص کنیم. اکنون این سوال را از خود می‌کنیم که چه مقدار از مقادیر شاخص آماری آزمون باعث رد شدن H_0 می‌گردد؟ اگر فرضیه صفر نادرست باشد، این نادرستی ممکن است از کمتر بودن یا بزرگتر بودن میانگین واقعی جمعیت از مقدار فرضی صورت گیرد. بنابراین مقادیر بسیار کوچک یا بسیار بزرگ شاخص آماری آزمون سبب رد شدن فرضیه صفر می‌گردند. همین مقادیر افراطی و تفریطی ناحیه رد آزمون فرضیه را نتیجه می‌دهد. مقدار ممکن از شاخص آماری آزمون چقدر باید کوچک یا بزرگ باشد تا مرز ناحیه رد فرضیه را مشخص سازد؟

پاسخ این سؤال به انتخاب سطح معنی‌دار آزمون، یعنی احتمال ارتکاب خطای نوع اول بستگی دارد. فرض کنید برای مثال اگر احتمال خطای نوع اول را $\alpha =0/05$ بگیریم. چون ناحیه رد فرضیه از دو قسمت تشکیل می‌شود، مقداری از α با مقادیر خیلی بزرگ و مقدار دیگر با مقادیر خیلی کوچک مربوط می‌شوند.

به نظر می‌رسد که مقدار α را بطور مساوی بین این دو قسمت تقسیم کنیم و $\alpha/2 =0/025$ را مقادیر خیلی کوچک و $\alpha/2 =0/025$ دیگر را با مقادیر خیلی بزرگ مربوط ساخته و در دنباله‌های بالایی و پایینی منحنی قرار دهیم.

نمونه‌برداری از جمعیت برخوردار از توزیع نرمال، اگر واریانس جمعیت نامعلوم باشد:

چگونگی تصمیم‌گیری رد یا قبول فرضیه همانند آزمون فوق می‌باشد.

شاخص آماری آزمون: چون واریانس جمعیت نامعلوم است، شاخص آماری آزمون برابر است با:

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

شاخص آماری آزمون از توزیع t با n-1 درجه آزادی، در صورتی که H_0 در واقع درست باشد، برخوردار است.

مثال:

برای بررسی میزان توانمندی افسران فراگیر، ابتدا مهارت‌های ۹ گانه (انتظامی، عملیاتی، اداری، حقوق و قضائی، اجتماعی، جسمانی، مدیریتی، بینش دینی و التزام عملی و بینش سیاسی) مورد اندازه‌گیری قرار گرفت و سپس میانگین کلی این مهارت‌ها به عنوان میزان توانمندی کلی افسران در نظر گرفته شد.

حال می‌توان فرض زیر را مورد آزمون قرار داد.

H_0: افسران فراگیر در اجرای وظایف محوله دارای توانمندی لازم نمی‌باشند.

H_1: افسران فراگیر در اجرای وظایف محوله دارای توانمندی لازم می‌باشند.

مقدار t محاسبه شده در آزمون با استفاده از نرم افزار spss مقدار ۹/۱۳ می‌باشد و با مقایسه با مقدار آن در جدول توزیع t می‌توان نتیجه گرفت در سطح معناداری ۹۵٪ فرض صفر رد می‌شود یعنی با ۹۵ درصد اطمینان می‌توان گفت، افسران فراگیر در اجرای وظایف خود، دارای توانمندی‌های لازم می‌باشند.

نمونه‌برداری از جمعیتی که از توزیع نرمال برخوردار نیست:

اگر نمونه‌ای که براساس آن فرضیه آماری را مورد آزمون قرار دهیم از جمعیت غیرنرمال انتخاب شود و نمونه انتخابی به حدکافی بزرگ باشد، شاخص آماری $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ استفاده می‌کنیم و چنانچه واریانس جمعیت معلوم نباشد واریانس نمونه را به عنوان برآورد آن به‌کار می بریم.

مثال:

مطالعه‌ای روی پرونده۶۴ مجرم در کلانتری یکی از مناطق انجام شده است. هر مجرم به طور متوسط در دو سال اخیر یک مرتبه به کلانتری مراجعه کرده است و انحراف معیار نمونه مذبور معادل ۱/۳ شده است.آیا می توان از این داده‌ها نتیجه گرفت که میانگین جمعیت تعداد دستگیری مجرمان در دو سال اخیر یک بوده است؟

همانگونه که بیان گردیده توزیع نمونه نامشخص ولی حجم نمونه بزرگ و انحراف معیار نمونه مشخص می‌باشد. پس همانگونه که اشاره شد از شاخص آماری $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ می‌توان استفاده کرد.

آزمون دو جمله‌ای

یکی از اساسی‌ترین موقعیت‌های پژوهشی هنگامی پیش می‌آید که با یک متغیر واحد که فقط دارای دو ارزش می‌باشد سروکار داشته باشیم. در این‌گونه موقعیت‌ها نسبتی از موارد موجود در جامعه که در داخل هر یک از دو مقوله قرار می‌گیرد مورد توجه ماست. مثلاً متغیر مورد مطالعه ممکن است یک متغیر طبقه‌ای با دو ارزش «قبول» و «مردود»، «حضور» و «عدم حضور» و مانند اینها باشد.

از آنجایی که متغیر موردنظر فقط دارای دو ارزش است هر یک از موارد جامعه باید در یکی از دو مقوله آن قرار گیرد. بنابراین اطلاع ما از نسبتی از جامعه که دارای ویژگی خاص می‌باشد p و برعکس آن $1 - p$ می‌باشد که با q نمایش می‌دهند.

مشخصه آماری این آزمون با استفاده از فرمول زیر محاسبه می‌گردد.

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{\pi}}$$

که در آن $\mu = np$ و $\sigma = \sqrt{npq}$

مثال:

تجربه نشان می‌دهد نیمی از نوجوانان بزهکار که از کانون اصلاح و تربیت پسران شهر تهران آزاد می‌شوند به سبب تکرار جرم در ظرف ۲ سال مجدداً دستگیر و روانه زندان می‌شوند. یک مددکار اجتماعی معتقد است با یک تکنیک گروه درمانی می‌تواند نسبت تکرار جرم را کاهش دهد. او تکنیک خود را در مورد ۳۶ آزمودنی (نوجوانان بزهکار) به‌کار می‌برد و بعد از ۲ سال با دقت کامل افراد ترخیص شده را زیر نظر می‌گیرد و ملاحظه می‌کند ۵ نفر از آنان قبل از پایان ۲ سال به علت تکرار جرم مجدّداً بازداشت شده‌اند.

پس فرض $H_0 : p = 0/5$ در برابر $H_1 : p < 0/5$ آزمون می‌گردد.

آزمون پارامتری معناداری تفاوت بین میانگین‌ها

پژوهشگران برای سنجش معنادار بودن تفاوت بین میانگین نمونه‌های انتخاب شده از جمعیت‌های مختلف از آزمون t استفاده می‌کنند. احتمال وجود تفاوت تصادفی بین میانگین نمونه‌ها در صورت درست بودن فرضیه صفر با استفاده از معادله (۱) محاسبه می‌شود. اگر بتوانیم فرض کنیم که واریانس دو جمعیت مساوی است خطای استاندارد را با استفاده از معادله (۲) محاسبه می‌کنیم. وقتی اندازه هر دو نمونه حداقل ۳۰ باشد، برای ارزیابی t از جدول منحنی نرمال استفاده می‌کنیم. اگر اندازه یکی از دو نمونه کمتر از ۳۰ باشد. برای ارزیابی t از جدول توزیع تی با $n_1 + n_2 - 2$ درجه آزادی استفاده می کنیم.

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} \quad (۱)$$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad (۲)$$

مثال:

برای بررسی میزان توانمندی افسران جزء و ارشد فراگیر، ابتدا مهارت‌های ۹ گانه (انتظامی، عملیاتی، اداری، حقوقی و قضائی، اجتماعی، جسمانی، مدیریتی، بینش دینی و التزام عملی و بینش سیاسی) مورد اندازه‌گیری قرار گرفت و سپس میانگین کلی این مهارت‌ها به عنوان میزان توانمندی کلی افسران در نظر گرفته شد.

حال می‌توان فرض زیر را مورد آزمون قرار داد.

H_0: افسران جز و ارشد فراگیر در اجرای وظایف دارای توانمندی یکسان می‌باشند.

H_1: افسران جز و ارشد فراگیر در اجرای وظایف دارای توانمندی یکسان نمی‌باشند.

مقدار t محاسبه شده در آزمون با استفاده از نرم افزار spss مقدار ۱۲/۱۳ می‌باشد و با مقایسه با مقدار آن در جدول توزیع t می‌توان نتیجه گرفت در سطح معناداری ۹۵ درصد فرض H_0 رد می‌شود یعنی با ۹۵ درصد اطمینان می‌توان گفت، افسران جزء و ارشد فراگیر در اجرای وظایف، دارای توانمندی‌های یکسان نمی‌باشند.

مقایسه زوج‌ها

در بحث گذشته که در رابطه با آزمون تفاوت بین میانگین‌های دو جمعیت صحبت کردیم این تصور وجود داشت که میانگین‌ها از هم مستقل هستند. روشی که برای ارزشیابی تأثیر نحوه معالجه با شیوه آزمایش خاص به‌کار می‌رود، روشی است که در آن از مشاهدات مربوط به نمونه‌های غیرمستقل استفاده می‌کنند. آزمون فرضیه‌ای که برمبنای این نوع داده‌ها قرار دارد به آزمون مقایسه زوج‌ها معروف است.

بسیار اتفاق می‌افتد که تفاوت واقعی بین دو جمعیت نسبت به متغیر موردنظر وجود ندارد، ولی وجود منابع خارجی پراکندگی، ممکن است سبب رد کردن فرضیه بی‌تفاوتی یا فرضیه موجود بشود. از طرف دیگر ممکن است تفاوت‌های واقعی با وجود عوامل خارجی پوشیده گردد. هدف از آزمون مقایسه زوج‌ها این است که با تشکیل زوج‌های شبیه به هم نسبت به متغیر مورد نظر حداکثر تعداد منابع خارجی پراکندگی را تا آنجا که امکان دارد از بین برد.

زوج مشاهدات از راههای گوناگون بدست می‌آید. موضوعات تحت مطالعه را قبل و بعد از اعمال رفتارهای خاص اندازه‌گیری می‌کنند.

فرضیه موجود که باید آزمون شود این است: $\mu_2 - \mu_1 \geq 0$ اگر $\mu_d = \mu_2 - \mu_1$ قرار دهیم، فرضیه ها را به صورت زیر می‌توان نوشت:

$$H_0 : \mu_d \geq 0$$
$$H_A : \mu_d < 0$$

با نگاهی به مفروضات مسئله ملاحظه می‌شود که شاخص آماری مناسب برای این آزمون عبارت است از:

$$t = \frac{\overline{d} - \mu_d}{s_{\overline{d}}}$$

در آن $s_{\overline{d}} = s_d / \sqrt{n}$

در این آزمون هرگز نباید نگران تساوی واریانس‌ها باشیم زیرا متغیّر ما تفاوت بین دو اندازه‌گیری را در همان شخص یا موضوع نشان می‌دهد. بنابراین تنها یک واریانس در محاسبات دخالت دارد.

اگر فرضیه صفر درست باشد، شاخص آماری آزمون از توزیع t با (n-1) درجه آزادی، برخوردار است.

مثال:

نمرات تست روانشناسی مجرمین را قبل و بعد از دوره بازپروری محاسبه کرده‌اند. دراین گونه موارد برای بررسی اینکه آیا اختلافی بعد و قبل از دوره بازپروری در مجرمین دیده شده، تفاوت بین زوج مشاهدات فردی را به عنوان متغیر مورد بررسی به‌کار می‌بریم.

آزمون معناداری برای همبستگی پیرسون (r): پژوهشگران برای آزمایش معناداری ضریب همبستگی می‌توانند از دو آزمون استفاده کنند. وقتی فرضیه صفر آن باشد که همبستگی در جمعیت صفر است $(\rho=0)$ برای تعیین معنادار بودن همبستگی می‌توان از توزیع t استفاده کرد که با معادله (۴) محاسبه می‌شود. پژوهشگران همچنین می‌توانند برای ارزیابی معناداری همبستگی از توزیع F استفاده کنند. نمرهF مبتنی است بر نسبت واریانس تبیین شده به واریانس تبیین نشده و با معادله (۵) محاسبه می‌شود.

$$t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (۴)$$

$$F=\frac{r^2}{1-r^2}(n-2) \qquad (۵)$$

مثال:

برای بررسی این فرضیه که آیا رضایت شغلی در توانمندسازی افسران تأثیر دارد و از آنجایی که توزیع جامعه نرمال و متغیر دارای مقیاس فاصله‌ای است می‌توان از آزمون ضریب همبستگی پیرسون استفاده کرد. بعد از ورود اطلاعات پرسشنامه، نرم‌افزار SPSS این مقدار را ۶۰۳٪ محاسبه کرده است. از آنجایی که ضریب همبستگی در فاصله منفی

یک تا یک مثبت تغییر می‌کند این عدد گویای این مطلب است که این دو متغیر از همبستگی نسبتاً خوبی برخوردار هستند و فرضیه $(\rho=0)$ رد می شود.

تجزیه و تحلیل واریانس[۱]

آزمون فرضیه صفر، یا فرضیه عدم تفاوت میانگین‌های جمعیت را ملاحظه کردیم. آزمون فرضیه صفر یا فرضیه عدم تفاوت بین میانگین چند جمعیت مختلف،برای پژوهشگر غیرمعمول نخواهد بود. محققی که برای نخستین بار با چنین مسئله‌ای روبرو می‌شود ممکن است بخواهد تمام زوج میانگین‌های نمونه را جداگانه به وسیله شاخص t آزمون کند. فرض کنید پنج جمعیت مورد مطالعه باشند. در آن صورت تعداد زوج میانگین‌های ممکن مربوط به نمونه‌های مختلف $10=\binom{5}{2}$ خواهد بود. چون حجم کار در اجرای این تعداد آزمون t مهم و اساسی است، اگر روش کاراتری را برای تجزیه واریانس بتوان جانشین آن کرد کار بسیار ارزشمندی خواهد بود. در عین حال، باید متوجه این نکته بود که احتمال زیادی وجود دارد که انجام این همه آزمون t ممکن است ما را به نتیجه‌گیری نادرستی سوق دهد.

فرض کنید پنج نمونه را از پنج جمعیت که دارای میانگین‌های مساوی هستند استخراج کنیم. چنان که دیدیم، اگر بنا باشد همه آزمون‌های ممکن را انجام دهیم. انجام ده آزمون لازم است. اگر سطح معنی‌دار $\alpha=$ ۰/۰۵ را برای هر آزمون در نظر بگیریم، احتمال آنکه فرضیه عدم تفاوت را در هر زمان رد نکنیم ۰/۹۵ خواهد بود. با استفاده از قاعده ضرب در حساب احتمالات، اگر آزمون‌ها از هم مستقل باشند، احتمال آن که فرضیه عدم تفاوت را در ده حالت رد نکنیم۰/۵۹۸۷= ^{۱۰}(۰/۹۵) می‌شود. از این رو، احتمال رد کردن حداقل یک فرضیه عدم تفاوت۰/۴۰۱۳= ۰/۵۹۸۷-۱ خواهد شد. چون از درستی فرضیه صفر در حالتی در این مثال توضیحی آگاه هستیم، رد کردن فرضیه صفر معادل با ارتکاب خطای نوع اول خواهد بود. از این رو، در درازمدت، در آزمون همه زوج

1. Analysis variance

میانگین‌های مربوط به پنج نمونه، مرتکب خطای نوع اول معادل ۴۰ درصد می‌شویم. در عمل مسئله از این هم مشکل‌تر خواهد شد؛ چون سه و بلکه بیشتر از سه تا از آزمونهای t از هم مستقل نیستند.

از این رو، واضح می‌گردد که روش دیگری برای آزمون تفاوت معنی‌دار بین چندین میانگین موردنیاز است. شیوه تجزیه و تحلیل واریانس چنین روش مناسبی را فراهم می‌آورد. به منظور توضیح چگونگی استفاده از روش تجزیه و تحلیل واریانس به مراحل زیر توجه کنید.

نسبت واریانس

آنچه باید انجام داد مقایسه دو برآورد σ^2 با یکدیگر است، که با محاسبه نسبت واریانس به صورت زیر میسر می‌شود:

$$V.R = \frac{\text{میانگین مربعات میان گروهی}}{\text{میانگین مربعات درون گروهی}}$$

چنانچه هر دو برآورد تقریباً با هم مساوی باشند، نسبت واریانس(V.R) به سمت یک میل خواهد کرد. نزدیکی نسبت با یک،گرایش به تأیید فرضیه برابری میانگین گروهی است. از سوی دیگر، اگر میانگین مربعات میان گروهی از میانگین مربعات درون‌گروهی بزرگتر باشند، V.R به مقدار قابل ملاحظه‌ای از یک بزرگتر خواهد بود. چنین مطلبی فرضیه برابری میانگینهای گروهی را دچار تردید خواهد کرد.

به خاطر تغییرات زیادی که در نمونه‌برداری وجود دارد، چنانچه حتی فرضیه صفر هم درست باشد، برابری میانگین مربعات درون گروهی و میان گروهی محتمل نیست. بنابراین، پیش از آنکه بتوانیم نتیجه‌گیری کنیم که این تفاوت مربوط به چیزی غیر از نوسانات نمونه‌برداری است، باید تصمیم‌گیری کرد که تفاوت مشاهده شده چقدر بزرگ باشد. به عبارت دیگر، مقدار V.R چقدر باید بزرگ باشد تا بتوان نتیجه گرفت که تفاوت مشاهده شده بین دو برآورد تنها در نتیجه شانس نیست؟

آزمون F

به منظور آن که به سؤال فوق پاسخ داده شود، آگاهی از توزیع نمونه‌برداری نسبت به دو واریانس نمونه مطلوب است. می‌دانیم که مقدار $(s_1^2/\sigma_1^2)(s_2^2/\sigma_2^2)$، در صورتی که واریانس‌های نمونه از نمونه‌های تصادفی و مستقل از هم و از دو جمعیت نرمال نمونه‌برداری شده باشند از توزیعی موسوم به توزیع F برخوردار است. در این مورد توزیع F برای تجزیه و تحلیل واریانس، توزیعی اساسی و بنیادی است.

می‌دانیم که در صورت برابری واریانس‌های دو جمعیت σ_1^2/σ_1^2 در عبارت فوق برابر یک شده و مقدار F به صورت s_1^2/s_2^2 باقی می‌ماند که خود نیز از توزیع F برخوردار است. توزیع F در واقع خانواده‌ای از توزیع‌ها می‌باشد و هر F به تعداد درجات آزادی مربوط به واریانس نمونه در صورت (درجه آزادی صورت) و تعداد درجات آزادی مربوط به واریانس نمونه در مخرج (درجه آزادی مخرج) بستگی دارد.

پس از آن که توزیع F مناسب تعیین شد، اندازه V.R مشاهده شده که سبب ردکردن فرضیه برابری واریانس‌های جمعیت می‌شود به سطح معنی‌دار انتخاب شده بستگی دارد. سطح معنی‌دار مقدار بحرانی F ، مقداری که ناحیه قبول را از ناحیه رد جدا می‌سازد، تعیین می‌کند.

همچنان که ملاحظه کردیم، V.R را در این گونه حالات با قرار دادن میانگین مربعات میان گروهی در صورت و میانگین مربعات درون گروهی در مخرج به گونه‌ای محاسبه کردیم که درجات آزادی صورت با $(k-1)$ تعداد گروهها منهای یک، و درجه آزادی مخرج با $\sum_{j=1}^{k}(n_j-1)=\sum_{j=1}^{k}n_j-k=N-k$ برابر خواهد بود.

محاسبات خلاصه شده به صورت جدول ۱ تهیه شده است، که جدول ANOVA نام دارد.

به منظور اخذ تصمیم باید مقدار محاسبه شدۀ V.R را با مقدار بحرانی F، با استفاده از جدول، مقایسه کرد.

جدول۱. تجزیه و تحلیل واریانس

منبع تغیرات	مجموع مربعات ss	درجهٔ آزادی d.F.	میانگین مربعات MS	نسبت واریانس V.R.
میان گروهی	$SS_{among} = \sum_{j=1}^{k} n_j (\bar{x}_{.j} - \bar{x}_{..})^2 = \sum_{j=1}^{k} \frac{T_{.j^2}}{n_j} - \frac{T_{..}^{2}}{N}$	$k-1$	$S_{among} = \frac{SS_{among}}{(k-1)}$	$\frac{MS_{among}}{MS_{within}}$ V.R=
درون گروهی	$SS_{within} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2 = \sum_{j=1}^{k} \sum_{i=1}^{nj} x_{ij}^{2} - \sum_{i=1}^{k} \frac{(T_{.j})^2}{n_j}$	$N-K$	$S_{within}=SS_{within}/(N-k)$	
	$SS_{total} = \sum_{j=1}^{k} \sum_{j=1}^{nj} (x_{ij} - \bar{x}_{..})^2 = \sum_{j=1}^{k} \sum_{i=1}^{nj} x_{ij}^{2} - \frac{T_{..}^{2}}{N}$	$N-1$		

مثال: در مطالعه‌ای توانمندسازی درجه داران، افسران جزء، افسران ارشد به تفکیک زن و مرد بعد از دوره آموزشی به شرح زیر بررسی گردیده است.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$
$$H_0 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6$$

- درجه داران مرد
- افسران جزء زن
- درجه داران زن
- افسران ارشد مرد
- افسران جزء مرد
- افسران ارشد زن

لازم به ذکر است که محاسبات فوق براساس داده‌ها انجام گرفته است. نسبت‌های F_0 حاصل در جدول زیر برای کد داده‌ها و داده‌های اصلی برابرند. در نتیجه برای آزمون فرضیه‌های بی‌اثر بودن عوامل، لزومی به برگردان این مقادیر به داده‌های اصلی نمی‌باشد. همان طور که در جدول بالا مشاهده شد اثرهای اعتیاد به موادمخدر و اثر جنسیت معنی‌دار می‌باشد. با توجه به داده‌ها کسانی که اعتیاد به موادمخدر داشته‌اند نسبت به کسانی که اعتیاد نداشته‌اند بیشتر مرتکب جرم شده‌اند. همچنین مردان نسبت به زنان بیشتر مرتکب جرم شده‌اند.

باید توجه داشت محاسبه داده‌ها با استفاده از نرم افزارهای آماری به راحتی امکان پذیر می‌باشد.

آزمون برای تفاوت‌های معنی‌دار بین هر یک از زوج میانگین‌ها

چنانچه تجزیه و تحلیل واریانس به رد کردن فرضیه صفر یا فرضیه عدم تفاوت میان میانگین‌های گروهی منجر شود، طبیعتاً این سؤال مطرح می‌شود که کدام زوج میانگین‌ها از هم‌دیگر متفاوت هستند. در واقع، پژوهشگر خواستار آن است که آزمون معنی‌دار روی هر یک از زوج میانگین رفتاری انجام دهد. برای مثال: در مثال فوق ممکن است خواستار آن باشیم که بدانیم پس از رد کردن فرضیه، $H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ کدام یک از ده فرضیه ممکن، باید رد شود. در عین حال، آزمایشگر در آزمون معنی‌دار تفاوت‌های بین هر زوج میانگین لازم است جانب احتیاط را نگه‌دارد و پیوسته مطمئن باشدکه شیوه او از اعتبار لازم برخوردار است. در استفاده از شیوه معتبر، نکته بحرانی سطح معنی‌دار است. گرچه در هر آزمون، مقدار α احتمال رد کردن فرضیه صفر را وقتی که واقعاً درست باشد کوچک می‌گیرند، اما احتمال رد کردن یک فرضیه صفر به واقع درست، در صورتی که زوج میانگین‌ها آزمون شود، چنان که خواهیم دید، از α بزرگتر خواهد شد. در طی سال‌های متمادی چندین شیوه برای انجام مقایسه‌های انفرادی پیشنهاد شده است. قدیمی‌ترین شیوه،و شاید یکی از مستعمل‌ترین آنها در گذشته،روش کمترین (کوچکترین) تفاوت معنی‌دار LSD است که فیشر برای نخستین بار آن را در سال ۱۹۳۵ در کتاب خود

تحت عنوان طرح آزمایش‌ها[1] بحث کرده است. شیوه LSD، که عبارت از آزمون دانشجویی t با استفاده از واریانس خطای ادغام شده می‌باشد، تنها زمانی معتبر خواهد بود که یا مقایسه‌ها به طور مستقل انجام گیرد یا قبل از تجزیه و تحلیل داده‌ها طرح‌ریزی گردد. تفاوت بین دو میانگین که از کمترین تفاوت معنی‌دار در سطح معنی‌دار معین بیشتر می‌شود در سطح معنی‌دار مورد استفاده در محاسبه LSD معنی‌دار خواهد بود. شیوه LSD معمولاً تنها وقتی به‌کار می‌رود که کل تجزیه و تحلیل واریانس به V.R معنی داری بیانجامد. نحوه استفاده از LSD را استیل و توری[2] با ذکر مثال آورده‌اند.

دانکن[3] سهم عظیمی در تحقیق موضوع مقایسه‌های چندگانه دارد و نتیجه کار او شیوه‌ای است که امروزه در وسعت زیادی به‌کار می‌رود و به آزمون دامنه چندگانه جدید دانکن موسوم است. بسط این آزمون به مواردی که اندازه‌های نمونه با هم مساوی نیستند را کرامر[4] بحث کرده است.

چنانچه هدف یک آزمایش مقایسه چندین رفتار با یک مشاهده باشد - و نه با یکدیگر- معمولاً شیوه‌ای موسوم به دانت[5] برای مقایسه مشاهده با هر یک از دیگر رفتارها دنبال می‌شود.

شیوه‌های دیگر مقایسه‌های چندگانه که مورد استفاده قرار می‌گیرند را توکی[6]، نیومن[7]، کویلز[8]، و شفه[9] طرح و پیشنهاد کرده‌اند. محاسن و عیوب شیوه‌های گوناگون توسط بانکرفت[10] و وینر[11] بحث شده است.

در مثال قبلی مقایسه دو به دو میانگین‌ها توسط هریک از آزمون‌های فوق قابل بررسی می‌باشد.

---

1. Design Of Experiments
2. Steel and Torrie
3. Duncan
4. Kramer
5. Dunnett
6. Tukey
7. Newman
8. Keuls
9. Scheffe
10. Bancroft
11. Winer

آزمون ناپارامتری

این آزمون نه مستلزم شرایط نرمال بودن است و نه مستلزم اندازه‌گیری فاصله‌ای، بیشتر آزمون‌های ناپارامتری نیز پیش‌فرض‌هایی دارند، ولی آن پیش‌فرضها از پیش‌فرض‌های آزمون‌های پارامتری ضعیف‌تر و کمتر هستند.

در عمل، نیازی نیست پژوهشگران به خود دردسر ساختن توزیع نمونه‌گیری را بدهند. در بسیاری موارد، توزیع‌های نمونه‌گیری را پژوهشگران پیشین ساخته و معلوم هستند. به علاوه، توزیع‌هایی وجود دارد که می‌توان از آنها به منزله تقریب‌هایی از توزیع‌های نمونه‌گیری خاص استفاده کرد. برای مثال توزیع نمونه‌گیری میانگین خیلی به توزیع منحنی نرمال نزدیک است؛ بنابراین پژوهشگران می‌توانند در آزمون فرضیه‌های مربوط به میانگین‌ها از توزیع منحنی نرمال استفاده کنند.

آزمون کای اسکوئر (χ^2)

«کای اسکویر» (یا کای دو)، آزمونی برای ارزیابی سطح معنی داربودن آماری و رابطه بین دومتغیر در سطح اسمی، ترتیبی و حتی فاصله ای، با استفاده از جدول بندی توافقی است. روش انجام این آزمون سه مرحله دارد:

(۱) توزیع فراوانی قابل انتظار برای هر خانه جدول بر مبنای این فرض که دو متغیر در جامعه، رابطه‌ای با یکدیگر ندارند، محاسبه می‌شود.

(۲) براساس تفاوت بین توزیع فراوانی قابل انتظار و فراوانی مشاهده شده در هر خانه جدول، آزمون «کای اسکوئر» محاسبه می‌شود. از آنجایی که فراوانی‌های قابل انتظار براساس فرض نبودن رابطه حساب شده‌اند، لذا هر چه تفاوت بین آنها و فراوانی‌های مشاهده شده بیشتر باشد، فاصله روابط مشاهده شده از فرضیه صفر بیشتر می‌شود و در نتیجه اعتماد به وجود رابطه بین دو متغیر افزایش می‌یابد.

(۳) مقدار «کای اسکوئر» محاسبه شده برای اطلاعات واقعی با جدول «کای اسکوئر» مقایسه و فرضیه صفر رد یا قبول می‌شود.

مثال:

یکی از کارشناسان مدیریت منابع انسانی یک سازمان دولتی تصور می‌کرد که در سطوح بالای مدیریت شایستگی وجود ندارد. او با توجه به کتاب اصول پیتر[1] متقاعد شده بود که عدم شایستگی در سطوح بالا بروز می‌کند و این عقیده را با یکی از همکارانش مطرح کرد و همکار وی درباره ماهیت ادعا با او به صحبت نشست.

برای اثبات این ادعا، او یک نمونه تصادفی ۴۰۰ نفری از پرونده‌های کارکنان را استخراج کرد. با توجه به تحصیلات و نمره آزمون ورودی افراد مذبور از لحاظ شایستگی، آنها را به سه سطح (پائین، متوسط، بالا) و با رعایت رتبه اداری و شرح شغل رسمی و از نظر مقام سازمانی به سه سطح (پائین، متوسط، بالا) تقسیم کرد.

اطلاعات مربوط به این دو متغیر در جدول ۲ مندرج است.

کارشناس مذبور می‌خواست بداند آیا می‌توان رابطه‌ای را که ممکن است بین شایستگی و مقام سازمانی در نمونه وجود دارد به تمام کارکنان سازمان تعمیم داد یا خیر. بر این اساس تصمیم گرفت که آزمون «کای اسکوئر» را انجام دهد.

۱- مفروضات وی عبارتند:

سطح اندازه گیری: مقیاس ترتیبی

فرضیه: H_0: بین شایستگی و مقام سازمانی رابطه‌ای وجود ندارد.

H_1: بین شایستگی و احراز مقام سازمانی رابطه مثبت برقراراست.

۲- باید دقت کرد که توزیع‌ها برحسب درجه آزادی تفاوت دارند. (۱- تعداد سطوح عامل اول) (۱- تعداد سطوح عامل دوم) مقدار درجه آزادی را مشخص می‌کند.

۳- سطح معنی دار بودن ۰/۰۵ و درجه آزادی ۴ ((۱-۳)(۱-۳))را انتخاب می‌کنیم.

۴- «کای اسکوئر» به شرح زیر محاسبه می‌شود:

الف- محاسبه فراوانی‌های قابل انتظار در هر خانه جدول بر مبنای فرضیه منفی که در جامعه بین شایستگی و مقام سازمانی رابطه‌ای وجود ندارد (باید دقت کرد که اگر جدول بر حسب درصد باشد، قبل از محاسبه باید فراوانی‌های قابل انتظار به ارقام خام تبدیل شوند).

1. The Peter Principle

جدول ۲. رابطه بین شایستگی و مقام سازمانی در هرم سلسله مراتب

مقام سازمانی	شایستگی			جمع
	پایین	متوسط	بالا	
پایین	۱۱۳	۶۰	۲۷	۲۰۰
متوسط	۳۱	۹۱	۳۸	۱۶۰
بالا	۸	۸	۲۴	۴۰
جمع	۱۵۲	۱۵۹	۸۹	۴۰۰

اگر دو متغیر با یکدیگر مربوط نباشند، انتظار می‌رود که توزیع مقام سازمانی در هر سطح از شایستگی، برابر باشد. در این حالت شایستگی، تأثیری روی مقام سازمانی ندارد و دو متغیر کاملاً بی ارتباط هستند.

با توجه به جدول ۲ ملاحظه می شود که از ۴۰۰ نفر نمونه، تعداد ۲۰۰ نفر یا ۵۰ درصد از لحاظ مقام سازمانی، در سطح پایین، ۱۶۰ نفر یا ۴۰ درصد در سطح متوسط و بقیه که ۱۰ درصد را تشکیل می‌دهند در سطح بالا رتبه بندی شده‌اند. به فرض اینکه فرضیه صفر مبتنی بر نبودن رابطه بین شایستگی و مقام سازمانی درست باشد، انتظار می‌رود که توزیع یکسان مقامات سازمانی در هر طبقه از شایستگی وجود داشته باشد. به عنوان مثال: از ۱۵۲ نفر کارمند در سطح پایین از لحاظ شایستگی، انتظار می‌رود (با توجه به ستون آخر) که ۵۰ درصد یا ۷۶ نفر درمقام سازمانی پایین (۷۶= ۱۵۲× ۰/۵۰)، ۴۰ درصد، یا ۶۰/۸ در مقام سازمانی متوسط (۶۰/۸= ۱۵۲× ۰/۴۰) و ۱۰ درصد یا ۱۵/۲ در مقام سازمانی بالا (۱۵/۲= ۱۵۲× ۰/۱۰) قرار گیرند. به همین نحو، فراوانی‌های قابل انتظار سایر طبقات مشخص می‌شوند. جدول ۳ محاسبات مشروح را نشان می‌دهد.

ب- مقدار کای اسکوئر محاسبه می‌شود. مقدارکای اسکوئر به وسیله:

(۱) محاسبه تفاوت بین فراوانی‌های قابل انتظار و مشاهده شده برای هر خانه جدول

(۲) مجذور کردن این تفاوت‌ها

(۳) تقسیم نتیجه حاصل به فراوانی قابل انتظار

(۴) جمع تمام ارقام حاصل در آخرین ستون

بدست می‌آید. به عنوان مثال: در خانه ۱ جدول (مربوط به شایستگی پایین و مقام سازمانی پایین) فراوانی مشاهده شده ۱۱۳ است و با ۷۶ (فراوانی قابل انتظار) مقایسه می‌شود. بنابراین ۱۸/۰۱۲= ۷۶: ۲(۷۶-۱۱۳) که در ستون آخر ردیف ۱ ملاحظه می‌شود، با سایر ارقام بعدی همان ستون که به همین ترتیب محاسبه می شوند جمع می‌شود و رقم ۸۹/۲۰ همان کای اسکوئر است.

ج- «کای اسکوئر» بدست آمده با «کای اسکوئر» مندرج در جدول توزیع کای اسکوئر مقایسه می‌شود. به منظور یافتن کای اسکوئر در جدول، دو چیز مورد نیاز است: ۱- درجه آزادی مربوط به جدول و ۲- سطح معنی دار بودن آماری، برای بدست آوردن درجه آزادی، از تعداد ردیف‌ها و ستون‌های جدول، عدد ۱ را کسر کرده و در هم ضرب می‌کنیم. در مثال فعلی چون تعداد ردیف‌ها وستون‌ها برابر ۳ است، لذا درجه آزادی برابر است با ۴= ۲×۲=(۱-۳)×(۱-۳).

۵- همانطور که جدول توزیع χ^2 نشان می دهد، با درجه آزادی ۴ و سطح معنی دار بودن آماری ۰/۰۵«کای اسکوئر» برابر ۹/۴۹ است. چون در این مثال، مقدار «کای اسکوئر» ۸۹/۲۰ خیلی زیادتر از مقدار احتمالی در جدول است (۹/۴۹<۸۹/۲۰) لذا نتیجه می گیریم که در جامعه بین دو متغیر (یعنی بین شایستگی و مقام سازمانی) رابطه وجود دارد.

لازم به توضیح است که فرمول محاسبه «کای اسکوئر» عبارت است از:

$$\mathrm{X}^2 = \sum \frac{(fo - fe)2}{fe}$$

که در آن: fo = فراوانی‌های مشاهده شده، fe = فراوانی‌های قابل انتظار می‌باشد.

برای محاسبه فراوانی‌های قابل انتظار در هرستون مجموع ردیف در مجموع ستون ضرب گردیده و بر تعداد تقسیم می نمائیم.

در مثال مذبور برای خانه ۱ جدول محاسبه بدین قرار است:

$$\frac{۱۵۲ \times ۲۰۰}{۴۰۰} = ۷۶$$

که در جدول ۳ نیز همین رقم ملاحظه می شود.

برای سایر خانه‌های جدول نحوه عمل به همین ترتیب است (۳).

جدول ۳. محاسبه فراوانی قابل انتظار و کای اسکوئر

خانه‌های جدول: شایستگی	خانه‌های جدول: مقام سازمانی	فراوانی مشاهده شده	فراوانی قابل انتظار	$\frac{(fo-fe)2}{fe}$
پایین	پایین	۱۱۳	٪۵۰×۱۵۲=۷۶	۱۸/۰۱
پایین	متوسط	۳۱	٪۴۰×۱۵۲=۶۰/۸	۱۴/۶۱
پایین	بالا	۸	٪۱۰×۱۵۲ =۱۵/۲	۳/۴۱
متوسط	پایین	۶۰	٪۵۰×۱۵۹=۷۹/۵	۴/۷۸
متوسط	متوسط	۹۱	٪۴۰×۱۵۹=۶۳/۶	۱۱/۸۰
متوسط	بالا	۸	٪۱۰×۱۵۹=۱۵/۹	۳/۳۹
بالا	پایین	۲۷	٪۵۰×۸۹=۴۴/۵	۶/۸۸
بالا	متوسط	۳۷	٪۴۰×۸۹=۳۵/۶	۰/۱۶
بالا	بالا	۲۴	٪۱۰×۸۹=۸/۹	۲۵/۶۲
جمع		۴۰۰	۴۰۰	۸۹/۲۰

محدودیت‌های کای اسکوئر

مثالی که ارایه شد، محدودیت‌های «کای اسکوئر» را به خوبی نشان می‌دهد. مراحل انجام دادن آزمون به نتیجه‌ای منجر شد که بین شایستگی و مقام سازمانی در جامعه رابطه وجود دارد. به هر حال، محقق در این مثال فرض کرده بود که رابطه برعکس است یعنی هر چه شایستگی کمتر باشد، مقام سازمانی بالاتر است. در حقیقت، وقتی که جدول‌بندی این دو متغیر از نمونه کارکنان (جدول ۲) به درصد تبدیل شود، رابطه بین شایستگی و مقام سازمانی مثبت است. درصدهای مربوط که در جدول۴ مندرج است نشان می‌دهد که با افزایش شایستگی، مقام سازمانی نیز در سلسله مراتب سازمانی بالا می‌رود. بنابراین در حالی که وجود رابطه می‌تواند در جامعه استنباط شود، اما رابطه در جهت معکوس فرضیه محقق است. بنابراین اندیشه او درست نیست.

نکته مهمی که این مثال نشان می‌دهد این نیست که آزمون «کای اسکوئر»منتج به اطلاعات غلط می‌شود، بلکه اطلاعات محدودی را نتیجه می‌دهد. آزمون صرفاً وجود یا فقدان رابطه را منعکس می‌سازد و در برابر ماهیت و در جهت رابطه که عملاً در جدول یافت می‌شود، حساسیت ندارد. محققان غالباً نتیجه می‌گیرند که مقدار «کای اسکوئر» حاصل از محاسبات منعکس کننده وجود رابطه در جامعه (به طریقی که فرض شده است) می‌باشد. ممکن است چنین باشد و احتمال دارد که چنین نباشد. تجزیه و تحلیل‌های آماری مکمل دیگری مانند درصدگیری و محاسبه مقیاس‌های وابستگی برای پاسخگویی به مسأله مذبور ضروری است.

جدول ۴. توزیع درصد

مقام سازمانی	شایستگی		
	پایین	متوسط	بالا
پایین	٪۷۴	٪۳۸	٪۳۰
متوسط	٪۲۱	٪۵۷	٪۴۳
بالا	٪۵	٪۵	٪۲۷
جمع	٪۱۰۰ n=۱۵۲	٪۱۰۰ n=۱۵۹	٪۱۰۰ n=۸۹

محدودیت‌های دیگر آزمون «کای اسکوئر» مربوط به کاربرد آن به عنوان آزمون معنی دار بودن آماری است. اولاً، آزمون «کای اسکوئر» مستلزم نمونه‌گیری از جامعه است (نمونه‌گیری تصادفی ساده) که بعضی از اوقات نمی‌تواند قانع کننده باشد. ثانیاً مقدار «کای اسکوئر» محاسبه شده از جدول‌بندی با اندازه نمونه متورم می‌شود. در نتیجه در نمونه‌های بزرگ وجود یک رابطه ضعیف، معمولاً از لحاظ آماری معنی‌دار یافت می‌شود. لذا آزمون نمی‌تواند تفاوت‌ها را نشان دهد.

بالاخره، با وجود اینکه آزمون «کای اسکوئر» اغلب به عنوان مقیاس رابطه قوی مورد سوء تعبیر قرار می‌گیرد، اهمیت روابط تجربی را ارزیابی نمی‌کند. این آزمون اطلاعاتی را فراهم می‌سازد که فقط احتمال وجود رابطه در جامعه را منعکس می‌سازد.

اطلاعات مذبور ارزشمند است، لکن موضوع اندازه رابطه را نادیده می‌گیرد. از این‌رو، توصیه می‌شود که آزمون کای اسکوئر با سایر روش‌های آماری، به خصوص آنهایی که مقدار قوت رابطه را نشان می دهد، به کار برده شود.

در پایان ذکر چند نکته لازم به نظر می‌رسد:

همان‌طور که ذکر شد، در آزمون معنی‌دار بودن آماری، ما با خطای نوع اول (رد کردن فرضیه‌ای که در واقع درست است) و خطای نوع دوم (رد نکردن فرضیه‌ای که عملاً غلط است) مواجه هستیم. چون غالباً محقّقان گرایش دارند که فرضیه صفر را رد کنند و در نتیجه فرضیه تحقیق مورد قبول واقع شود، لذا ممکن است ریسک خطای نوع دوم وجود داشته باشد. بنابراین، همیشه باید درباره انتخاب ناحیه بحرانی[1] جنبه‌های احتیاط را رعایت کرد تا رد کردن فرضیه صفر محدودتر شود. همان‌طور که ذکر شد، سطوح معنی‌دار بودن آماری که معمولاً در تحقیقات اجتماعی به کار می‌روند، ۰/۰۵ و ۰/۰۲ یا ۰/۰۱ هستند. باید متوجه بود که این سطوح جنبه مطلق ندارند. با وجود اینکه یک محقق معمولاً در استفاده از این سطوح احتیاط می‌کند، اگر مایل نباشد که فرضیه صفر را رد کند، شاید سطح ۰/۱۰ یا ۰/۲۰ را انتخاب کند و خطای نوع دوم را کاهش دهد.

همچنین، معنی‌دار بودن آماری[2] را نباید با معنی‌دار بودن عملی[3] اشتباه کرد. معنی‌دار بودن آماری فقط به ما می‌گوید که اگر تفاوت‌هایی در جامعه نباشد، تفاوت‌هایی در نمونه نمی‌تواند خیلی به طور تصادفی اتفاق افتد. معنی‌دار بودن آماری درباره اهمیت و درجه قوت این تفاوت‌ها چیزی نمی‌گوید. عاملی که تفاوت‌هایی را در یک نمونه کوچک به وجود می‌آورد، از عاملی که تفاوت‌های ناچیزی را در یک نمونه بزرگ ایجاد می‌کند (و هر دو از نظر آماری معنی‌دار نیز هستند) برای محققان ارزش خیلی زیادتری دارد.

چنانچه در یک آزمون، سطح معنی‌دار بودن ۰/۰۰۱ و در آزمون دیگر سطح معنی‌دار بودن۰/۰۵ باشد، ما تقریباً از وجود رابطه در حالت اول مطمئن هستیم، اما در مورد دوم زیاد اطمینان نداریم. در هر حال، باید به خاطر داشت که سطح معنی‌دار بودن حاصل، بستگی به

1- Critical Reaion
2- Statistical significance
3- Pracitical Significance

اندازه نمونه دارد. اگر نمونه‌ها خیلی بزرگ باشند استقرار سطح معنی‌دار بودن ساده است؛ حتی اگر رابطه خیلی ضعیف باشد. برای نمونه‌هایی بزرگ، یک سؤال مهم این است که به فرض اینکه رابطه وجود داشته باشد، این رابطه تا چه حد قوی است؟

برای روشن شدن بحث فوق، به ذکر مثالی از «کای اسکوئر» (یا کی دو) مبادرت می‌شود. دو جدول زیر (b و a) را در نظر بگیرید.

(a

۳۰	۲۰	۵۰
۲۰	۳۰	۵۰
۵۰	۵۰	۱۰۰

(b)

۶۰	۴۰	۱۰۰
۴۰	۶۰	۱۰۰
۱۰۰	۱۰۰	۲۰۰

«کای اسکوئر» برای جدول a برابر ۴ و برای جدول b برابر ۸ است. یعنی اگر نسبت‌های خانه‌های جدول تغییر نکند، مقدار «کای اسکوئر» با افزایش تعداد موارد زیاد می‌شود. اگر تعداد موارد دو یا سه برابر شود، مقدار «کای اسکوئر» نیز دو یا سه برابر می‌شود. لذا در نمونه‌های کوچک، آزمون‌های معنی‌دار بودن خیلی مهمتر است. سطح معنی‌دار بودن به دو عامل بستگی دارد: ا- درجه قوت یا استحکام رابطه ۲- اندازه نمونه. معنی‌دار بودن می‌تواند با یک رابطه خیلی قوی و نمونه‌های خیلی کوچک، یا با یک رابطه خیلی ضعیف و نمونه‌های خیلی بزرگ به دست آید.

آزمون علامت

آزمون بسیار معمول غیرپارامتری که به مفروضات آزمون t بستگی ندارد، یا اندازه‌گیری آن مقیاس ترتیبی است، آزمون علامت نام دارد. آزمون علامت بر میانه به جای میانگین به عنوان شاخص تمایل مرکزی یا محل تمرکز تکیه دارد. در توزیع‌های متقارن میانه و میانگین با هم برابر خواهند بود. تنها فرض آزمون فوق متصل بودن متغیر مورد مطالعه است.

نام آزمون علامت از این اصل گرفته می‌شود که علامت‌های مثبت و منفی، به جای مقادیر عددی داده‌های خام در محاسبات مورد استفاده قرار می‌گیرند. آزمون علامت را نخست با ارائه مثالی در یک نمونه شرح می‌دهیم.

مثال:

یک گروه شامل ۱۰نفر از افسران ارشد را در نظر گرفته و طی آزمونی سطح رضایتمندی آنان از محیط ستادی نسبت به اجرایی اندازه گیری می شود حال با استفاده از نمرات محاسبه شده فرضیه صفر را که میانه جمعیت مورد نمونه برداری برابر ۵ می‌باشد را آزمون می کنیم.

نخستین گام در آزمون فرضیه بررسی داده‌ها به منظور تعیین نمرات بالا و پایین میانه نظری است. اگر علامت + برای نمرات بالا میانه و علامت − برای نمرات زیرمیانه در نظر گرفته شود

نفرات	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
نمره رضایتمندی	۴	۵	۸	۸	۹	۶	۱۰	۷	۶	۶
علامت نمره براساس میانه	−	۰	+	+	+	+	+	+	+	+

چنانچه فرضیه صفر درست باشد، یعنی، اگر میانه به واقع برابر ۵ باشد، انتظارمان این است که تعداد نمرات بالا و پایین عدد ۵ تقریباً با هم برابر باشند. این طریق استدلال راه دیگری را برای ارائه فرضیه صفر پیشنهاد می کند. یعنی، احتمال علامت مثبت با احتمال علامت منفی برابر و مساوی ۵ خواهد بود. فرضیه صفر را به صورت نمادی زیر می توان نشان داد:

$$H_0 : p(+) = p(-) = 0.5$$

به عبارت دیگر، اگر H_0 درست باشد، انتظار داریم که تقریباً همان تعداد علامت‌های مثبت با همان تعداد علامت‌های منفی در جدول برابر باشد.

نگاهی به جدول فوق فزونی مثبت ها را بر منفی ها نشان می‌دهد. به طور دقیق، هشت مثبت، یک منفی، و یک صفر را مشاهده می‌کنیم. نمره صفر دقیقاً میانه داده‌ها را نمایش می‌دهد. روش معمول این است که از تجزیه و تحلیل، صفرها را حذف کرده و در نتیجه n، اندازه نمونه را مطابق آن کاهش دهیم. چنانچه شیوه فوق را در مورد داده‌های مورد بحث اعمال کنیم تعداد مشاهدات به ۹ کاهش می‌یابد که از آن میان ۸ تا مثبت و یک منفی خواهیم داشت.

چون تعداد علامت‌های مثبت و منفی یکسان نیست، از ناموزونی توزیع مثبت ها و مشکوک شدن به فرضیه صفر تعجب نخواهیم کرد. به بیان دیگر، در صورتی که فرضیه صفر درست باشد، اگر تعداد کم علامت‌های منفی بر حسب تصادف رخ داده باشد؛ یا عاملی غیر از تصادف (فرضیه صفر نادرست باشد) سبب این امر گردد موجب شگفتی نخواهد بود. اگر k تعداد علامت‌های منفی را نشان دهد، توزیع نمونه‌برداری k، در صورتی که فرضیه صفر درست باشد، توزیعی دوجمله‌ای با پارامتر $P=0.5$ خواهد بود.

احتمال مشاهده x یا کمتر از آن علامت منفی در صورتی که نمونه‌ای n تایی با پارامتر p ارائه شده باشد با محاسبه عبارت زیر ممکن می‌شود:

$$P(k \leq x \vdots n, p) = \sum_{k=0}^{x} \binom{n}{k} p^k q^{n-k}$$

در مثال مورد بحث خواهیم داشت:

$$\binom{9}{0}(0.5)^0(0.5)^{9-0} + \binom{9}{1}(0.5)^1(0.5)^{9-1} = 0.0195$$

آزمون میانه

یکی دیگر از آزمون‌های ناپارامتری آزمون میانه است. این آزمون تقریباً همتای پارامتری آزمون‌های t و z است. یعنی وقتی که دو گروه نمونه از میان دو جامعه مستقل با توزیع‌های یکسان انتخاب شده‌اند می‌توان آن را به کار برد. البته، به شرطی که مقیاس اندازه‌گیری حداقل ترتیبی باشد طبیعی است این موضوع وقتی مشکل‌ساز می‌شود که نمره تکراری

درست در میانه مشترک قرار گیرد. زیرا سایر نمره‌های همرتبه اثری در خود مشخصه آزمون ندارد. در این صورت، باید نمره‌های همرتبه با میانه مشترک داخل هر گروه را به گونه‌ای به مقوله‌ها اختصاص داد که تعداد نمره‌های بالاتر و پایین‌تر از میانه را هر چه ممکن است با هم نزدیک‌تر سازد. و تفاوت آنها را کاهش دهد. در این آزمون لزومی ندارد که حتماً حجم گروههای نمونه با یکدیگر برابر باشد. همچنین این آزمون را می‌توان برای مقایسه توزیع های چندجامعه نیز به کار برد و حتی می‌توان کاربرد آن را به مواردی که گروهها با یکدیگر وابسته‌اند (مانند زوج‌های جور شده) گسترش داد.

برای انجام آزمون، داده‌های نمونه را روی هم می‌ریزیم و میانه مشترک را پیدا می‌کنیم. سپس گروههای نمونه را از هم جدا و معلوم می‌کنیم که در هر یک از گروههای مطالعه شده چه تعدادی از مشاهده‌ها بالاتر و چه تعدادی پایین‌تر از میانه مشترک است. آنگاه فراوانی‌های بالاتر و پایین‌تر از میانه مشترک را برای گروههای مختلف در یک جدول دوبعدی تنظیم می‌کنیم و با استفاده از آزمون x^2 مشخص می‌کنیم که آیا جامعه‌ها با میانه یکسان انتخاب شده‌اند یا نه؟

درجه آزادی در اینجا برابر است با d.f=k-1 که در آن k تعداد گروههاست.

نام گروه وضعیت	۱	۲	۳	K
بالاتر از میانه مشترک ($\rangle md$)	n_{11}	n_{12}	n_{13}	n_{1k}
پایین‌تر یا مساوی میانه مشترک($\leq md$)	n_{21}	n_{22}	n_{23}	n_{2k}

مثال: فرض کنید می‌خواهیم بدانیم که آیا نمره کارایی افسران ارشد بیش از افسران جزء است؟ بدین منظور ۱۴ افسر ارشد و ۱۸ افسر جزء را به طور تصادفی از میان کلیه کارکنان اداره‌ای انتخاب و یک آزمون کارائی در هر دو گروه اجرا کرده‌ایم.

داده‌های حاصل از اجرای این آزمایش از این قرار است.

نمرات افسران جزء	۴۱-۴۷-۵۵-۴۶-۶۵-۶۶-۴۷-۴۵-۴۰-۶۶-۶۱-۴۶-۵۳-۶۸-۵۸-۴۵-۴۳-۴۷
نمرات افسران ارشد	۶۲-۶۳-۴۹-۷۰-۵۴-۵۱-۶۲-۵۲-۴۰-۵۷-۷۰-۷۲-۶۷-۴۲

این دو گروه را با هم ترکیب و آنها را مرتب می‌کنیم و سپس میانه آنها را به دست می‌آوریم

$$\frac{N}{2}+1=17 \quad ، \quad \frac{N}{2}=16 \quad ، \quad N=18+14=32$$

در نتیجه مقدار میانه مشترک عبارت است از: $Md=\frac{53+54}{2}=53/5$

حال دو گروه نمونه را از هم جدا و معلوم می‌کنیم که در هر یک از گروههای نمونه چه تعدادی از مشاهده‌ها بالاتر و چه تعدادی پایین‌تر از میانه مشترک است.

اکنون فراوانی‌های بالاتر و پایین تر از میانه مشترک را برای هر دو گروه نمونه در یک جدول ۲×۲ تنظیم می‌کنیم. اگر گروههای نمونه از جامعه‌هایی با توزیع یکسان یعنی جامعه‌هایی با میانه یکسان انتخاب شده باشند، انتظار داریم که در هر گروه نیمی از نمره‌ها پایین تر از میانه مشترک و نیم دیگر بالاتر از آن باشد. چنانچه شرایط مربوط به اندازه گروههای نمونه و فراوانی‌های نظری برای جدول توافقی ۲×۲ برقرار باشد، فرضیه یکسان بودن دو توزیع جامعه را می‌توان با به کار بردن آزمون x^2 با درجه آزادی d.f=k-1 ، که در آن k تعداد گروههای نمونه است،آزمون نمود. در اینجا، k= ۲ در نتیجه d.f =۲-۱=۱

جنس / وضعیت	زن	مرد	جمع
بالاتر از میانه مشترک	۷	۹	۱۶
پایین‌تر یا مساوی میانه مشترک	۱۱	۵	۱۶
جمع	۱۸	۱۴	۳۲

x^2 را محاسبه می‌کنیم: $x^2{}_c=1/14$ و $x^2{}_{(1,0/05)}=3/84$

چون $x^2{}_c \langle x^2{}_t$ است پس فرض H_0 مورد قبول است. به بیان دیگر، از لحاظ کارایی بین افسران جزء و افسران ارشد فرق معنی‌دار وجود ندارد.

آزمون من-ویتنی[1]

این آزمون معادل ناپارامتری آزمون زوج شده مقایسه میانگین‌ها (T-test) است. در این آزمون هر مقدار مشاهده با رتبه آن جایگزین می‌شود. اگر $x_1, \ldots, x_n$ نمونه تصادفی n تایی از جامعه اول و $y_1, \ldots, y_m$ نمونه تصادفی m تایی از جامعه دوم باشند، رتبه‌هایی از یک تا m+n را به ترکیب این مشاهدات نسبت می‌دهیم. $R(x_i), R(y_j)$ رتبه‌هایی فرض می‌شوند که به x_i و y_j نسبت داده شده‌اند. اگر چند نمونه مقداری دقیقاً مساوی داشته باشند (هم‌رتبه باشند) به هر یک از آنها مقدار متوسط رتبه‌ای را نسبت می‌دهند که در صورت برابر نبودن مشاهدات به آنها نسبت داده می‌شد. شرایط اجرای آزمون به صورت زیر است:

- هر دو نمونه، تصادفی انتخاب شده باشند.
- نمونه‌ها نسبت به یکدیگر و در داخل خود مستقل باشند.
- مقیاس اندازه‌گیری نمونه‌ها، ترتیبی یا فاصله‌ای باشند.
- اگر تفاوتی بین توزیع‌ها وجود داشته باشد، در مکان توزیع‌هاست، یعنی در صورت تفاوت $F(x), G(x), F(x) = G(x+c)$ ، c مقداری ثابت است. با این فرضیات آزمون فرض به صورت دو دنباله‌ای یا یک دنباله‌ای با فرضیه صفر $H_0 : F_x(x) = G_y(x)$ یا $H_0 : E(x) = E(y)$ اجرا می‌شود. اگر مجموع رتبه‌های جامعه اول

$$T = \sum_{i=1}^{n} R(x_i)$$

فرض شود، می‌توان T را به عنوان آماره آزمون به کار برد. اگر تعداد همرتبه‌ها زیاد باشد، با کم کردن میانگین از T و تقسیم کردن آن بر انحراف معیار، آماره

$$T_1 = \frac{T - n\frac{N+1}{2}}{\sqrt{\frac{nm}{N(N-1)}\sum_{i=1}^{N} R^2{}_i - \frac{nm(N+1)^2}{4(N-1)}}}$$

1 -Mann-Whitney U

جایگزین T می‌شود، که در آن $\sum R^2{}_i$ مجموع مربعات N رتبه (رتبه‌های متمایز یا متوسط گیری شده) در نمونه‌هاست.

آزمون کروسکال- والیس[1]

به ازای هر آزمون پارامتری معمولاً چندین آزمون ناپارامتری متناظر وجود دارد. حال آزمون کروسکال والیس که با آزمون F متناظر است بررسی می‌کنیم این آزمون در مواردی به کار برده می‌شود که مانند آزمون F تعداد گروهها بیش از دو باشد. علاوه بر آن بتوان مشاهدات را بین گروهها مستقل رتبه‌بندی کرد. یعنی مقیاس اندازه‌گیری، حداقل، ترتیبی باشد. به سخن دیگر، در مواردی که متغیرها رتبه‌ای باشد، به جای آزمون F این آزمون را به کار می‌برند. کارایی این آزمون در حدود ۹۵ درصد آزمونF است. اما محدودیت آن کمتر است. این آزمون متفاوت بودن یا نبودن بیش از دو گروه مستقل را از لحاظ رتبه‌بندی مورد ارزیابی قرار می‌دهد. شکل توزیع کروسکال- والیس تقریباً به صورت توزیع χ^2 با درجه آزادی d.f= k-1 است. توجه داشته باشید که اگر هیچ تفاوتی در بین طبقات(یا گروهها) وجود نداشته باشد رتبه‌ها را باید به طور تصادفی در هر گروه توزیع کرد. در این صورت باید میانگین آنها در هر گروه تقریباً برابر باشد. به عبارتی اگر تفاوتی وجود داشته باشد، همه جامعه‌ها دارای میانگین یکسان نخواهد بود. (H_1). مساله مهم که باید بدان توجه داشت این است که فرضیه‌ها مخالف در این آزمون همواره بدون جهت است، یعنی نشان می‌دهد که گروهها با یکدیگر تفاوت دارند. اما هرگز نمی‌گوید که کدام گروه بالاتر و کدام گروه پایین‌تر است.

مشخصه آماری مشاهده شده در تحلیل واریانس رتبه‌ای کروسکال- والیس را با H یا T نشان می‌دهند. ولی از آنجایی که توزیع کروسکال والیس شبیه توزیع x^2 است. بنابراین در تفسیر آزمون کروسکال والیس از جدول x^2 استفاده می‌شود، علامت فرمول کروسکال - والیس را با $x^2{}_{k.w}$ نیز نمایش می‌دهند. بنابراین فرمول آزمون کروسکال والیس عبارت است از:

1. kruskal Wallis

$$\chi^2_{k.w} = H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R^2{}_i}{n_i} - 3(N+1)$$

در این فرمول: n_i بارت است از تعداد رتبه گروه i ام و N تعداد کل رتبه‌ها یا به طور دقیق‌تر تعداد کل افراد جامعه، یعنی $N=\sum_{i=1}^{k} n_i$ و K تعداد گروهها و سرانجام $R^2{}_i$ مجموع مجذور رتبه‌های گروه iام است. مقدار H به دست آمده با استفاده از جدول x^2 با درجه آزادی k-1 قضاوت و تفسیر می‌شود. توجه داشته باشید که در اینجا H_0 عبارت است از اینکه K تابع توزیع جامعه‌ها یکسان هستند. همچنین H_1 عبارت است از اینکه حداقل یکی از جامعه‌ها نسبت به حداقل یکی از جامعه‌های دیگر دارای مشاهداتی بزرگتر است.

مثال:

محققی می‌خواهد سه مقطع آموزش ستادی را بررسی کند. برای این کار با کلیه اعضای آموزشی این سه مقطع - که ۱۷ نفر از درجه داران می‌باشند - مصاحبه کرده است. در نتیجه این مصاحبه، براساس جواب‌های اعضای آموزشی به آنها طبق جدول زیر رتبه داده است. می‌خواهیم بدانیم آیا از لحاظ رتبه‌بندی این سه مقطع با هم فرق معنی‌دار دارند یا نه؟

رتبه‌های مقطع سوم	رتبه‌های مقطع دوم	رتبه‌های مقطع اول
	۱۱	۱۲
۴	۱۶	۱۴
۳	۵	۱۰
۸	۷	۱۷
۱	۶	۱۵
۹	۲	۱۳
R_3 = ۲۵	R_2 = ۴۷	R_1 = ۸۱
$\overline{x}_3$ =۵/۰	$\overline{x}_2$ =۷/۸۳	$\overline{x}_1$ =۱۳/۵

$$H = \frac{12}{N(N+1)} \sum \frac{R^2{}_i}{n_i} - 3(N+1)$$

برای محاسبه H ابتدا باید $\sum \frac{R^2{}_i}{n_i}$ را حساب کنیم.

$$\sum_{i=1}^{3} \frac{R^2{}_i}{n_i} = \frac{(81)^2}{6} + \frac{(47)^2}{6} + \frac{(25)^2}{5} = 1093/5 + 368/17 + 125 = 1586/67$$

اگر این مقادیر را در فرمول بالا قرار دهیم، خواهیم داشت:

$$H = \frac{12}{17(17+1)} \times 1586/67 - 3(17+1) = 8/22$$

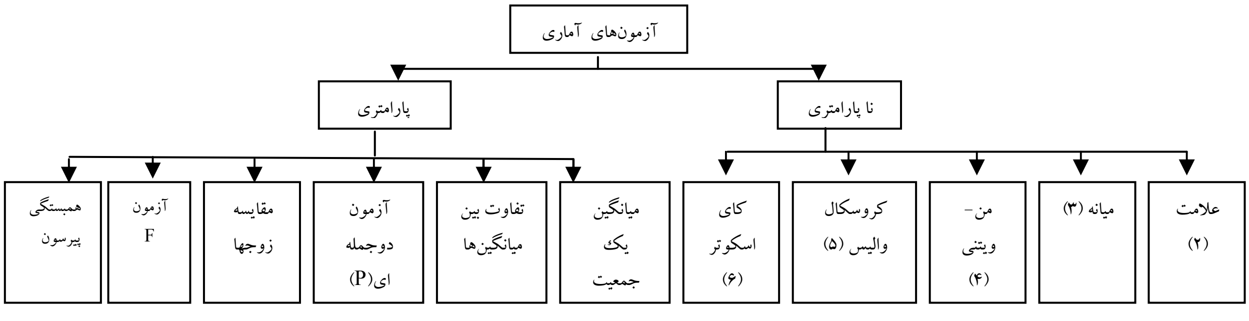
چون H تقریباً به صورت توزیع x^2 با درجه آزادی d.f=k-1 است.

پس با x^2 جدول مقایسه می‌کنیم. d.f=k-1=3-1=2

ملاحظه می‌شود که در سطح ۰/۰۲ معنی‌دار است.($x^2{}_{(2,0.02)} = 7/82$)

بنابراین، رتبه‌ها تصادفی نیستند و سه مقطع آموزشی با هم فرق معنی‌دار دارند.

نتایج



(۱) این آزمون نه مستلزم شرایط نرمال بودن است و نه مستلزم اندازه‌گیری فاصله‌ای

(۲) مقیاس این آزمون ترتیبی است و معادل آزمون میانگین در آزمون پارامتری است.

(۳) این آزمون همتای آزمون‌های t و z برای تفاوت بین دو جامعه مستقل با توزیع‌های یکسان می‌باشد.

(۴) این آزمون معادل ناپارامتری آزمون زوج شده مقایسه میانگین‌ها است.

(۵) این آزمون با آزمون F متناظر است و در مواردی که متغیرها رتبه‌ای باشند به کار برده می شود.

(۶) این آزمون معادل آزمون همبستگی پیرسون در آزمون پارامتری است و رابطه بین دو متغیر در سطح اسمی، ترتیبی و حتی فاصله‌ای را آزمون می‌کند.

منابع فارسی

پاک‌گوهر، علیرضا. (۱۳۸۸). *کاربرد آمار در مدیریت ترافیک*. دانشگاه علوم انتظامی: دانشکده راهنمایی و رانندگی.

جزینی، علیرضا. (۱۳۸۵). *عوامل مؤثر در توانمندسازی افسران ستاد ناجا*. رساله کارشناسی ارشد، دانشکده علوم انسانی دانشگاه تربیت مدرس.

چالمرز، آلن. (۱۳۷۳). *علم چیست؟* ترجمه محمد مشایخی،تهران: شرکت سهامی انتشار.

چاوا فرانکفورد/ دیوید. (۱۳۸۱). *روشهای پژوهش در علوم اجتماعی*. ترجمه دکتر فاضل لاریجانی و رضا فاضلی تهران: سروش.

عرب‌محبی، احمد. (۱۳۸۵). *بررسی رابطه آموزش فراگیران رسته انتظامی با اثربخشی عملکرد آنان با کلانتریهای فرماندهی انتظامی تهران بزرگ*. رساله کارشناسی ارشد، دانشکده فرماندهی و ستاد دانشگاهی علوم انتظامی.

گوری ک. باتاچاریا، ریچارد ا. جانسون. (۱۳۷٤). *مفاهیم و روشهای آماری*. جلد دوم، ترجمه مرتضی ابن شهر آشوب، فتاح میکائیلی، تهران: مرکز نشر دانشگاهی.

واین. و. دانیل. (۱۳۷۸). *اصول و روشهای آمار زیستی*. ترجمه دکتر سیدمحمدتقی آیت‌اللهی، تهران: مؤسسه انتشارات امیرکبیر.

ایران نژاد پاریزی، مهدی. (۱۳۷۸). *روشهای تحقیق در علوم اجتماعی*. تهران:نشر مدیران.

هومن، حیدرعلی. (۱۳۷۸). *استنباط آماری در پژوهش رفتاری*. تهران: نشر پارسا.

منابع لاتین

Barker, c. pistrang, N. & Elliot, R.ResearchMethods in clinical and Counselling psy choloay. By John Wiley & sons. (1994).

Cohen, L : Maniion. L0, & Morrison.Research Methods in Education, by Routledge Falmer, 5 th edition. (2001).