

انتخاب صحیح آزمون آماری

آیا قصد دارید تحقیقی را انجام دهید؟ و یا اینکه در حال مطالعه یک تحقیق می باشید؟ چگونه می توانید از صحت روش تجزیه و تحلیل داده ها اطمینان حاصل فرمائید؟ شاخه های مختلف علوم برای تجزیه و تحلیل داده ها از روش های مختلفی مانند روش های ذیل استفاده می نمایند:

- الف) روش تحلیل محتوا
- ب) روش تحلیل آماری
- ج) روش تحلیل ریاضی
- د) روش اقتصادسنجی
- ه) روش ارزشیابی اقتصادی

و...

تمرکز این نوشتار بر روش های تجزیه و تحلیل سیستمهای اقتصادی اجتماعی و بویژه روش های تحلیل آماری می باشد. آمار علم طبقه بندی اطلاعات، علم تصمیم گیری های علمی و منطقی، علم برنامه ریزی های دقیق و علم توصیف و بیان آن چیزی است که از مشاهدات می توان فهمید. هدف ما آموزش درس آمار نیست زیرا اینگونه مطالب تخصصی را میتوان در مراجع مختلف یافت. هدف اصلی ما ارائه یک روش دستیابی سریع به بهترین روش آماری می باشد. یکی از مشکلات عمومی در تحقیقات میدانی انتخاب روش تحلیل آماری مناسب و یا به عبارتی انتخاب آزمون آماری مناسب برای بررسی سوالات یا فرضیات تحقیق می باشد. در آزمون های آماری هدف تعیین این موضوع است که آیا داده های نمونه شواهد کافی برای رد یک حدس یا فرضیه را دارند یا خیر؟

انتخاب نادرست آزمون آماری موجب خدشه دار شدن نتایج تحقیق می شود. دکتر غلامرضا جندقی استادیار دانشگاه تهران در مقاله ای کاربرد انواع آزمون های آماری را با توجه به نوع داده ها و ویژگی های نمونه آماری و نوع تحلیل نشان داده است که در این بخش به نکات کلیدی آن اشاره می شود:

قبل از انتخاب یک آزمون آماری بایستی به سوالات زیر پاسخ داد:

- ۱- چه تعداد متغیر مورد بررسی قرار می گیرد؟
 - ۲- چند گروه مقایسه می شوند؟
 - ۳- آیا توزیع ویژگی مورد بررسی در جامعه نرمال است؟
 - ۴- آیا گروه های مورد بررسی مستقل هستند؟
 - ۵- سوال یا فرضیه تحقیق چیست؟
 - ۶- آیا داده ها پیوسته، رتبه ای و یا مقوله ای Categorical هستند؟
- قبل از ادامه این مبحث لازم است مفهوم چند واژه آماری را یاد آور شوم که زیاد وقت گیر نیست.

۱- **جامعه آماری:** به مجموعه کاملی از افراد یا اشیاء یا اجزاء که حداقل در یک صفت مورد علاقه مشترک باشند، گفته می شود.

۲- **نمونه آماری:** نمونه بخشی از یک جامعه آماری تحت بررسی است که با روشی که از پیش تعیین شده است انتخاب می شود، به قسمی که می توان از این بخش، استنباطهایی درباره کل جامعه بدست آورد.

۳- **پارامتر و آماره:** پارامتر یک ویژگی جامعه است در حالی که آماره یک ویژگی نمونه است. برای مثال میانگین جامعه یک پارامتر است. حال اگر از جامعه نمونه گیری کنیم و میانگین نمونه را بدست آوریم، این میانگین یک آماره است.

۴- **برآورد و آزمون فرض:** برآوردیابی و آزمون فرض دو روشی هستند که برای استنباط درمورد پارامترهای مجهول دو جمعیت به کار می روند.

۵- **متغیر:** ویژگی یا خاصیت یک فرد، شیئ و یا موقعیت است که شامل یک سری از مقادیر با دسته بندیهای متناسب است. قد، وزن، گروه خونی و جنس نمونه هایی از متغیر هستند. انواع متغیر می تواند کمی و کیفی باشد.

۶- **داده های کمی:** مانند قد، وزن یا سن درجه بندی می شوند و به همین دلیل قابل اندازه گیری می باشند. داده های کمی نیز خود به دو دسته دیگر تقسیم می شوند:

الف: داده های فاصله ای (Interval data)

ب: داده های نسبتی (Ratio data)

۷- **داده های فاصله ای:** به عنوان مثال داده هایی که متغیر IQ ضریب هوشی) را در پنج نفر توصیف می کنند عبارتند از: ۸۰، ۱۱۰، ۷۵، ۹۷ و ۱۱۷، چون این داده ها عدد هستند پس داده های ما کمی اند اما می دانیم که IQ نمی تواند صفر باشد و صفر در اینجا فقط مبنایی است تا سایر مقادیر IQ در فاصله ای منظم از صفر و یکدیگر قرار گیرند پس این داده ها فاصله ای اند.

۸- **داده های نسبتی:** داده های نسبتی داده هایی هستند که با عدد نوشته می شوند اما صفر آنها واقعی است. اکثریت داده های کمی این گونه اند و حقیقتاً دارای صفر هستند. به عنوان مثال داده هایی که متغیر طول پاره خط بر حسب سانتی متر را توصیف می کنند عبارتند از: ۲۰، ۱۵، ۲۵، ۸ و ۳۳، چون این داده ها عدد هستند پس داده های ما کمی اند و چون صفر در اینجا واقعاً وجود دارد این داده نسبتی تلقی می شوند.

۹- **داده های کیفی:** مانند جنس، گروه خونی یا ملیت فقط دارای نوع هستند و قابل بیان با استفاده از واحد خاصی نیستند. داده های کیفی خود به دو دسته دیگر تقسیم می شوند:

الف: داده های اسمی (Nominal data)

ب: داده های رتبه ای (Ordinal data)

۱۰- **داده های رتبه ای:** Ordinal مانند کیفیت درسی یک دانش آموز (ضعیف، متوسط و قوی) و یا رتبه بندی هتل ها (یک ستاره، دو ستاره و...) است.

۱۱- **داده های اسمی:** (nominal) که مربوط به متغیر با خواص کیفی مانند جنس یا گروه خونی است و بیانگر عضویت در یک گروه category خاص می باشد. (داده مقوله ای)

۱۲- **متغیر تصادفی کسسته و پیوسته:** به عنوان مثال تعداد تصادفات جاده‌ای در روز یک متغیر تصادفی گسسته است ولی انتخاب یک نقطه به تصادف روی دایره‌ای به مرکز مبدأ مختصات و شعاع ۳ یک متغیر تصادفی پیوسته است.

۱۳- **گروه:** یک متغیر می تواند به لحاظ بررسی یک ویژگی خاص در یک گروه و یا دو و یا بیشتر مورد بررسی قرار گیرد. نکته ۱: دو گروه می تواند وابسته و یا مستقل باشد. دو گروه وابسته است اگر ویژگی یک مجموعه افراد قبل و بعد از وقوع یک عامل سنجیده شود. مثلاً میزان رضایت شغلی کارکنان قبل و بعد از پرداخت پاداش و همچنین اگر در مطالعات تجربی افراد از نظر برخی ویژگی ها در یک گروه با گروه دیگر همسان شود.

۱۴- **جامعه نرمال:** جامعه ای است که از توزیع نرمال تبعیت می کند.

۱۵- **توزیع نرمال:** یکی از مهمترین توزیع ها در نظریه احتمال است. و کاربردهای بسیاری در علوم دارد.

فرمول این توزیع بر حسب دو پارامتر امید ریاضی و واریانس بیان می شود. منحنی رفتار این تابع تا حد زیادی شبیه به زنگ های کلیسا می باشد. این منحنی دارای خواص بسیار جالبی است برای مثال نسبت به محور عمودی متقارن می باشد، نیمی از مساحت زیر منحنی بالای مقدار متوسط و نیمه دیگر در پایین مقدار متوسط قرار دارد و اینکه هرچه از طرفین به مرکز مختصات نزدیک می شویم احتمال وقوع بیشتر می شود.

سطح زیر منحنی نرمال برای مقادیر متفاوت مقدار میانگین و واریانس فراگیری این رفتار آنقدر زیاد است که دانشمندان اغلب برای مدل کردن متغیرهای تصادفی که با رفتار آنها آشنایی ندارند، از این تابع استفاده می کنند. به عنوان مثال در یک امتحان درسی نمرات دانش آموزان اغلب اطراف میانگین بیشتر می باشد و هر چه به سمت نمرات بالا یا پایین پیش برویم تعداد افرادی که این نمرات را گرفته اند کمتر می شود. این رفتار را بسهولت می توان با یک توزیع نرمال مدل کرد.

اگر یک توزیع نرمال باشد مطابق قضیه چی بی شف ۶۸،۲۶٪ مشاهدات در فاصله میانگین، مثبت و منفی یک انحراف معیار قرار دارد. و ۹۵،۴۴٪ مشاهدات در فاصله میانگین، مثبت و منفی دو انحراف معیار قرار دارد. و ۹۹،۷۳٪ مشاهدات در فاصله میانگین، مثبت و منفی سه انحراف معیار قرار دارد.

نکته ۱: واضح است که داده های رتبه ای دارای توزیع نرمال نمی باشند.

نکته ۲: وقتی داده ها کمی هستند و تعداد نمونه نیز کم است تشخیص نرمال بودن داده ها توسط آزمون **کولموگروف - اسمیرنوف** مشکل خواهد شد.

۱۶- **آزمون پارامتریک:** آزمون های پارامتریک، آزمون های هستند که توان آماری بالا و قدرت پرداختن به داده های جمع آوری شده در طرح های پیچیده را دارند. در این آزمون ها داده ها توزیع نرمال دارند. (مانند آزمون تی).

۱۷- **آزمون های غیر پارامتری:** آزمون هایی می باشند که داده ها توزیع غیر نرمال داشته و در مقایسه با آزمون های پارامتری از توان تشخیصی کمتری برخوردارند. (مانند آزمون من - ویتنی و آزمون کروسکال و والیس)

نکته ۳: اگر جامعه نرمال باشد از آزمون های پارامتریک و چنانچه غیر نرمال باشد از آزمون های غیر پارامتری استفاده می نمائیم.

نکته ۴: اگر نمونه بزرگ باشد، طبق قضیه حد مرکزی حتی اگر جامعه نرمال نباشد می توان از آزمون های پارامتریک استفاده نمود.

حال به کمک جدول زیر براحتی می توانید یکی از ۲۴ آزمون مورد نظر خود را انتخاب کنید:

داده های کیفی اسمی Categorical	داده رتبه ای و یا داده کمی غیر نرمال	داده کمی و دارای توزیع نرمال	هدف
آزمون نسبت	آزمون میانه	آزمون میانگین و انحراف معیار	توصیف یک گروه
آزمون خی - دو یا آزمون دو جمله ای	آزمون ویلکاکسون	آزمون یک نمونه ای	مقایسه یک گروه با یک مقدار فرضی
آزمون دقیق فیشر (آزمون خی دو برای نمونه های بزرگ)	آزمون من - ویتنی	آزمون برای نمونه های مستقل	مقایسه دو گروه مستقل
آزمون مک - نار	آزمون کروسکال	آزمون زوجی	مقایسه دو گروه وابسته
آزمون خی - دو	آزمون والیس	آزمون آنالیز واریانس یک راهه	مقایسه سه گروه یا بیشتر (مستقل)
آزمون کوکران	آزمون فریدمن	آزمون آنالیز واریانس با اندازه های مکرر	مقایسه سه گروه یا بیشتر (وابسته)
آزمون ضریب توافق	آزمون ضریب همبستگی اسپرمن	آزمون ضریب همبستگی پیرسون	اندازه همبستگی بین دو متغیر
آزمون رگرسیون لجستیک	آزمون رگرسیون نا پارامتریک	آزمون رگرسیون ساده یا غیر خطی	پیش بینی یک متغیر بر اساس یک یا چند متغیر

در رویکردی دیگر بر مبنای تعداد متغیر، تعداد گروه و نرمال بودن جامعه نیز می توان به الگوریتم آزمون آماری مورد نظر دست یافت:

یک متغیر:

انتخاب آزمون آماری برای یک متغیر	یک متغیر در یک گروه	یک متغیر در دو گروه	یک متغیر در سه گروه یا بیشتر
متغیر نرمال	آزمون میانگین و انحراف معیار	آزمون تی	آزمون آنالیز واریانس ANOVA
متغیر غیر نرمال	آزمون نسبت (دو جمله ای)	آزمون خی - دو	آزمون ناپارامتریک

دو متغیر:

انتخاب آزمون آماری برای دو متغیر	هر دو متغیر پیوسته هستند	یک متغیر پیوسته و دیگری گسسته است	هر دو متغیر مقوله ای هستند
	آزمون همبستگی	آزمون آنالیز واریانس ANOVA	آزمون خی - دو

سه متغیر و بیشتر:

انتخاب آزمون آماری برای سه متغیر و بیشتر	یک گروه	دو گروه و بیشتر
	آنالیز کواریانس	تحلیل ممیزی
	آنالیز واریانس با اندازه های مکرر	آنالیز واریانس چند متغیره
	تحلیل عاملی و رگرسیون چند گانه	

قابل ذکر است قبل از ورود به الگوریتم انتخاب آزمون آماری بهتر است به سوالات زیر پاسخ دهیم:

- ۱- آیا اختلافی بین میانگین (نسبت) یک ویژگی در دو یا چند گروه وجود دارد؟
- ۲- آیا دو متغیر ارتباط دارند؟
- ۳- چگونه می توان یک متغیر را با استفاده از متغیر های دیگر پیش بینی کرد؟
- ۴- چه چیزی می توان با استفاده از نمونه در مورد جامعه گفت؟