

کدام آزمون آماری را انتخاب کنیم؟*

دکتر غلامرضا جندقی^۱

چکیده

یکی از مشکلات محققین در تحقیقات میدانی انتخاب روش تحلیل آماری مناسب برای بررسی سوالات یا فرضیات تحقیق می‌باشد. در این مقاله کاربرد انواع آزمون‌های آماری با توجه به نوع داده‌ها و ویژگیهای نمونه آماری و نوع تحلیل مورد نیاز بررسی می‌شود و چگونگی انتخاب آزمون مناسب به روش الگوریتمیک و نموداری نشان داده شده و کاربرد آزمونهای آماری به تفکیک انواع پارامتری و ناپارامتری بررسی می‌گردد.

واژه‌های کلیدی

آمار پارامتری^۲ - آمار ناپارامتری^۳ - آزمون‌های آماری^۴

۱ - مقدمه: در بسیاری از مطالعات میدانی در مدیریت، محقق حجم معنابهی از داده‌های آماری را جمع‌آوری می‌نماید که بایستی به نوعی از طریق آنها به سوالات یا فرضیات تحقیق خود پاسخ دهد. در این راه مهمترین ابزار محقق استفاده از علم آمار

* تاریخ دریافت ۸۳/۸/۲۱ تاریخ پذیرش ۸۳/۱۰/۱۵

۱. استادیار دانشگاه تهران - پردیس قم

2. Parametric statistics
3. Non parametric statistics
4. statistical tests

است. سوال تحقیق معمولاً بر اساس هدف کلی و اهداف جزئی ساخته می‌شود. به عنوان مثال اگر هدف کلی یکی تحقیق بررسی میزان رضایت کارکنان یک سازمان باشد؛ سوال تحقیق می‌تواند به صورت "میزان رضایت کارکنان چقدر است؟" و یا "آیا میزان رضایت کارکنان زن و مرد یکسان است؟" باشد.

در آمار استنباطی روش‌های گوناگونی برای تجزیه و تحلیل داده‌های آماری وجود دارد که هر یک برای نوع خاصی از داده‌ها و همچنین نوع خاصی از سوالات تحقیق مورد استفاده قرار می‌گیرند. مشکل اصلی محقق انتخاب روش آماری مناسب برای پاسخگویی به سوالات تحقیق است و غالباً عدم مشورت محقق با متخصص آمار موجب استفاده نادرست از روش‌های آماری و خدشه‌دار شدن نتایج تحقیق می‌گردد. در این مقاله سعی بر این است که با معرفی آزمون‌های گوناگون آماری الگویی برای انتخاب درست آنها ارائه گردد. البته فرض این است که خواننده آشنایی نسبتاً خوبی با روش‌های آماری دارد.

۲ - **انتخاب آزمون آماری:** یک آزمون آماری وسیله‌ای است برای تصمیم‌گیری‌های کمی در مورد یک یا چند فرآیند. در آزمون‌های آماری هدف تعیین این مطلب است که آیا داده‌های نمونه شواهد کافی برای رد یک حدس یا فرضیه در مورد فرآیند ارائه می‌دهد؟ در تحقیقات میدانی مربوط به علم مدیریت غالباً محقق داده‌های زیادی جمع‌آوری نموده و برای پاسخگویی به سوالات تحقیق بایستی آزمون مناسبی را انتخاب کند. انتخاب آزمون آماری به عوامل مختلفی مثل طبیعت داده‌ها و ویژگی‌های نمونه و نوع استنباط مورد نیاز بستگی دارد. تعداد متغیرها و نوع داده‌ها از لحاظ کمی کیفی، پیوسته یا گسسته بودن بیانگر طبیعت داده‌هاست.

ویژگی‌های نمونه شامل تعداد گروه‌های مورد بررسی و نوع نمونه از نظر نرمال بودن یا نبودن جامعه‌ای که نمونه از آن گرفته شده می‌باشد و نوع استنباط مورد نیاز می‌تواند تنها توصیف جامعه مورد بررسی یا مقایسه اختلاف دو یا چند گروه و یا سنجش رابطه بین دو یا چند متغیر باشد. بنابراین قبل از انتخاب یک آزمون آماری بایستی به سوالات زیر پاسخ داد (لی ویلاگی، ۱۹۹۶).

۱ - چه تعداد متغیر مورد بررسی قرار می‌گیرد؟

۲ - چند گروه مقایسه می‌شوند؟

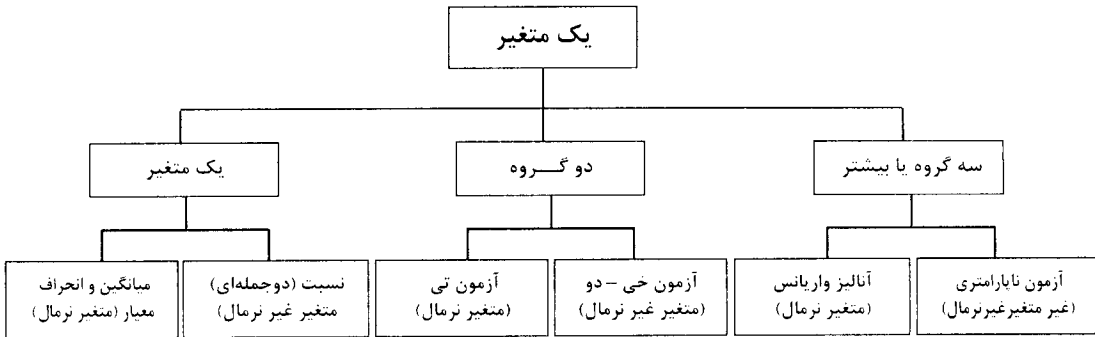
۳ - آیا توزیع ویژگی مورد بررسی در جامعه نرمال است؟

۴ - آیا گروه‌های مورد بررسی مستقل‌اند؟

۵ - سوال یا فرضیه تحقیق چیست؟

۶ - آیا داده‌ها پیوسته، رتبه‌ای و یا مقوله‌ای^۱ هستند؟

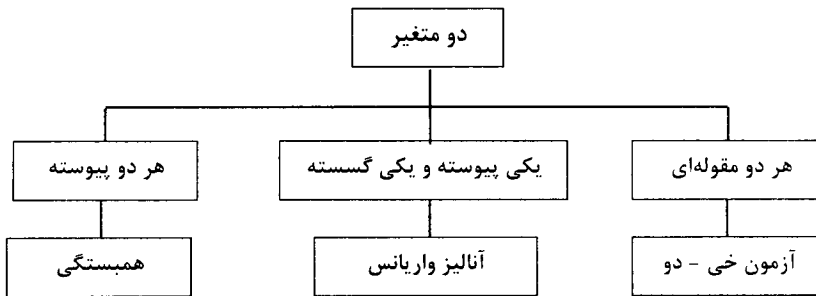
نمودار (۱) چگونگی انتخاب آزمون آماری برای یک متغیر را نشان می‌دهد. محقق ممکن است یک متغیر را در یک یا چند گروه بررسی نماید. قبل از هر چیز بایستی از وضعیت نرمال بودن توزیع آن متغیر اطمینان حاصل کند. در حالت نرمال بودن توزیع متغیر مورد بررسی از شاخصهای میانگین و انحراف معیار از مون تی^۲ و یا آنالیز واریانس^۳ استفاده می‌نماید. اگر متغیر مورد نظر نرمال نباشد از آزمون‌های موسوم به ناپارامتری کمک می‌گیرد که در ادامه به بررسی آن‌ها خواهیم پرداخت.



نمودار (۱): انتخاب آزمون آماری برای یک متغیر

- 1 . Categorical
- 2 . T - Test
- 3 . Anova

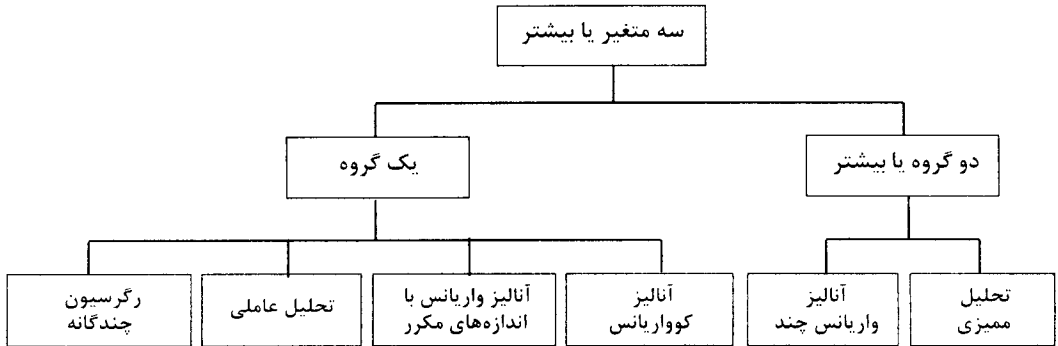
نمودار (۲) روش تشخیص آزمون مورد نیاز برای بررسی رابطه بین دو متغیر را نشان می‌دهد. دو متغیر مورد بررسی می‌توانند هر دو از نوع پیوسته باشند که در این صورت از ضریب همبستگی^۱ استفاده می‌شود. اگر یکی از دو متغیر گسسته و دیگری پیوسته باشد از آنالیز واریانس و هنگامی که دو متغیر از نوع مقوله‌ای (کیفی) باشند از آزمون خی - دو^۲ بهره می‌گیریم.



نمودار ۲: انتخاب آزمون آماری برای بررسی ارتباط و متغیر

نمودار (۳) چگونگی انتخاب آزمون برای بررسی سه متغیر یا بیشتر را نشان می‌دهد. اگر متغیرهای مورد نظر برای یک گروه بررسی می‌شود بسته به مورد از روش‌های رگرسیون چندگانه^۳ تحلیل عاملی^۴ و یا آنالیز واریانس با اندازه‌های تکراری^۵ استفاده می‌شود. در حالتی که متغیرهای مورد نظر برای دو یا چند گروه مورد تجزیه و تحلیل قرار می‌گیرد می‌توان از آنالیز واریانس چند متغیره^۶ و یا تحلیل ممیزی^۱ سود برد (سریواستاوا^۲، ۲۰۰۲).

- 1 . Correlation Coefficient
- 2 . Chi - Square
- 3 . Multiple Regression
- 4 . Factor Analysis
- 5 . Anova With Repeated Measures
- 6 . Multivariate Anova



نمودار ۳: انتخاب آزمون آماری برای سه متغیر و بیشتر

۳ - آزمون‌های پارامتری و ناپارامتری^۳: بسیاری از آزمون‌های آماری بر اساس

این فرض استوارند که داده‌ها از جامعه‌ای نرمال نمونه‌گیری شده‌اند. چنین آزمون‌هایی را آزمون‌های پارامتری گویند. یکی از آزمون‌های پارامتری شایع آزمون تی می‌باشد. آزمون‌هایی که بر اساس فرض نرمال بودن جامعه استوار نیستند به آزمون‌های ناپارامتری موسومند. آزمون‌های مبتنی بر رتبه‌ها مانند آزمون من - ویتنی^۴ و آزمون کروسکال والیس^۵ (هاروی موتولسکی^۶، ۱۹۹۵) از آزمون‌های ناپارامتری به شمار می‌روند. تشخیص این‌که در یک تحقیق به آزمون ناپارامتری نیازمندیم یا نه در پاره‌ای از مواقع آسان در مواردی مشکل و در مواردی اهمیت چندانی ندارد.

۳-۱ - مواردی که تشخیص آسان است: اگر متغیر مورد بررسی از توزیع

نرمال یا تقریباً نرمال پیروی کند بایستی از آزمون‌های پارامتری استفاده نمود اما اگر

- 1 . Discriminant Analysis
- 2 . Srivastava
- 3 . Non - Parametric Tests
- 4 . Mann - Whitney
- 5 . Kruskal - Walis
- 6 . Harvey Motulsky

متغیر مورد بررسی رتبه‌ای باشد واضح است که توزیع نرمال ندارد مثل رتبه‌بندی دانشجویان از نظر کیفیت درسی (ضعیف - متوسط - قوی) و یا رتبه‌بندی هتل‌ها (یک ستاره - دو ستاره و...) که در این مواقع بایستی از آزمون‌های ناپارامتری بهره گرفت.

۳-۲ - مواردی که تشخیص مشکل است: وقتی داده‌ها کمی هستند و تعداد

نمونه کم است تشخیص نرمال بودن داده‌ها توسط آزمون کولموگروف - اسمیرنف^۱ مشکل است زیرا با حجم نمونه کم توان تشخیص آزمون کولموگروف - اسمیرنف کاهش می‌یابد. بنابراین برخی محققین از آزمون‌های پارامتری استفاده می‌کنند چون مطمئن نیستند که نرمال بودن داده‌ها خدشه‌دار شده است و برخی دیگر از آزمون‌های ناپارامتری به کار می‌برند چون از نرمال بودن داده‌ها مطمئن نیستند. (لی^۲، ۱۹۹۶).

۳-۳ - مواردی که استفاده از هر دو نوع آزمون تفاوت چندانی ندارد:

وقتی نمونه بزرگ باشد قضیه حد مرکزی (جندقی و همکاران، ۱۳۷۹) تضمین می‌کند که آزمون‌های پارامتری می‌توانند مورد استفاده قرار گیرند حتی اگر جامعه نرمال نباشد. اما این که چقدر یک نمونه بایستی بزرگ باشد محل اختلاف آمار دانان است. اگر توزیع متغیر مورد بررسی خیلی غیر طبیعی نباشد در یک نمونه ۳۰ تایی و بالاتر می‌توان از آزمون‌های پارامتری استفاده کرد. در نمونه‌های بزرگ حتی اگر از جامعه‌ای نرمال انتخاب شده باشد می‌توان از آزمون‌های ناپارامتری استفاده نمود تنها تفاوت این دو نوع آزمون این است که آزمون‌های ناپارامتری توان تشخیصی کمتر دارند. بنابراین در نمونه‌های بزرگ به کار گرفتن هر یک از دو نوع آزمون مشکلی ایجاد نمی‌کند.

۴- گروه‌های مستقل و گروه‌های وابسته (همبسته): وقتی دو گروه از نظر یک

یا چند متغیر مقایسه می‌شوند محقق باید بداند که آیا این گروه‌ها مستقل هستند یا وابسته.

دو گروه وابسته به دو صورت به وجود می‌آید:

1 . Kolmogrov - Smirnov

2 . Lee

۴- ۱- اگر ویژگی یک مجموعه از افراد قبل و بعد از وقوع یک عامل سنجیده شود و هدف مقایسه آن ویژگی قبل و بعد از وقوع آن عامل باشد اصطلاحاً دو گروه وابسته داریم. مقایسه رضایت کارمندان یک سازمان قبل و بعد از انجام یک برنامه مثالی از دو گروه وابسته است یعنی دو اندازه‌گیری بر روی یک مجموعه از افراد تشکیل دو گروه وابسته می‌دهند. (میلتون و سوکاس^۱، ۱۹۹۲).

۴- ۲- گاهی در مطالعات تجربی افراد را از نظر برخی ویژگیها در یک گروه با گروه دیگر همسان^۲ می‌کنند به طوری که یک تناظر یک به یک بین افراد دو گروه برقرار می‌شود. در این صورت نیز اگر متغیری برای دو گروه اندازه‌گیری شود تشکیل دو گروه وابسته می‌دهند. بنابراین همواره تعداد افراد در دو گروه وابسته برابر است.

وقتی هیچ یک از دو حالت فوق رخ ندهد دو گروه را مستقل گوئیم. تشخیص استقلال یا وابستگی گروه‌های مورد مطالعه نقش اساسی در انتخاب آزمون آماری دارد.

۵- کمی یا کیفی بودن داده‌ها: داده‌های آماری نتایج حاصل از اندازه‌گیری، شمارش، آزمایش، مشاهده و گاهی مطالعه افراد یک جامعه به دست می‌آیند. داده‌ها یا متغیرهای یک تحقیق به روش‌های مختلف اندازه‌گیری و ثبت می‌شوند (جندقی و همکاران ۱۳۷۸). نوع مقیاس اندازه‌گیری در انتخاب آزمون مناسب تعیین کننده است. برخی از متغیرها با مقیاس کیفی اندازه‌گیری می‌شوند. داده کیفی داده‌ای است غیر عددی که برای نامگذاری افراد یا اشیا به کار می‌رود. مقیاسهای کیفی یا اسمی هستند مانند "رشته تحصیلی" و "رتبه‌ای مثل" رتبه فرد در کلاس. برخی دیگر از متغیرها به صورت کمی ثبت می‌شوند. داده‌های کمی به دو صورت پیوسته و گسسته مشاهده می‌شوند. "نمره رضایتمندی" و "سن افراد" پیوسته و "تعداد فرزندان" از نوع گسسته می‌باشند. بدیهی است که به تناسب نوع داده روش تجزیه و تحلیل تفاوت خواهد داشت.

1 . Milton & Tsokos

2 . Matching

۶- الگوریتم انتخاب آزمون آماری: با در دست داشتن داده‌های آماری محقق ابتدا باید سوال خود را مشخص کند و اگر بیش از یک سوال مطرح است یک به یک به آنها پاسخ دهد. معمولاً سوال محقق به یکی از صورتهای زیر مطرح می‌شود.

۱ - آیا اختلافی بین میانگین (نسبت) یک ویژگی در دو یا چند گروه وجود دارد؟

۲ - آیا دو متغیر رابطه دارند؟

۳ - چگونه می‌توان یک متغیر را با استفاده از متغیرهای دیگر پیش‌بینی کرد؟

۴ - چه چیزی می‌توان با استفاده از نمونه در مورد جامعه گفت؟ (سایت، استات الگوریتم^۱).

۶-۱ - تحلیل اختلاف بین دو یا چند گروه: اگر متغیر مورد بررسی کمی و توزیع آن در جامعه نرمال است برای مقایسه دو گروه مستقل از آزمون تی و برای دو گروه وابسته از آزمون تی - زوجی استفاده می‌شود. در صورت نرمال نبودن جامعه آزمون من - ویتنی برای دو گروه مستقل و آزمون ویلکاکسون^۲ برای دو نمونه وابسته به کار برده می‌شود. برای مقایسه چند گروه بسته به این که جامعه نرمال باشد یا نباشد به ترتیب از آنالیز واریانس، آزمون کروسکال - والیس، آزمون فریدمن و کوکران استفاده می‌شود.

۶-۲ - رابطه دو متغیر: گاهی هدف بررسی رابطه دو متغیر است. اگر دو متغیر مورد بررسی از نوع کیفی اسمی باشند از آزمون خی^۲ استفاده می‌شود. برای آزمون رابطه (همبستگی) بین دو متغیر کیفی رتبه‌ای ضریب همبستگی اسپیرمن^۳ را می‌توان به کار برد. برای بررسی وجود همبستگی بین دو متغیر کمی از ضریب همبستگی پیرسون^۴ استفاده می‌شود. البته ضرایب همبستگی دیگر مثل ضریب همبستگی کندال^۵ و ضریب

-
- 1 . StatAlgorithm Homepage
 - 2 . Willcoxon Test
 - 3 . Spearman Rank Correlation Coefficient
 - 4 . Pearson
 - 5 . Kendall

همبستگی کرامر^۱ نیز وجود دارند که بر حسب این که دو متغیر مورد نظر یکی کیفی اسمی و دیگر کیفی رتبه‌ای باشد مورد استفاده قرار می‌گیرند. لذا محقق می‌تواند برای تفصیل بیشتر به کتابهای آمار تحلیلی از جمله (اگرستی^۲، ۲۰۰۲) مراجعه نماید.

۶ - ۳ - پیش‌بینی یک متغیر با استفاده از یک یا چند متغیر دیگر: چنانچه محقق قصد داشته باشد روند تغییرات یک متغیر را بر اساس یک یا چند متغیر دیگر پیش‌بینی نماید، از رگرسیون^۳ ساده و یا چندگانه استفاده می‌کند. در مدل رگرسیونی، متغیر پیش‌بینی کننده را متغیر مستقل و متغیر پیش‌بینی شونده را متغیر وابسته گوئیم (نرو و اسرمن^۴، ۱۹۸۹).

به عنوان مثال اگر بخواهیم میزان حقوق کارمندان را از روی سابقه خدمت آنها پیش‌بینی نماییم، متغیر میزان حقوق را متغیر وابسته و متغیر سابقه خدمت را متغیر مستقل می‌نامیم. مدل رگرسیونی ممکن است به صورت خطی^۵ یا غیر خطی باشد. چنانچه متغیر وابسته از نوع متغیرهای دو مقداری^۶ باشد بایستی از مدل رگرسیون لجستیک^۷ استفاده نمود.

۶ - ۴ - رابطه جامعه و نمونه: در بسیاری از مواقع محقق داده‌ها را از طریق سرشماری از تمام افراد جامعه مورد بررسی به دست آورده است. در این صورت ممکن است فقط نیاز به توصیف داده‌ها داشته باشد. به عبارت دیگر تنها از روشهای آمار توصیفی شامل شاخصها و نمودارها استفاده می‌نماید.

جدول (۱) نحوه انتخاب آزمون آماری مناسب را نشان می‌دهد.

-
- 1 . Cramer
 - 2 . Agresti
 - 3 . Regression
 - 4 . Neter & Wasserman
 - 5 . Linear
 - 6 . Binary
 - 7 . Logistic

جدول (۱): انتخاب آزمون آماری مناسب بر حسب نوع داده و هدف آزمون

هدف نوع داده	کمی و دارای توزیع نرمال	رتبه‌ای و یا کمی غیر نرمال	مقوله‌ای دو تایی
توصیف یک گروه	میانگین و انحراف معیار	میانه	نسبت
مقایسه یک گروه با یک مقدار فرضی	آزمون یک نمونه‌ای	آزمون ویلکاکسون	آزمون خی-دو یا دو جمله ای
مقایسه دو گروه مستقل	آزمون برای نمونه‌های مستقل	آزمون من-ویتنی	آزمون دقیق فیشر (آزمون خی-دو برای نمونه‌های بزرگ)
مقایسه دو گروه وابسته	آزمون زوجی	آزمون کروسکال	آزمون مک-نمار
مقایسه سه گروه یا بیشتر (مستقل)	آنالیز واریانس یک راهه	والیس	آزمون خی-دو
مقایسه سه گروه یا بیشتر (وابسته)	آنالیز واریانس با اندازه‌های مکرر	آزمون فریدمن	آزمون کوکران
اندازه‌گیری همبستگی بین دو متغیر	ضریب همبستگی پیرسون	ضریب همبستگی اسپرمن	ضرایب توافق
پیش‌بینی یک متغیر بر اساس یک یا چند متغیر دیگر	رگرسیون ساده یا غیر خطی	رگرسیون نابارامتری	رگرسیون لجستیک

منابع

- ۱ - جندقی، غلامرضا و صادقی، احمد (۱۳۷۸): آمار توصیفی برای کلیه رشته‌های علوم انسانی، قم، انتشارات حضور.
- ۲ - جندقی و همکاران (۱۳۷۹): آمار و احتمال مهندسی، تهران، انتشارات دانشگاه امام حسین(ع)
- 3 - Agresti A.(2002): Categorical Data Analysis. USA. John Willey Company
- 4 - Lee Willoughby T. (1996): What Staistical Test is Appropriata?
- 5 - Milton J.S., Tsokos J.O (1996): Statistical Methods in the Biological and Health Sciences, USA. NcGraw Hill Company
- 6 - Motulsky H.(1995): Intuitive Biostatistics, Osford University Press
- 7 - Neter J., Wasserman W. (1989): Applied Linear Statistical Models, USA, Richard D. Irwin Inc.
- 8 - Srivastava M.S (2002): Methods of Multivariate Statistics. USA. John Willey Company
- 9 - Statalgorithm Webpage: A: Simple Algorithm to Help Decide Statistical Test to Use [www. Edu.rcsed. ac. Uk/statistics](http://www.Edu.rcsed.ac.Uk/statistics)