بسم الله الرحمن الرحیم

آزمون های ناپارامتری

استاد:خانم دکتر فاطمه حسینی

دانشجویان:

فاطمه جهادی

سحر خیرخواه

آرتا روحی

دانشگاه سمنان

پاییز ۹۸

فهرست

چکیده

در حوزه تجزیه و تحلیل آماری داده‌ها[1]، توزیع جامعه آماری که نمونه از آن گرفته شده، مهم است زیرا هر چه اطلاعات بیشتر در زمینه رفتار داده‌ها و شکل پراکندگی و توزیع آن‌ها وجود داشته باشد، نتایج قابل اعتمادتر و دقیق‌تر خواهند بود. در مقابل، وجود اطلاعات کم از توزیع جامعه آماری مربوط به نمونه، باعث کاهش اعتماد به نتایج حاصل از روش‌های معمول (پارامتری[2]) آماری می‌شود. بنابراین در این حالت مجبور به استفاده از روش‌های ناپارامتری[3] هستیم که برای اجرای آن‌ها فرضیاتی در مورد توزیع داده‌ها وجود ندارد. بنابراین اگر توزیع جامعه آماری نامشخص باشد و از طرفی حجم نمونه نیز کوچک باشد به‌طوری که نتوان از قضیه حد مرکزی[4] برای تعیین توزیع حدی یا مجانبی جامعه آماری، استفاده کرد، از تحلیل‌های ناپارامتری استفاده می‌شود، زیرا در این حالت کارآمدتر از روش‌های پارامتری هستند. به این ترتیب در زمانی که توزیع جامعه مشخص نباشد و یا حجم نمونه کم باشد، روش‌ها و آزمون‌های ناپارامتری نسبت به روش‌ها و آزمون‌های پارامتری از توان آزمون بیشتری برخوردارند و نسبت به آن‌ها ارجح هستند.

مقدمه

آزمون‌های آماری بر اساس ویژگی‌های داده‌ها یا متغیرها به دو دسته آزمون های پارامتری و ناپارامتری تقسیم می‌شوند. در جریان رشد روش‌های آماری مدرن، اولین تکنیک‌های استنباطی که پدید آمدند، تکنیک هایی بودند که برای جمعیتی که نمرات از آن برگرفته می‌شود مفروضات فراوانی قائل بودند. از آن‌جا که مقادیر جمعیت «پارامتر» اند، این تکنیک‌های آماری را پارامتری می‌خوانند. این دسته از تکنیک‌ها به نتیجه‌گیری‌هایی می‌انجامد که با قید و شرط همراه‌اند. به عبارت ساده‌تر می‌توان گفت داده‌های پارامتری به نمونه‌ای گفته می‌شود که از توزیع جامعه آماری آن مطلع هستیم. معمولا این توزیع آماری برای داده‌های کمی، نرمال یک یا چند متغیره در نظر گرفته می‌شود. در این حالت از آزمون‌های آماری پارامتری مثل آزمونT ، آزمون F و یا آزمون Z استفاده می‌کنیم. همچنین برای اندازه‌گیری میزان همبستگی بین متغیرهای دو یا چند بعدی نیز از ضریب همبستگی پیرسون استفاده خواهیم کرد. اما در بسیاری از مسائل کاربردی در دنیای واقعی شرایطی وجود دارد که امکان استفاده از روش‌های متداول مبتنی بر توزیع نرمال (روش‌های پارامتری) وجود ندارد. در صورت برقرار نبودن پیش‌فرض‌های آزمون‌های پارامتری، بدون نگرانی برای از دست رفتن کارایی و حتی با کارایی بالاتر، می‌بایست از آزمون‌های ناپارامتری استفاده کرد . به همین دلیل در سال‌های اخیر تعداد زیادی تکنیک استنباطی ابداع شده است که مفروضات سختی را درباره پارامترها پیش نمی‌کشند. این

[1] Statistical data analysis
[2] Parametric
[3] Nonparametric
[4] Central limit theorem

تکنیک‌های جدیدتر، ناپارامتری هستند به طوری که نتیجه‌گیری‌ها در آن مستلزم هیچ قید و شرطی نیست. همچنین لازم به ذکر است که برای داده‌های کیفی، توزیع دوجمله‌ای[5] یا چند جمله‌ای[6] در نظر گرفته می‌شود.

با توجه به خصوصیات توزیع نرمال و خوش‌تعریف بودن آن از لحاظ ریاضی[1]، توزیع نرمال مبنای بسیاری از روش‌های آماری می‌باشد. همان‌طور که در بحث آمار توصیفی نیز بدان اشاره می شود، نرمال بودن توزیع داده‌ها برای متغیرهای کمی سبب انتخاب میانگین به عنوان پارامتر مورد بررسی در داده ها می‌شود. که استفاده از میانگین سبب سهولت استنتاج‌های ریاضی در روش‌های آماری و خواص بهینه در آن‌ها نظیر افزایش حساسیت آزمون ها می‌گردد.

به هر حال، همیشه نرمال بودن توزیع داده‌ها را نمی‌توان به عنوان یک پیش‌فرض قابل قبول پذیرفت و در نتیجه امکان استفاده از روش‌های آماری مبتنی بر توزیع نرمال وجود ندارد و یا در برخی از مطالعات پیامدها به صورت رتبه‌ای و اسمی ارزیابی می شوند. اما روش‌های مزبور همه موقعیت‌های مطالعاتی را پوشش نمی‌دهند.

در دو حالت فوق نیاز به روش‌هایی است که در چنین موقعیت‌هایی بتوانند اختلاف‌ها و یا رابطه‌های مورد نظر در اهداف مطالعات را بررسی نمایند. بنابراین در این روش‌ها دیگر نیازی نیست نگران نرمال بودن یا نبودن توزیع داده‌ها و یا شاخص‌های مورد نیاز در آزمون بود . به‌عبارت دقیق‌تر به روش‌هایی نیاز است فارغ از توزیع داده‌ها اهداف مورد نظر را بررسی نمایند. به این روش‌ها اصطلاحا روش های آزاد توزیع[7] و یا ناپارامتری نیز گفته می شود. آزمون های کلاسیک آماری با در نظر گرفتن توزیع آمار مشخص برای داده‌ها، پارامتر مشخصی از این توزیع نظیر میانگین را مورد آزمون قرار می‌دهند که به این دلیل به آن‌ها آزمون های پارامتری گفته می شود. در مقابل آزمون های که نیاز به مشخص بودن توزیع احتمالی داده‌ها ندارند به دلیل در تقابل قرار گرفتن با آزمون‌های پارامتری به آن‌ها آزمون‌های ناپارامتری گفته می شو د. عبارت های ناپارامتری و آزاد توزیع که به‌طور معادل برای این شکل از آزمون‌ها به کار می‌روند یکسان نیستند. آزمون ناپارامتری براساس نام آن، قاعدتاً باید آزمونی باشد که فرضیه‌ای درباره مقدار پارامتریک توزیع احتمالی در نظر نمی‌گیرد ولی یک آزمون آزاد توزیع به آزمونی گفته می‌شود که درباره شکل توزیعی داد ه‌ها فرضی در نظر نمی‌گیرد و ممکن است روی پارامتر مشخصی در آن توزیع، آزمون ر ا انجام دهد. به هر حال برای متغیرهای کمی، با وجود این تفاوت در نام، آزمون های ناپارامتری یا آزاد توزیع، زمانی استفاده می‌شوند که پیش‌فرض اساسی نرمالیتی توزیع داده‌ها به هر دلیلی نظیر ماهیت داده ها، طراحی پیچیده مطالعه یا طرح نمونه گیری نامناسب و ... برقرار نباشد.کارایی نسبی آزمون‌های ناپارامتری در مقایسه با آزمون های معادل پارامتری آزمون های پارامتری با در نظرگرفتن پیش فرض نرمالیتی داد ه‌ها انجام می‌شوند و چون با در نظرگرفتن این پیش فرض آزمون متمرکز روی پارامتری خاص نظیر میانگین می‌گردد، به آن عنوان پارامتری داده می‌شود. معادل ناپارامتری این آزمون ها

[5] Binomial distribution
[6] Multinomial distribution
[7] Distribution-Free

یعنی‌آزمون‌هایی که بدون در نظر گرفتن پیش فرض در نرمـالیتی داده هـا انجـام می‌شوند. زمانی که شـرایط آزمـون‌هـای پـارامتری برقـرار باشند، آزمون های پارامتری به دلیل استفاده از اطلاع توزیـع نرمال در داده‌ها در مقایسه با آزمون‌های ناپارامتری تـوان بیشتری دارند ولی زمانی که شرایط و یا پـیش‌فـرض‌هـای آزمون‌های پارامتری برقرار نباشد، توان وکارایی آزمون‌های ناپارامتری بیشتر هم باشد[2].

آزمون هـای ناپـارامتری از لحـاظ اجـرا و تفـسیر سـاده‌تـر از آزمون های پارامتری معـادل‌شـان نیـستند حتـی یکـی از نقدهایی که به آزمون های پارامتری وارد می‌شود این اسـت که این آزمون ها جامعه یا فرآیند مشاهده شده را بیش ازحد ساده می‌کنند و در واقع آزمون‌های پارامتری نه به دلیل این که بهینه‌تر هستند بلکه به دلیل سهولت به کارگیری آن‌هـا بـه‌صورت متداول‌تری استفاده می‌شوند. یکی از مشکلاتی که در استفاده از آزمون‌هـای ناپـارامتری وجـود دارد، ذهنیـت منفی کاربران غیر متخصص می‌باشد. تصور عامـه بـر ایـن است که آزمون های ناپارامتری در مقایـسه بـا آزمـون‌هـای پارامتری متناظر آن‌ها خیلی ضـعیف‌تـر عمـل مـی‌کننـد و تلاش آن‌ها بـر ایـن اسـت کـه بـه هـر ترتیـب ممکـن از آزمون‌های پارامتری استفاده نمایند. یکی از این روش‌ها اعمال تبدیل روی داده‌هاست. مشهور ترین تبدیل‌ها تبدیل باکس-کاکس [3] است که تنها روی داده‌های اکیدا مثبت انجام می‌شود. در واقع این تبدیل نزدیک‌کننده توزیع متغیرها تا احد امکان به توزیع نرمال چند متغیره می باشد. حتی در برخی شرایط کارشناس آمار حاضر است به جای بکارگیری روش‌های ناپارامتری دو مرتبه از تبدیل باکس- کاکس استفاده کند، تا بتواند باید آزمون های پارامتری را به کـار ببرد. ولی به هر حال موقعیت‌هایی وجود دارنـد کـه ممکـن است تبدیل روی داده‌ها سـبب نرمـالیتی آن‌هـا نگـردد. در چنین موقعیتی باید از آزمـون ناپـارامتری مناسـب اسـتفاده شود.

مسئله ضعیف‌تر عمل نمودن آزمون ناپارامتری به حساسیت یا توان آزمـون آن هـا بـر مـی‌گـردد و اغلـب یـک آزمـون ناپارامتری به جای اصل داده‌ها از رتبه‌های متنـاظر بـا آن‌هـا استفاده می کند و این موضوع سبب کاهش اطلاع در داده‌ها و در نتیجه کاهش توان آزمون می گردد. حال مسئله اساسی این است که این موضوع تا چه اندازه می‌تواند موثر باشد و آیا در همه شرایط موضوع کاهش اطلاع رخ می دهـد و یـا این‌که در شرایط خاص باید این موضوع را مد نظـر قـرار داد. زمانی که شرایط یک آزمون پارامتری برقرار باشد و به اشتباه از یک آزمون ناپارامتری معادل اسـتفاده گـردد، ایـن موضوع سبب کاهش توان تصمیم‌گیری می‌شود. به‌عبـارت دیگـر، در زمـان برقـراری نرمـالیتی داده‌هـا، یـک آزمـون ناپارامتری در مقایسه با آزمون‌های پارامتری تـوان کمتـری دارند.

بحث نرمال بودن توزیع داد هها صرفا برای متغیرهـای کمی (فاصله‌ای و نسبتی) مطـرح اسـت و در نتیجـه آزمون‌های پارامتری که به نرمالیتی توزیع داده‌ها نیـاز دارنـد محـدود بـه ایـن نـوع از متغیرهـای هـستند. درحالی که علاوه بر این نـوع از متغیرهـا، متغیرهـای اسمی و رتبه ای هم وجود دارند که در این موارد نیـز نیاز است رو ش‌های آمـاری ناپـارامتری بـه کـار رود. بـه عبـارت دقیـق تر، آزمـون هـای ناپـارامتری شـامل آزمـون هـایی مـی‌شـود کـه بـرای متغیرهـای کمـی غیرنرمال، رتبه‌ای و اسمی به کار می رود و پیش فرضی را در مورد توزیع زیربنایی داده ها در نظر نمی‌گیرد.

مشاهده می‌شود برخی از کاربران به اشتباه آزمون های ناپارامتری را صرفاً مخصوص داده‌های با مقیاس اسمی یا رسته ای می دانند و یا اینکه برخی از کاربران آن را به کار می‌برند چون حجم نمونه مطالعه آن‌ها کوچک است. در حالی که روش‌های ناپارامتری برای انواع مقیاس‌های متغیرها قابل به کارگیری است و ازطرف دیگر ملاک استفاده از این آزمون ها حجم نمونه نیست و پیش فرض‌های زیربنایی برای انتخاب آزمون تعیین‌کننده هستند. به‌عبارت دیگر، اگر حجم نمونه کوچک باشد ولی توزیع داده ها برای یک متغیرکمی نرمال باشد، باز هم می توان از آزمون ها پارامتری برای بررسی رابطه یا اختلاف استفاده نمود.

همچنین ملاحظه می شود برخی از کاربران چون حجم نمونه در مطالعه آن‌ها بیشتر از عدد ثابتی (معمولاً ۳۰) می باشد به اشتباه با استناد به قضیه حد مرکزی [4]. آزمون های پارامتری را به‌جای آزمون های ناپارمتری برای متغیرهای کمی غیرنرمال به‌کار می برند. در حالی که در قضیه حد مرکزی بیان می شود که در صورت داشتن نمونه بزرگ، توزیع میانگین به نرمال میل می‌کند [5] که در این مورد باید به دو نکته دقت کرد؛ اول این که عدد مشخص‌کننده نمونه بزرگ همیشه عدد ۳۰ نیست و در صورت کمی بودن متغیر و پراکندگی کم و نبودن داده های پرت در اصل داده‌ها عدد ۳۰ می تواند قابل قبول باشد و در شرایطی غیر از این، عدد خیلی بزرگتر برای برقراری شرایط این قضیه نیاز است. دوم این که شرایط استفاده از این قضیه برای آزمون‌هایی است که شاخص آزمون آن‌ها ر وی میانگین‌ها تعریف شده (و برای آزمون‌هایی نظیر t باشد) نظیر آزمون های تحلیل واریانس و همبستگی قابل به کارگیری نیست. بنابراین در شر ایط حجم نمونه بالا نمی‌توان معادل پارامتری را برای همه آزمون‌های ناپارامتری به‌کار برد و فقط برای مقایسه میانگین‌ها با یک عدد ثابت، در دو گروه مستقل و وابسته در شرایط حجم نمونه بالا و نبودن داده‌های پرت موثر، به‌ترتیب استفاده از آزمون های t یک نمونه‌ای، دونمونه‌ای مستقل و وابسته ممکن است.

البته اگر حجم نمونه بزرگ باشد، در اکثر موارد، نتایج حاصل از آزمون‌های پارامتری و ناپارامتری با یکدیگر همخوانی دارند.

انواع داده‌ها

به منظور ادامه این بخش لازم است که انواع داده ها را تعریف نماییم.

داده‌های 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[۸]

ب: داده های نسبتی[۹]

[8] Interval data

[9] Ratio data

- **داده‌های فاصله ای**: به عنوان مثال داده‌هایی 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 کمی‌اند 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 فاصله‌ای اند.
- **داده‌های نسبتی**: داده‌های نسبتی داده‌هایی هستند که با عدد نوشته می شوند اما صفر آن‌ها واقعی است. اکثریت داده‌های کمی این گونه‌اند و حقیقتاً دارای صفر هستند. به عنوان مثال داده‌هایی که متغیر طول پاره‌خط بر حسب سانتی‌متر را توصیف می‌کنند عبارتند از: ۲۰، ۱۵، ۳۵، ۸ و ۲۳، چون این داده‌ها عدد هستند پس داده های ما کمی‌اند و چون صفر در این‌جا واقعاً وجود دارد این داده نسبتی تلقی می‌شوند.
- **داده‌های 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[۱۰]
  - ب: داده های رتبه ای[۱۱]

داده های رتبه ای[۱۲]: مانند کیفیت درسی یک دانش آموز (ضعیف، متوسط و قوی) و یا رتبه بندی هتل ها (یک ستاره، دو ستاره و ...)

از آنجایی که در بیشتر روش‌های ناپارامتری به جای داده‌ها، ترتیب آن‌ها به کار گرفته می‌شود، بهتر است با مفهوم رتبه بیشتر آشنا شویم. در ادامه به معرفی رتبه و کاربردهای آن در آمار ناپارامتری می‌پردازیم.

استفاده از رتبه‌ها به جای مقدارها، یکی از ویژگی‌های روش‌های ناپارامتری است. برای مثال همانطور که دیده‌اید ضریب همبستگی اسپیرمن یک روش ناپارامتری برای اندازه‌گیری همبستگی بین مقدارها است. برای محاسبه ضریب همبستگی اسپیرمن به جای استفاده از مقدارها، رتبه‌هایشان ملاک قرار می‌گیرد و ضریب همبستگی عادی (پیرسون) به جای مقدارها از روی رتبه‌ها محاسبه می‌شود.

برای ایجاد رتبه‌ها کافی است که آن‌ها را به ترتیب چیده و از کمترین تا بیشترین مقدار، برچسب‌های از ۱ تا n را نسبت دهیم. این برچسب‌ها رتبه را نشان می‌دهد.

داده‌های اسمی[۱۳] که مربوط به متغیر یا خواص کیفی مانند جنس یا گروه خونی است و بیانگر عضویت در یک گروها[۱۴] خاص می باشد. (داده مقوله ای)

10 Nominal data
11 Ordinal data
12 Ordinal (Rank)
13 Nominal
14 Category

بسته به نوع داده‌ها و شرایط مختلف آزمون‌های ناپارامتری متفاوتی وجود دارند که در ادامه به بررسی آن‌ها می‌پردازیم.

آزمون علامت[15]

این آزمون هنگامی به کار برده می‌شود که نمونه های جفت، مورد نظر باشد (مثل زن و شوهر و یا خانه های فرد و زوج و . . .). زیرا در این آزمون یافته‌ها به صورت جفت جفت بررسی می‌شوند. به عبارت دیگر برای تشخیص اختلاف بین زوج مشاهدات مناسب است. اندازه مقادیر در آن بی اثر است و فقط علامت مثبت و منفی و یا در واقع جهت پاسخ‌ها و یا بیشتر و کمتر بودن پاسخ‌های جفت‌های گروه مورد تحقیق (نمونه آماری) در نظر گرفته می‌شود. به بیان دیگر اگر متغیرهای X و Y عددی و از نوع پیوسته باشند، می‌توان آزمون علامت را به صورت یک فرض آماری در نظر گرفت که نشان می‌دهد مقدار میانه اختلاف مربوط به این دو متغیر برابر با صفر است یا خیر. از طرفی از این آزمون برای قضاوت در مورد میانه یک نمونه تصادفی نیز می‌توان استفاده کرد. به این ترتیب می‌توانیم تصمیم بگیریم که آیا میانه اعداد با مقدار مشخصی برابر است یا خیر

هنگامی که ارزشیابی متغیر مورد مطالعه با روش‌های عادی قابل اندازه گیری نباشد و قضاوت در مورد نمونه های آماری (که به صورت جفت ها هستند) فقط با علامت بیشتر $(+)$ و کمتر $(-)$ مورد نظر باشد ، از این آزمون می‌توان استفاده کرد. شکل توزیع می‌تواند نرمال و یا غیر نرمال باشد و یا از یک جامعه و یا دو جامعه باشند (مستقل و یا وابسته). توزیع باید پیوسته باشد. این آزمون فقط تفاوت‌های زوج‌ها را مورد بررسی قرار می‌دهد و در صورت مساوی بودن نظرات هر زوج (مشابه بودن) آن‌ها را از آزمون حذف می‌کند. چون مقادیر در این آزمون نقشی ندارند، شدت و ضعف و اندازه بیشتر یا کمتر بودن نظرات پاسخگویان (جفت‌ها) در این آزمون بی‌اثر است و در واقع نقص این آزمون حساب می‌شود.

نکته: همان‌طور که گفته شد، آماره آزمون علامت براساس روابط بزرگتر یا کوچکتر و یا مساوی شکل می‌گیرد. ولی اگر رابطه بین زوج مشاهدات را بتوان به صورت عددی (مانند آزمون T) نشان داد دقت و توان آزمون بیشتر از توان آزمون علامت خواهد بود.

شکل اولیه این آزمون برای اولین بار توسط جان آربوتنات[16] در سال ۱۷۱۰ مطرح شد[6]. او برای بررسی درستی یا نادرستی نظریه «پسران متولد شده بیشتر از دختران متولد شده است» که در آن زمان مطرح بود، از این آزمون استفاده نمود. برای انجام این آزمون او به بررسی تعداد تولدها در شهر لندن از تاریخ ۱۶۲۹ تا ۱۷۱۰ پرداخت. او در مقاله‌ای که در آن دوره به چاپ رساند اولین بار واژه «آزمون بامعنایی[17]» را به کار برد. پس از آربونات، نیکولاس برنولی[18] بین سال‌های

[15] Sign test
[16] John Arbuthnot
[17] Significance Tests
[18] Nicholas Bernoulli

۱۷۱۰ تا ۱۷۱۳ تحلیل‌ها و نتایج آماری او را بررسی کرده و نتیجه گرفت که بیشتر تغییرات موالید در سال مربوط به تولد پسران است که بوسیله توزیع دو جمله‌ای توصیف می‌شود. به نظر می‌رسد اولین بار در بررسی‌های «برنولی» است که از توزیع دوجمله‌ای برای برازش داده‌های واقعی استفاده شده است. البته تفکرات آربوتنات علاوه بر برنولی، کونور[19] و اسپرنت[20] نیز را علاقمند نمود. آن‌ها نیز با آزمون فرض آماری نظریه او را که نشان می‌داد نرخ تولد پسران با دختران یکسان نیست، مورد بررسی قرار دادند.

شکل ۱: سمت راست جان آربوتنات و سمت چپ نیکولاس برنولی

به منظور استفاده از آزمون علامت باید زوج‌های نمونه، تصادفی باشند. به این معنی که هر یک از مولفه‌های تشکیل دهنده زوج مرتب (X, Y) باید به صورت تصادفی از جامعه گرفته شده باشند و از طرفی با یکدیگر مرتبط باشند تا مفهوم زوج وجود داشته باشد. بنابراین با درنظر گرفتن مشاهده‌های نمونه تصادفی دو بعدی به صورت

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_{n^*}, Y_{n^*})$$

به‌طوری‌که n^* ، تعداد زوج‌های مشاهده است. برای زوج کردن مشاهده‌ها باید یک مبنای طبیعی وجود داشته باشد، در غیر این‌صورت X ها و Y ها مستقل هستند و آزمون من-ویتنی مناسب تر خواهد بود. درون هر زوج مشاهده مقایسه ای انجام می‌شود و زوج‌ها به‌صورت زیر رده‌بندی می‌شوند.

- اگر $X_i < Y_i$ ها باشد، زوج مثبت $(+)$ در نظر گرفته می شود.
- اگر $X_i > Y_i$ ها باشد، زوج منفی $(-)$ در نظر گرفته می شود

[19] Conover
[20] Sprent

- در صورتی که $X_i = Y_i$ زوج را صفر یا هم‌مرتبه در نظر گرفته می شود. البته برخی در این حالت می‌گویند (به اصطلاح) گره[21] اتفاق افتاده است و در این آزمون نمونه‌هایی که گره‌دار هستند حذف کرده و از مابقی نمونه برای انجام آزمون استفاده می‌کنند.

پذیره‌ها

- ✓ متغیرهای تصادفی دو متغیره $(X_i, Y_i),\ \ i = 1, \dots, n^*$ دو به دو مستقل‌اند.
- ✓ مقیاس اندازه‌گیری در هر زوج، حداقل ترتیبی است، یعنی می توان تعیین کرد که هر زوج (X_i, Y_i) مثبت یا منفی و یا صفر است.
- ✓ زوج‌های (X_i, Y_i) به‌صورت درونی سازگارند، به این صورت که اگر برای یک زوج (X_i, Y_i) ، $P(+) > P(-)$ ، آن‌گاه برای همه‌ی زوج‌ها $P(+) > P(-)$ است. این مورد برای دو حالت $P(+) < P(-)$ و $P(+) = P(-)$ نیز صادق است.

فرض‌ها

۱) آزمون دوطرفه (دو دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: P(+) = P(-) \\ \mathrm{H}_1: P(+) \neq P(-) \end{cases}$$

۲) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: P(+) \leq P(-) \\ \mathrm{H}_1: P(+) > P(-) \end{cases}$$

۳) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: P(+) \geq P(-) \\ \mathrm{H}_1: P(+) < P(-) \end{cases}$$

آزمون علامت برای فرض‌های بالا ناریب و سازگار است[7]. البته می‌توان آزمون‌هایی هم‌ارز آزمون‌های بالا به ازای همه i ها به صورت زیر نوشت:

۱) آزمون دوطرفه (دو دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: E(X_i) = E(Y_i) \\ \mathrm{H}_1: E(X_i) \neq E(Y_i) \end{cases}$$

۲) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: E(X_i) \leq E(Y_i) \\ \mathrm{H}_1: E(X_i) > E(Y_i) \end{cases}$$

[21] Kont

۳) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: E(X_i) \geq E(Y_i) \\ \mathrm{H}_1: E(X_i) < E(Y_i) \end{cases}$$

آزمون‌های بالا وضعیت میانگین‌های دو متغیر نسبت به یکدیگر مورد بررسی قرار می‌دهد. لازم به ذکر است که آزمون‌های بالا اریب و ناسازگار است مگر این‌که در مورد توزیع (X_i, Y_i) پذیره‌های بیشتری در نظر گرفته شود.

آزمون‌های فوق می‌تواند براساس میانه، یعنی به صورت زیر نیز بیان شود.

۱) آزمون دوطرفه (دو دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: Median(X_i) = Median(Y_i) \\ \mathrm{H}_1: Median(X_i) \neq Median(Y_i) \end{cases}$$

۲) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: Median(X_i) \leq Median(Y_i) \\ \mathrm{H}_1: Median(X_i) > Median(Y_i) \end{cases}$$

۳) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: Median(X_i) \geq Median(Y_i) \\ \mathrm{H}_1: Median(X_i) < Median(Y_i) \end{cases}$$

آماره آزمون

فرض کنید T آماره آزمون، برابر تعداد زوج‌های مثبت $(+)$ باشد، یعنی T برابر تعداد زوج‌هایی است که X_i آن‌ها از Y_i شان کوچکتر است. به عبارت دیگر

$$T = Sum\ (+) = \sum_{i=1}^{n^*} P\ (X_i < Y_i)$$

قاعده تصمیم‌گیری

با حذف زوج‌های هم‌مرتبه و تعریف n به‌صورت زیر

$$n = \sum_{i=1}^{n^*} P\ (X_i < Y_i) + \sum_{i=1}^{n^*} P\ (X_i > Y_i)$$

و با تعریف α به عنوان سطح معنی‌دار و همچنین با توجه به این واقعیت آماره آزمون از توزیع دوجمله‌ای به صورت زیر تبعیت می‌کند. یعنی

$$T \sim Binomial(n, 0.5)$$

به سادگی آزمون انجام می‌شود و قاعده تصمیم‌گیری سه آزمون به صورت زیر خواهد بود.

۱) فرض H_0 رد می‌شود هر گاه

$$T \geq n-t \qquad or \qquad T \leq t$$

۲) فرض H_0 رد می‌شود هر گاه

$$T \geq n-t$$

۱) فرض H_0 رد می‌شود هر گاه

$$T \leq t$$

به طوری که در آن

$$t = \frac{1}{2}(n + z_{\alpha/2}\sqrt{n})$$

و $z_{\alpha/2}$ عکس تابع توزیع نرمال استاندارد (جدول نرمال استاندارد) است.

نظریه

پیشامد $X_i < Y_i$ را می‌توان به صورت $Y_i - X_i > 0$ نیز نمایش داد. و این گزاره نشان‌دهنده این است که اختلاف X_i ها و Y_i ها مثبت است. همچنین $(-)$ و 0 نیز نشان می‌دهد که $Y_i - X_i$ به ترتیب منفی و یا صفر اند. بنابراین آزمون علامت، آزمونی برای مقایسه احتمال یک اختلاف مثبت با احتمال یک اختلاف منفی است. در آزمون دوجمله‌ای این احتمال‌ها به ترتیب «رده اول» و «رده دوم» نامیده می‌شود. با حذف هم‌مرتبه ها داریم

$$P(+) + P(-) = 1$$

بنابراین فرض

$$\mathrm{H}_0: P(+) = P(-)$$

هم‌ارز با فرض

$$\mathrm{H}_0: P(+) = \frac{1}{2}$$

است. بنابراین از همان روش آزمون دوجمله‌ای استفاده می‌شود ، اگر چه تقارن

$$p^* = \frac{1}{2} = 1 - p^*$$

تسهیل اندکی نتیجه می‌شود.

آزمون ویلکاکسون

بیشتر روش‌های آمار ناپارامتری برپایه رتبه‌ها و توزیع آن‌ها است. توسعه آمار ناپارامتری و روش‌های آن توسط فرانک ویلکاکسون[22] شیمی‌دان و آمارشناس آمریکای در سال ۱۹۴۵ طی مقاله‌ای که منتشر کرد رخ داد . این مقاله و مقاله‌ای که هنری مَن[23] و دونالد ویتنی[24] در سال ۱۹۴۷ منتشر کردند، باعث شد که توجه به آمار ناپارامتری و انجام تحلیل بر روی جوامع آماری که توزیع احتمالی آن‌ها نرمال یا گاوسی نیست، گسترش پیدا کند. از جمله این روش‌های می‌توان به تحلیل‌های ناپارامتری، آزمون ویلکاکسون جمعی رتبه‌ای[25] و آزمون ویلکاکسون رتبه علامت‌دار[26] اشاره کرد. در این‌جا به به بررسی آزمون رتبه علامت‌دار می‌پردازیم.

شکل ۲: فرانک ویلکاکسون

آزمون ویلکاکسون رتبه علامت‌دار به منظور انجام بررسی دو نمونه وابسته یا انطباق بین دو نمونه به کار گرفته می‌شود. این آزمون را می‌توان مشابه آزمون t برای گروه‌های وابسته برای میانگین جامعه غیرنرمال، در نظر گرفت. هر چند این آزمون، ناپارامتری بوده و به توزیع داده‌ها بستگی ندارد ولی برای انجام این آزمون باید فرضیات زیر مورد بررسی قرار گیرد.

- داده‌ها باید به صورت زوجی و از یک جامعه گرفته شده باشند.
- هر مولفه از زوج‌ها به صورت تصادفی انتخاب شده و مستقل از نمونه‌های دیگر باشند.

22 Frank Wilcoxon
23 Henry Mann
24 Donald Ransom Whitney
25 Wilcoxon Sum-Rank Test
26 Wilcoxon Signed-Rank Test

- نوع یا مقیاس داده‌ها باید به صورت فاصله‌ای یا نسبتی باشند تا بتوان تفاوت بین مقدار آن‌ها را بدست آورد و این تفاضل‌ها را رتبه‌بندی کرد. به این ترتیب باید مطمئن شد که تفاضل مقدارهای زوج‌ها به صورت مقیاس ترتیبی باشند.

داده‌ها در این آزمون به صورت

$$(x_1, y_1), (x_2, y_2), \dots, (x_{n^*}, y_{n^*})$$

از متغیرهای تصادفی

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_{n^*}, Y_{n^*})$$

هستند. حال باید قدر مطلق تفاضل‌های زوج‌ها به صورت زیر برای همه زوج‌ها محاسبه شود.

$$D_i = |Y_i - X_i|, \qquad i = 1,2, \dots, n^*$$

از بررسی زوج‌هایی که اختلاف آن‌ها صفر است (یعنی $Y_i = X_i$ یا $D_i = 0$) صرف نظر می‌کنیم. فرض کنید تعداد زوج‌های باقی‌مانده برابر n باشد، بنابراین $n \leq n^*$. به هر یک از n زوج براساس اندازه نسبی قدر مطلق تفاضل‌ها رتبه‌های 1 تا n را به صورت زیر تخصیص می‌دهیم.

- رتبه 1 به زوج (X_i, Y_i) با کوچکترین D_i ها تخصیص داده می‌شود.
- رتبه 2 به زوجی تخصیص داده می‌شود که دارای دومین کوچکترین D_i ها باشد.

  و به همین ترتیب ادامه می‌یابد تا
- رتبه n به زوجی تخصیص می‌یابد که دارای بزرگترین D_iها باشد.

نکته: اگر چندین زوج دارای D_i های مساوی باشند، میانگین رتبه‌هایی را به آن‌ها تخصیص می‌دهیم که در صورت مساوی نبودن به آن‌ها تعلق می‌گرفت. به طور مثال اگر رتبه‌های 3 ، 4 ،5 و 6 به چهار زوج تعلق بگیرد، اما به دلیل این‌که چهار قدر مطلق تفاضل مربوط به آن زوج‌ها دقیقا با یکدیگر مساوی‌اند نتوانیم تعیین کنیم که کدام رتبه به کدام یک از چهار قدر مطلق باید اختصاص داده شود، آن‌گاه به هر زوج میانگین رتبه‌ها یعنی

$$\frac{3+4+5+6}{4} = 4.5$$

اختصاص داده می‌شود.

پذیره‌ها

- ✓ توزیع هر یک از D_i باید متقارن باشد.
- ✓ D_i ها دو به دو مستقل هستند.
- ✓ D_i ها همه دارای یک میانه‌اند.
- ✓ مقیاس اندازه‌گیری D_i ها حداقل، فاصله است.

فرض‌ها

۱) آزمون دوطرفه (دو دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: Median(D_i) = 0 \\ \mathrm{H}_1: Median(D_i) \neq 0 \end{cases}$$

۲) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: Median(D_i) \leq 0 \\ \mathrm{H}_1: Median(D_i) > 0 \end{cases}$$

۳) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: Median(D_i) \geq 0 \\ \mathrm{H}_1: Median(D_i) < 0 \end{cases}$$

البته اگر مدل را کمی تغییر دهیم ، این فرض‌ها را می‌توان به صورت قابل ملاحظه وسعت داد. این تغییر عبارت است از افزودن پذیره

✓ « (X_i, Y_i) به ازای $i = 1, \dots, n$ نمونه‌ای تصادفی و ودو متغیره را تشکیل دهند.»

در صورت وجود $E(X)$ و $E(Y)$ می‌توان فرض‌های بالا را به صورت زیر بیان نمود.

۱) آزمون دوطرفه (دو دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: E(X) = E(Y) \\ \mathrm{H}_1: E(X) \neq E(Y) \end{cases}$$

۲) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: E(X) \leq E(Y) \\ \mathrm{H}_1: E(X_i) > E(Y) \end{cases}$$

۳) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} \mathrm{H}_0: E(X) \geq E(Y) \\ \mathrm{H}_1: E(X) < E(Y) \end{cases}$$

آماره آزمون

مقدار R_i ، که رتبه علامت‌دار نامیده می‌شود، که برای هر زوج (X_i, Y_i) به صورت زیر تعریف می‌شود.

R_i : رتبه‌ای که به ازای $D_i > 0$ به (X_i, Y_i) اختصاص داده می‌شود.

R_i : منفی رتبه‌ای که به ازای $D_i < 0$ به (X_i, Y_i) اختصاص داده می‌شود.

آماره آزمون به صورت زیر تعریف می‌شود.

$$W = \frac{\sum_{i=1}^{n} R_i}{\sqrt{\sum_{i=1}^{n} R_i}}$$

در حالتی که هیچ هم‌مرتبه‌ای وجود نداشته باشد، عبارت بالا به صورت زیر که ساده‌تر است، تغییر خواهد کرد.

$$W = \frac{\sum_{i=1}^{n} R_i}{\sqrt{n(n+1)(2n+1)/6}}$$

همچنین با استفاده از لم ۲.۴.۱ موجود در کتاب [10] در حالتی که زوج هم مرتبه وجود نداشته باشد، می‌توان نوشت

$$W^{+} = \sum_{i=1}^{n} (R_i) \times I(D_i > 0)$$

به طوری که در آن I تابع نشانگر است.

قاعده تصمیم‌گیری

خوشبختانه برای توزیع W جدول‌هایی نظیر جدول ۱ تهیه شده است که از طریق آن می‌توان ناحیه بحرانی را تشخیص داده و نسبت به رد یا تایید فرض صفر اقدام کرد. با فرض این‌که w_p چندک p ام باشد، در این صورت قاعده تصمیم‌گیری سه آزمون به صورت زیر خواهد بود.

۲) فرض H_0 رد می‌شود هر گاه

$$W \leq w_{\alpha} \qquad \text{یا} \qquad W \geq w_{1-\alpha}$$

۳) فرض H_0 رد می‌شود هر گاه

$$W \geq w_{1-\alpha}$$

۴) فرض H_0 رد می‌شود هر گاه

$$W \leq w_{\alpha}$$

جدول ۱: جدول نقاط بحرانی آماره ویلکاکسون

n	alpha values						
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
5	--	--	--	--	--	0	2
6	--	--	--	--	0	2	3
7	--	--	--	0	2	3	5
8	--	--	0	2	3	5	8
9	--	0	1	3	5	8	10
10	--	1	3	5	8	10	14
11	0	3	5	8	10	13	17
12	1	5	7	10	13	17	21
13	2	7	9	13	17	21	26
14	4	9	12	17	21	25	31
15	6	12	15	20	25	30	36
16	8	15	19	25	29	35	42
17	11	19	23	29	34	41	48
18	14	23	27	34	40	47	55
19	18	27	32	39	46	53	62
20	21	32	37	45	52	60	69

آزمون من- ویتنی[27]

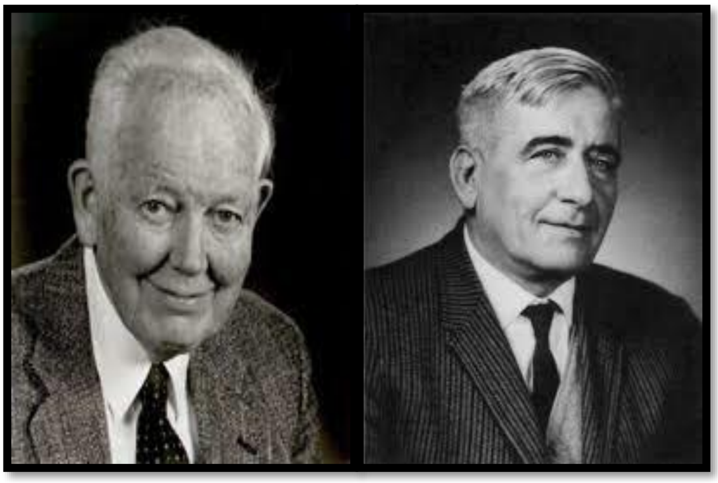
آزمون من–ویتنی یکی از قوی‌ترین آزمون‌های ناپارامتری است. در حقیقت آزمونی برای مقایسه وضعیت دو گروه مستقل است و وقتی داده‌های یک مطالعه به صورت کیفی ترتیبی باشند بهتر است از این آزمون که یک آزمون ناپارامتری و معادل آزمون دو نمونه مستقل t است، استفاده کرد. در این حال از آزمون t دو نمونه مستقل استفاده نمی کنیم زیرا میانگین متغیری که در مقیاس ترتیبی اندازه‌گیری شده باشد، به علت یکسان نبودن فاصله واحدها، معنی و مفهوم واقعی میانگین را نخواهد داشت. همان‌طور که ذکر شد آزمون‌های ناپارامتری تحت شرایط زیر باید به معادل‌های آزمون تی ترجیح داده شوند[8]:

- هنگامی که داده‌ها فقط به صورت مقیاس اندازه‌گیری ترتیبی هستند.
- هنگامی که داده‌ها فاصله‌ای یا نسبی، اما دارای توزیع غیر طبیعی هستند. (مثلاً دارای کجی شدید و غیر نرمال هستند)
- هنگامی که داده‌ها فاصله‌ای یا نسبی هستند، اما واریانس‌های دو نمونه در آزمون واریانس برابر نیستند.

در واقع آزمون من ویتنی برای محاسبه تفاوت های بین دو گروه مورد بررسی، مقادیر مربوط به هر دو گروه یا توزیع را به صورت یک مجموع واحد و بدون توجه به اینکه هر مقدار به کدام گروه تعلق دارد، از بیشترین صفت تا کمترین رتبه بندی می کند. اگر در زمان رتبه بندی مقادیر یکسان یا تکراری وجود داشته باشد، رتبه های مربوط به آن مقدارها با یکدیگر جمع شده و مجموع بدست آمده بر تعداد آن ها تقسیم می شود و رتبه بندی مشترکی برای تمام آن ها لحاظ می شود.

[27] Mann- Whitney test

این آزمون توسط توسط دانشمندان آماری هنری مَن[۲۸] و دونالد ویتنی[۲۹] در سال ۱۹۴۷ طی مقاله‌ای ارائه شده است[9]. به همین دلیل این آزمون من-ویتنی نام گرفته است. آماره من-ویتنی را با W و یا U نمایش می‌دهند.



شکل۳:سمت راست هنری من، سمت چپ دونالد ویتنی

شرایط آزمون

- متغیر وابسته[۳۰] باید به صورت ترتیبی یا سطوحی از متغیر عددی باشد. برای مثال می‌توان متغیرهایی با مقیاس طیف لیکرت[۳۱] را (مثلا با ۵ سطح) به عنوان متغیر وابسته در نظر گرفت. البته در این آزمون می‌توانید از متغیرهای عددی مانند درآمد، وزن، بهره هوشی و ... نیز استفاده کنید.
- متغیر مستقل[۳۲] باید بیانگر دو طبقه یا دو وضعیت باشد. در نتیجه متغیرهای طبقه‌ای (ترتیبی یا اسمی) که به صورت دو وضعیتی باشند، می‌توانند به عنوان متغیر مستقل معرفی شوند. برای مثال جنسیت (زن- مرد)، وضعیت اشتغال (شاغل- بیکار) از گروه متغیرهای طبقه‌ای یا دو وضعیتی هستند که قابلیت استفاده در آزمون من ویتنی را دارند.
- نمونه‌ها از هر دو جامعه باید به صورت تصادفی باشد. به این ترتیب نمونه‌گیری باید به شکل تصادفی در هر گروه انجام شود. بنابراین نباید یک مشاهده بیش از یک بار و یا در بیش از یک گروه به کار گرفته شود. از طرفی استقلال یا عدم وابستگی بین گروه یا جامعه‌ها نیز از فرضیه اصلی این آزمون محسوب می‌شود.
- از آنجایی که این آزمون یک روش ناپارامتری است، فرض بر این است که داده‌ها دارای توزیع نرمال نیستند یا حداقل در مورد توزیع آن‌ها اطلاعاتی در دست نیست. زیرا اگر داده‌ها دارای توزیع نرمال یا تقریبا نرمال باشند، روش‌های پارامتری بهتر از آزمون من ویتنی پاسخ داده و توان آزمون بزرگتری دارند. بنابراین بهتر است از آزمون‌ها پارامتری برای انجام مقایسه توزیع دو جامعه استفاده کرد.

---

[28] Henry Mann
[29] Donald Ransom Whitney
[30] Dependent Variable
[31] Likert spectrum
[32] Independent Variable

فرض‌ها

زمانی که توزیع داده‌ها را در بین دو گروه بررسی می‌کنید، بسیار بعید است که توزیع احتمالی برای آن‌ها به مانند یکدیگر باشد ولی ممکن است که اختلاف در بین توزیع احتمالی دو گروه ناشی از اختلاف در معیار مرکزی باشد. به این ترتیب به نظر می‌رسد که دو گروه در شکل توزیع یکسان ولی در پارامتر مکان اختلاف داشته باشند. فرض یکسان بودن توزیع در استنباط آماری به عنوان فرض صفر در نظر گرفته می‌شود. در آزمون من ویتنی، با استفاده از مقایسه میانگین رتبه‌ها، می‌توان هم‌توزیع بودن و یکسان بودن پارامتر مکان را مشخص کرد.

$$\begin{cases} \mathrm{H_0}: & \text{بین دو گروه تفاوتی وجود ندارد} \\ \mathrm{H_1}: & \text{بین دو گروه تفاوت وجود دارد} \end{cases}$$

برای انجام این آزمون باید با داشتن نمونه‌های تصادفی

$$X_1, X_2, \dots, X_{n_1} \quad \text{و} \quad Y_1, Y_2, \dots, Y_{n_2}$$

زوج‌های (X_i, Y_i) را شناسایی کرد. تعداد کل زوج‌ها $n_1 \times n_2$ است. پس از شناسایی زوج‌ها باید آن‌ها را رتبه‌گذاری نمود تا بتوان آماره را یافت. همچنین در نظر بگیرید که مقادیر این دو گروه در یک مجموعه قرار گرفته و رتبه‌های مربوط به مقدارها، از کم به زیاد تعیین شده باشد. در این صورت اگر R_1 مجموع رتبه‌های گروه اول و R_2 مجموعه رتبه‌های گروه دوم باشد آماره آزمون U به صورت زیر تعریف می‌شود.

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

نکته: مشخص است که اگر گروه اول دارای مقدارهای کوچکتر از گروه دوم باشد همه رتبه‌های کوچکتر به آن گروه تعلق خواهد داشت و به این ترتیب مجموع رتبه‌های مربوط به گروه اول براساس دنباله حسابی به صورت زیر خواهد شود.

$$R_1 = 1 + 2 + \cdots + n_1 = \frac{n_1(n_1+1)}{2}$$

در نتیجه هر چقدر مقدار U_1 کوچک باشد رای به یکسان نبودن توزیع دو گروه خواهیم داد. از طرفی می‌توان ملاک را گروه ۲ در نظر گرفت. در نتیجه می‌توان مجموع رتبه‌ها را برحسب این گروه محاسبه کرد.

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2}$$

با توجه به کوچک بودن مقدار U_2 نیز به رای عدم یکسان بدون توزیع‌ها در بین دو گروه خواهیم داد. اگر فرض صفر را به صورت یکسان بودن توزیع دو گروه در نظر گرفته شود، به این ترتیب می‌توان گفت که کوچک بودن U_1 یا U_2 دلیلی بر رد صفر خواهد بود. باید توجه داشت که مجموع U_1 و U_2 ثابت و برابر با خواهد بود، زیرا:

$$U_1 + U_2 = R_1 - \frac{n_1(n_1+1)}{2} + R_2 - \frac{n_2(n_2+1)}{2}$$

بنابراین اگر $N = n_1 + n_2$ باشد، واضح است که

$$R_1 + R_2 = \frac{1}{2}N(N+1)$$

است. با اندکی محاسبات جبری خواهیم داشت:

$$U_1 + U_2 = \frac{N(N+1)}{2} - \frac{n_1(n_1+1)}{2} - \frac{n_2(n_2+1)}{2} =$$

$$\frac{1}{2}(N^2 + N - n_1^2 - N - n_2^2) = \frac{1}{2}\left((n_1+n_2) - n_1^2 - n_2^2\right) =$$

$$\frac{1}{2}(2n_1n_2) = n_1n_2$$

به این ترتیب می‌توان مقدار U_1 را برحسب U_2 یا برعکس محاسبه کرد. از طرفی برای آماره آزمون می‌توان هم از U_1 و هم از U_1 استفاده کرد.

توجه: آزمون ویلکاکسون جمعی رتبه‌ای درست به مانند آزمون من- ویتنی است و این مسئله در قالب یک قضیه اثبات می‌شود.

آزمون مك نمار[33]

از آزمون‌های آماری ناپارامتری است که برای ارزیابی همانندی دو نمونه وابسته بر حسب متغیر دو جوابی استفاده می‌شود. متغیرها می‌توانند دارای مقیاس‌های اسمی یا رتبه‌ای باشند. این آزمون در طرح‌های ماقبل و مابعد می‌تواند مورد استفاده قرار گیرد (یک نمونه در دو موقعیت مختلف). برای نمونه می توان به بررسی نظرات افراد در مورد مشارکت در انتخابات ریاست جمهوری پیش از سخنرانی در آن مورد و پس از سخنرانی، اشاره نمود. پس اجرای آزمون مک نمار مستلزم وجود دو مجموعه از داده های دو مقوله ای است. این داده ها می توانند به طور ذاتی دو مقوله ای باشند نظیر «بلی یا خیر» و یا به صورت طبقه ای، رتبه ای یا فاصله ای باشند که پژوهشگر آن ها را به صورت دو مقوله ای کدبندی می کند نظیر گرایش سیاسی که به صورت فاصله ای اندازه گیری شود، اما پژوهشگر آن ها را به صورت «مثبت یا منفی» مقوله بندی نماید.

جدول زیر داده های مربوط به این آزمون را نمایش می دهد :

جدول ۲: نمود دیداری جدول 2×2 برای آزمون مک- نمار

[33] McNemar's test

داده سنجش اول / داده سنجش دوم	مقوله اول	مقوله دوم
مقوله اول	A	B
مقوله دوم	C	D

آزمون مک نمار به افرادی که تغییر عقیده یا عملکرد نداده اند، یعنی مقادیر **A** و**D** اهمیت نمی دهد و تنها با مقادیری سروکار دارد که با تغییر عقیده مربوط می شوند یعنی **B** و**C** . منطق اصلی انجام این آزمون این است که اگر تدبیر یا عامل مورد نظر که پس از سنجش اول اعمال می شود مؤثر نباشد، می بایست یا در عقیده ها و عملکردها تغییری صورت نگیرد و یا تغییر در میان آزمودنی ها به سوی یکی از طرف ها (مثبت یا منفی، مخالف یا موافق و...) برابر با طرف دیگر باشد. بنابراین فرضیه پژوهش در این آزمون بیان می دارد که بین دو دسته از داده ها تفاوت معناداری ایجاد می شود.

این آزمون توسط کوئین مک نمار[۳۴] روان‌شناس و آماردان آمریکایی در سال ۱۹۴۷ معرفی شد.

شکل ۴: کوئین مک نمار

فرض‌ها

۱) آزمون دوطرفه (دو دنباله‌ای)

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

۲) آزمون یک طرفه (یک دنباله‌ای)

$$\begin{cases} H_0: \mu_1 \leq \mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases}$$

۳) آزمون یک طرفه (یک دنباله‌ای)

[34] Quinn Michael McNemar

$$\begin{cases} H_0: \mu_1 \geq \mu_2 \\ H_1: \mu_1 < \mu_2 \end{cases}$$

آماره آزمون

آماره آزمون مک نمار به صورت زیر است.

$$Q(B,C) = \frac{(B-C)^2}{B+C}$$

زمانی‌که نمونه کوچک است با استفادهاز تصحیح پیوستگی این آماره به صورت زیر است.

$$Q_C(B,C) = \frac{(|B-C|-1)^2}{B+C}$$

این آماره به طور تقریبی دارای توزیع کای‌دو با یک درجه آزادی است.

آزمون کراسکال-والیس

آزمون کراسکال-والیس[۳۵] برای مقایسه سه یا بیش از سه گروه مستقل که در سطح رتبه‌ای اندازه گیری می‌شوند، مورد استفاده قرار می‌گیرد[12]. می‌توان گفت که این آزمون تعمیمی از آزمون من-ویتنی که تنها برای دو گروه استفاده می‌شد است. در واقع این آزمون معادل ناپارامتری آزمون F مستقل در روش تحلیل واریانس یکطرفه، می‌باشد. زمانی که فرض‌های بنیادی تحلیل واریانس مانند نرمال بودن توزیع داده‌ها و برابری واریانس گروه‌ها برقرار نباشد، از آزمون کروسکال والیس استفاده می‌شود. به همین دلیل گاهی به این آزمون «تحلیل واریانس رتبه‌ای» نیز گفته می‌شود. این آزمون می‌تواند در مورد داده‌های پیوسته (فاصله ای یا نسبی) نیز به کار برده شود، در این حالت باید توجه شود که داده‌ها به صورت داده‌های رتبه‌ای تبدیل شده و مورد استفاده قرار می‌گیرد.

یکی از کاربردهای این آزمون است که گاه به دنبال مقایسه ی میانگین یک متغیر کمی در سطوح مختلف یک متغیر کیفی چند سطحی (بیش از دو سطح) هستیم. یعنی در این حالت یک متغیر کمی و یک متغیر کیفی چند سطحی داریم. به عنوان مثال زمانی که درصدد بررسی میانگین فشار خون در گروه‌های سنی مختلف (کمتر از ۲۵ سال، ۲۵ تا ۴۵ سال، ۴۵ تا ۶۵ سال و بالاتر از ۶۵ سال) هستیم، فشارخون متغیر کمی و گروه های سنی متغیر کیفی چند سطحی خواهند بود. تا زمانی که میانگین فشارخون از توزیع نرمال پیروی می کند، مقایسه ی میانگین‌ها از طریق آنالیز واریانس[۳۶] (ANOVA) امکان پذیر خواهد بود. مشکل زمانی رخ می‌دهد که متغیر کمی مورد بررسی ما نرمال نباشد. در این حالت به دلیل عدم برقراری فرض نرمال بودن که یکی از پیش فرض‌های استفاده از آنالیز واریانس به حساب می‌آید به سراغ آزمون ناپارامتری هم خانواده با ANOVA یعنی آزمون کروسکال والیس می‌رویم.

[35] Kruskal–Wallis test
[36] Analysis of variance

این آزمون توسط ویلیام کروسکال[37] و آلن والیس[38] معرفی شده است[11]. کروسکال یک اقتصاددان و آماردان آمریکایی والیس یک ریاضی‌دان و آماردان آمریکایی بود .

شکل۵: سمت راست ویلیام کروسکال ، سمت چپ آلن والیس

فرض

کروسکال- والیس این فرضیه را که k گروه نمونه از یک جامعه آماری مشترک یا جامعه آماری شبیه به هم که با توجه به میانگین ها استخراج شده‌اند، آزمون می‌کند. آنالیز واریانس یک طرفه کروسکال-والیس با استفاده از رتبه‌ها آزمون فوق العاده مفیدی برای تصمیم‌گیری درباره این است که آیا k گروه نمونه مستقل از جامعه‌های آماری مختلف آمده‌اند یا نه؟ بدیهی است که نمونه‌ها بدون استثنا اختلافاتی با یکدیگر دارند ولی سوال این است که آیا اختلافات مشاهده شده در نمونه‌ها نماینده اختلافات موجود در جوامع هستند یا ناشی از شانس و تصادف اند؟ فرضیه صفر در این آزمون بر خلاف فرض مقابل آن، تاکید بر عدم اختلاف بین گروه‌ها دارد. این فرضیه با توجه به میانگین‌ها، مبنا را بر شباهت k نمونه از یک جامعه مشترک می‌گیرد. یعنی دو فرضیه صفر و یک به صورت زیر مطرح می‌شوند.

$$\begin{cases} \mathrm{H}_0: \mu_1 = \mu_2 = \cdots = \mu_k \\ \mathrm{H}_1: \mu_i \neq \mu_j \qquad i \neq j \end{cases}$$

در این آزمون همه k نمونه را روی هم ریخته (تا n مشاهده به دست آید) سپس هر یک از n مشاهده را به صورت رتبه در می‌آوریم. کوچک‌ترین مقدار رتبه یک و بیش‌ترین مقدار، آخرین رتبه (رتبه n ام) را به خود اختصاص می‌دهد. اینک برای هر یک از K گروه، مجموع رتبه‌ها را محاسبه می‌کنیم. آزمون کروسکال والیس معلوم می‌کند که آیا این مجموعه رتبه‌ها چنان با یکدیگر تفاوت دارند که نتوان گفت آن‌ها از یک جامعه آماری مشترک استخراج شده اند.

37 William Kruskal
38 W. Allen Wallis

آماره آزمون

آماره آزمون کراسکال-والیس به صورت زیر است.

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

که در آن k تعداد گروه‌ها، n_j تعداد اعضای گروه j ام ، n تعداد کل (مجموع اعضای تمام گروه‌ها) و R_j مجموع رتبه مربوط به گروه j ام است.

قاعده تصمیم‌گیری

آماره بالا برای نمونه های بزرگ (اگر تعداد اعضای هر گروه بیشتر از ۵ نفر باشد) تقریبا دارای توزیع کای دو با $k-1$ درجه آزادی است که با توجه به سطح معنی‌داری تعیین شده برای رد یا پذیرش فرض صفر می‌توان از جدول توزیع کای‌دو استفاده کرد. ولی اگر تعداد اعضای هر نمونه کمتر از ۵ و تعداد گروه ها هم ۳ تا باشند باید از جدول مربوط به آزمون کروسکال-والیس استفاده کرد[13].

جدول ۳: جدول آزمون کراسکال والیس

Sample Size	$H_{0.9000}$	$H_{0.9500}$	$H_{0.9750}$	$H_{0.9900}$	$H_{0.9950}$	$H_{0.9975}$	$H_{0.9990}$
5,5,2,1	5.5648	6.5341	7.2725	8.3077	9.0198	9.4352	9.7582
5,5,2,2	5.7943	6.7714	7.6457	8.6286	9.2914	9.8800	10.3429
5,5,3,1	5.6476	6.7371	7.6286	8.5962	9.2743	9.7619	10.2191
5,5,3,2	5.9150	6.9417	7.8750	8.9467	9.6350	10.1667	10.8200
5,5,3,3	6.0118	7.1176	8.0588	9.1882	9.9176	10.5529	11.2353
5,5,4,1	5.6625	6.7800	7.7625	8.8625	9.5500	10.1025	10.5900
5,5,4,2	5.9338	7.0279	8.0162	9.1500	9.8868	10.5154	11.1904
5,5,4,3	6.0523	7.2157	8.2092	9.3562	10.1307	10.7895	11.5739
5,5,4,4	6.0684	7.2895	8.3421	9.5351	10.3281	11.0228	11.8439
5,5,5,1	5.6824	6.8294	7.8000	9.0176	9.7588	10.3941	10.9588
5,5,5,2	5.9451	7.0745	8.0941	9.2863	10.0980	10.7451	11.5137
5,5,5,3	6.0433	7.2456	8.2889	9.4959	10.3193	10.9930	11.8257

آزمون فریدمن[39]

این آزمون نیز هماننند آزمون کراسکال-والیس برای مقایسه چند گروه از نظر میانگین رتبه های آن‌ها است و معلوم می‌کند که آیا این گروه‌ها می‌توانند از یک جامعه باشند یا نه؟

مقیاس در این آزمون باید حداقل رتبه ای باشد. این آزمون متناظر غیر پارامتری آزمون *F* است و معمولا در مقیاس های رتبه ای به جای *F* به کار می‌رود و جانشین آن می‌شود (چون در *F* باید همگنی واریانس ها وجود داشته باشد که در مقیاس‌های رتبه ای کمتر رعایت می‌شود). همچنین آزمون فریدمن تعمیم یافته آزمون ویلکاکسون است و معادل ناپارامتریک آزمون اندازه‌های مکرّر است.

آزمون فریدمن برای تجریه واریانس دو طرفه (برای داده‌های ناپارامتری) از طریق رتبه بندی به کار می‌رود و نیز برای مقایسه میانگین رتبه بندی گروه های مختلف. تعداد افراد در نمونه‌ها باید یکسان باشند که این از معایب این آزمون است. نمونه‌ها باید همگی جور شده باشند. این آزمون برای طرح‌های درون گروهی (نمونه‌های وابسته) مناسب است.

این آزمون توسط میلتون فریدمن[40] اقتصاددان، آماردان و نویسنده آمریکایی معرفی شد.

از معایب این آزمون این است که تعداد افراد در نمونه‌ها باید یکسان باشند و نمونه‌ها باید همگی جور شده باشند.

آزمون کوکران[41]

آزمون کوکران یک آزمون ناپارامتری که به بررسی متغیرهای دو ارزشی در نزد بیش از ۲ گروه می پردازد. در اصل در این آزمون را می‌توان به عنوان آزمون تعمیم یافته‌ی آماره ی مک نمار حساب آورد ، اما با این تفاوت که این روش در زمان‌هایی به کار برده می‌شود که تعداد گروه‌ها یا تکرارها ۳ یا بیشتر از ۳ باشد . این روش زمانی کاربرد دارد که متغیر پاسخ یا وابسته به صورت دو ارزشی نظیر موافق یا مخالف، شکست یا موفقیت، بلی یا خیر، و نظایر این‌ها بوده و بتوان به پاسخ کدهای صفر و ۱ داد. در این نظام نمره‌گذاری به وجود صفت مورد نظر نظیر موافق، موفقیت، قبولی و نظایر این ها نمره‌ی ۱ و به عدم وجود صفت مورد نظر نظیر مخالف، شکست، مردودی و نظایر این‌ها نمره ی صفر داده می‌شود.

[39] Friedman test
[40] Milton Friedman
[41] Cochran's C test

این آزمون توسط ویلیام کوکران[42] معرفی شد.

شکل ۷: ویلیام کوکران

فرضیه ها در آزمون کوکران

فرضیه ی صفر در این آزمون بیان می دارد که میان گروه ها یا تکرارهای مختلف در زمینه ی متغیر آزمونی یا وابسته برابری وجود دارد و یا تفاوت معنی داری وجود ندارد و فرضیه‌ی پژوهش وجود تفاوت معنی دار میان گروه های مورد بررسی یا تکرارها را می رساند.

آماره آزمون

آماره آزمون به صورت زیر تعریف می شود:

$$Q = \frac{k(k-1)\sum_{i=1}^{k}(x_i - T)^2}{k\sum_{j=1}^{n} y_j - \sum_{j=1}^{n} {y_j}^2}$$

که در آن k تعداد گروه ها یا تکرارها، n تعداد آزمودنی ها، x_i جمع مقادیر مربوط به آزمودنی i ام ام در گروه ها (تکرارها) ی مختلف، y_j جمع پاسخ های تمام آزمودنی ها در گروه (تکرار j ام) و همچنین

$$T = \frac{\sum_{i=1}^{k} x_i}{k}$$

، میانگین مربوط به کل گروه ها می باشند.

به منظور بررسی نتیجه حاصل از این آماره از جدول توزیع کای‌دو استفاده می‌شود. با توجه به سطح خطای مورد نظر که به طور معمول ۰.۰۵ درنظر گرفته می شود، مقدار کای‌دو بحرانی را با $k-1$ درجه آزادی از جدول آن بدست

[42] William Gemmell Cochran

می‌آوریم. چنانچه مقدار آماره Q بزرگتر از مقدار بحرانی جدول باشد، فرضیه صفر یعنی «عدم وجود تفاوت بین گروه ها (تکرارها)» تأیید نمی شود.

ضریب همبستگی[43]

در بسیار ی از مسائل کاربردی کاربر به رابطه بین دو متغیر علاقه‌مند است. در آمار، از کوواریانس[44] و همبستگی برای کمّی کردن رابطه بین متغیرها بهره می‌برند. کوواریانس به واحد اندازه‌گیری داده‌ها بستگی دارد. در نتیجه نمی‌توان بزرگی کوواریانس دو متغیر را با بزرگی کوواریانس دو متغیر دیگر بدون در نظر گرفتن واحد اندازه‌گیریشان، مقایسه کرد. ضریب همبستگی که شاخصی بدون واحد است، این مشکل را حل می‌کند. ضریب همبستگی ابزاری آماری برای تعیین نوع و درجه‌ی رابطه‌ی یک متغیر کمی با متغیر کمی دیگر است و یکی از معیارهای مورد استفاده در تعیین همبستگی دو متغیر است که شدت رابطه و همچنین نوع رابطه (مستقیم یا معکوس) را نشان می‌دهد.
با توجه به نوع داده‌ها، شیوه‌های مختلفی برای اندازه‌گیری ضریب همبستگی وجود دارد. به همین دلیل آن‌ها به سه دسته ضریب همبستگی پیرسون[45]، ضریب همبستگی اسپیرمن[46] و ضریب همبستگی کندال[47] تقسیم می‌شوند.

ضریب همبستگی پیرسون

یکی از مشهورترین شیوه‌های اندازه‌گیری وابستگی بین دو متغیر کمی، محاسبه ضریب همبستگی پیرسون است. این شاخص توسط کارل پیرسون[48] آماردان انگلیسی در سال ۱۹۰۰ طی مقاله‌ای معرفی شد[14]. او از این شاخص برای بررسی علمی روی علوم زیستی و حتی جمعیتی استفاده کرد و به نتایج جالب توجهی رسید. شیوه محاسبه برای ضریب همبستگی پیرسون در ادامه دیده می‌شود.

شکل ۸: کارل پیرسون

43 Correlation coefficient
44 Covariance
45 Peasron Correlation Coefficient
46 Spearman Correlation Coefficient
47 Kendall Correlation Coefficient
48 Karl Pearson

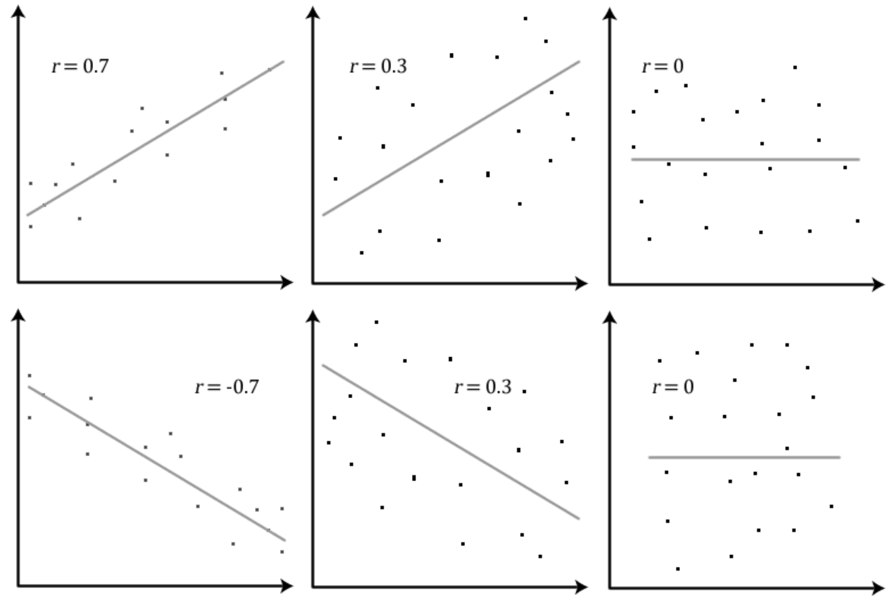
ضریب همبستگی پیرسون بین دو متغیر تصادفی (کمی) X و Y به صورت زیر تعریف می‌شود.

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X\sigma_Y}$$

با استفاده از نامساوی کوشی-شوارتز نیز می‌توان نشان داد که قدر مطلق ضریب همبستگی هرگز بزرگتر از ۱ نخواهد بود. پس می‌توان نوشت

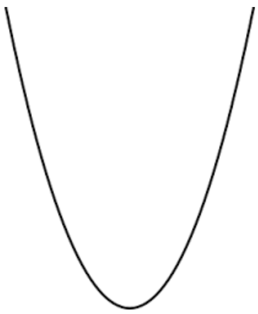
$$-1 \leq \rho(X,Y) \leq 1$$

در صورت عدم وجود رابطه بین دو متغیر، برابر صفر است. (به شکل ۹ توجه کنید.)



شکل ۹: نمودارهای همبستگی دو متغیر

نکته: $r=0$ بیانگر عدم وجود ارتباط بین دو متغیر تصادفی X و Y نیست بلکه تنها نشان‌دهنده عدم وجود ارتباط **خطی** بین دو متغیر تصادفی X و Y است. این دو متغیر می‌توانند ارتباط غیر خطی داشته باشند و یا نداشته باشند. به عنوان مثال به شکل ۱۰ که بیانگر وجود ارتباط درجه ۲ بین دو متغیر تصادفی X و Y است می‌توان اشاره کرد.



شکل۱۰: ارتباط درجه ۲

نکته: اگر ضریب همبستگی دو پارامتر با یکدیگر مثبت باشد، به این معناست که در فضایی که مطالعه و بررسی انجام شده، افزایش یک پارامتر با افزایش پارامتر دیگر و نیز کاهش آن پارامتر با کاهش پارامتر دیگر همراه است و اگر ضریب همبستگی دو پارامتر با یکدیگر منفی باشد، به این معناست که در فضایی که مطالعه و بررسی انجام شده، افزایش یک پارامتر با کاهش پارامتر دیگر و کاهش آن پارامتر با افزایش پارامتر دیگر همراه است.

خصوصیات

- بدون واحد بودن: ضریب همبستگی پیرسون به واحد اندازه‌گیری داده‌ها بستگی ندارد. یعنی شاخصی بدون واحد است. حتی گاهی آن را به صورت درصدی نیز بیان می‌کنند.
- تقارن ضریب همبستگی پیرسون: واضح است که ضریب همبستگی پیرسون دارای تقارن است. زیرا

$$Cov(X,Y) = Cov(Y,X)$$

- اندازه‌گیری ارتباط خطی بین دو متغیر: هر چه مقدار ضریب همبستگی به ۱ یا ۱- نزدیک شود، وجود رابطه خطی بین دو متغیر بیشتر می‌شود.
- تعیین جهت همبستگی: چنانچه ضریب همبستگی مثبت باشد، رابطه بین دو متغیر را مستقیم و اگر منفی باشد، رابطه بین دو متغیر معکوس خواهد بود.
- استقلال دو متغیر: اگر دو متغیر مستقل باشند، ضریب همبستگی پیرسون برابر با صفر خواهد بود. البته عکس این موضوع صحیح نیست. یعنی ممکن است ضریب همبستگی پیرسون برای دو متغیر برابر با صفر باشد در حالیکه آن دو متغیر مستقل نیستند.
- استقلال برای دو متغیر نرمال: اگر دو متغیر X و Y دارای توزیع نرمال باشند، آنگاه صفر بودن ضریب همبستگی می‌تواند دلیلی برای استقلال دو متغیر تصادفی نرمال باشد.

ضریب همبستگی نمونه‌ای

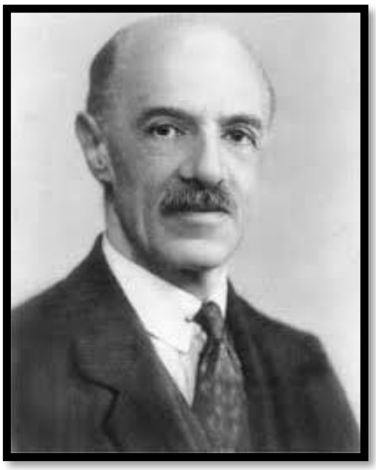
اگر یک نمونه تصادفی به حجم n از متغیرهای تصادفی X و Y به صورت (x_i, y_i) داشته باشیم، ضریب همبستگی نمونه‌ای پیرسون[۴۹] به شکل زیر محاسبه می‌شود:

$$r(x,y) = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\left(n\sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right)^{\frac{1}{2}} \left(n\sum_{i=1}^{n} {y_i}^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)^{\frac{1}{2}}}$$

ضریب همبستگی رتبه‌ای اسپیرمن

با توجه به اینکه ضریب همبستگی پیرسون براساس میانگین و واریانس محاسبه می‌شود، ممکن است در مقابل داده‌های دورافتاده، منحرف شده و میزان همبستگی را به درستی نشان ندهد. در چنین مواقعی از ضریب همبستگی رتبه‌ای اسپیرمن استفاده می‌شود. در واقع ضریب همبستگی اسپیرمن آماره‌ای ناپارامتری برای سنجش ضریب همبستگی بین دو متغیر تصادفی است. ضریب همبستگی رتبه‌ای اسپیرمن به مانند ضریب همبستگی پیرسون، نشان می‌دهد تمایل یک متغیر به پیروی کردن از مقدارهای متغیر دیگر چقدر است. به عبارت دیگر مقدار این ضریب نشاندهندهٔ قابلیت بیان یک متغیر به صورت تابعی یکنوا از متغیر دیگر است. ضریب همبستگی اسپیرمن درواقع ضریب همبستگی پیرسونی است که بین داده‌های رتبه‌بندی شده تعریف می‌شود.

این ضریب همبستگی توسط چارلز اسپیرمن[۵۰] دانشمند روانشناس انگلیسی در سال ۱۹۰۴ معرفی شد. او با استفاده از این ضریب همبستگی توانست تئوری‌هایش در زمینه شناخت و هوش را توسعه دهد[15].



شکل ۱۱: چارلز اسپیرمن

49 Pearson Sample Correlation Coefficient
50 Charles Spearman

در این ضریب همبستگی به جای محاسبه روی مقدارها، از رتبه‌ها استفاده می‌شود. به همین دلیل به آن ضریب همبستگی رتبه‌ای نیز می‌گویند. بنابراین اگر

$r_{x_1}, r_{x_2}, \dots, r_{x_n}$ رتبه‌های مربوط به $x_1, x_2, \dots, x_n$

$r_{y_1}, r_{y_2}, \dots, r_{y_n}$ رتبه‌های مربوط به $y_1, y_2, \dots, y_n$

باشند، ضریب همبستگی رتبه‌ای اسپیرمن که به صورت r_s نشان داده می‌شود طبق رابطه زیر قابل محاسبه است.

$$r_s = \frac{Cov(r_x, r_y)}{s_{r_x} s_{r_y}}$$

ضریب همبستگی اسپیرمن شدت رابطه خطی را اندازه‌گیری نمی‌کند. به این معنی که ممکن است ضریب همبستگی رتبه‌ای اسپیرمن برابر ۱ باشد در حالی که رابطه خطی بین دو متغیر وجود نداشته باشد.

تفاوت ضریب همبستگی رتبه ای اسپیرمن و ضریب همبستگی پیرسون

- ضریب همبستگی پیرسون برای محاسبه ی همبستگی دو متغیر فاصله ای یا نسبی به کار برده می شود، ولی ضریب اسپیرمن ، همبستگی موجود بین دو متغیر ترتیبی را نشان می دهد.
- به کمک ضریب همبستگی اسپیرمن روابط غیرخطی بررسی می شود در حالیکه ضریب همبستگی پیرسون به منظور بررسی یک رابطه ی خطی بکار برده می شود.
- کارایی ضریب همبستگی رتبه ای اسپیرمن کمتر از ضریب همبستگی پیرسون است.
- محاسبه ی ضریب همبستگی اسپیرمن ساده تر بوده و نیاز به پیش فرض های کمتری نسبت به ضریب پیرسون دارد.

ضریب همبستگی کندال

ضریب همبستگی کندال[51] نیز مانند ضریب همبستگی اسپیرمن، به جای مقدار از ترتیب مقدارها برای اندازه‌گیری میزان وابستگی استفاده می‌کند. ضریب همبستگی رتبه‌ای کندال به تای کندال[52] نیز مشهور است. این روش در واقع یک آماره‌ی ناپارامتری است که برای سنجش همبستگی آماری میان دو متغیر تصادفی به کار می‌رود.

[51] Kendall rank correlation coefficient
[52] Kendall's tau coefficient

این ضریب همبستگی توسط مرسیه کندال[53] ، دانشمند انگلیسی علم آمار در سال ۱۹۳۸ معرفی شد[16]. او بوسیله این شاخص، میزان همخوانی رتبه‌ها را اندازه‌گیری کرد.

شکل ۱۲: مرسیه کندال

فرض کنید زوج‌های $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ مشاهدات متغیرهای تصادفی X و Y را تشکیل می‌دهند. در ادامه زوج هماهنگ و ناهمانگ تعریف کنیم.

زوج هماهنگ: زوج (x_i, y_i) و (x_j, y_j) را هماهنگ[54] می‌گویند اگر اگر داده‌های این زوج‌ها را براساس مولفه اول یا دوم مرتب کنیم، دارای رتبه‌های یکسانی خواهند بود. یعنی اگر $x_i < x_j$ باشد، آن‌گاه $y_i < y_j$.

زوج ناهماهنگ: زوج (x_i, y_i) و (x_j, y_j) را ناهماهنگ[55] می‌گویند اگر اگر داده‌های این زوج‌ها را براساس مولفه اول یا دوم مرتب کنیم، دارای رتبه‌های یکسانی خواهند بود. یعنی اگر $x_i > x_j$ باشد، آن‌گاه $y_i > y_j$.

حال براساس تعریف هماهنگ و ناهماهنگ برای زوج‌ها، اگر تعداد زوج‌های هماهنگ را با con و تعداد زوج‌های ناهماهنگ را نیز با dis نشان دهیم، ضریب همبستگی کندال به صورت زیر محاسبه می‌شود:

$$\tau = \frac{con - dis}{n(n-1)/2}$$

شکل دیگری نیز برای محاسبه ضریب همبستگی کندال وجود دارد که از تابع علامت[56] استفاده می‌کند، به صورت زیر تعریف می‌شود.

---

[53] Maurice George Kendall
[54] Concordant
[55] Disconcordant
[56] Sign function

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} sign(x_i - x_j) sign(y_i - y_j)$$

تابع علامت به صورت زیر تعریف می شود.

$$sign(x_i - x_j) = \begin{cases} -1 & x_i < x_j \\ 0 & x_i = x_j \\ 1 & x_i > x_j \end{cases}$$

خصوصیات

از آنجایی که مخرج این کسر تعداد انتخاب‌های زوجها از بین n مشاهده است، همیشه از صورت بزرگتر است. پس ضریب همبستگی کندال از ۱ کوچکتر و از ۱- بزرگتر است.

اگر همه زوجها با هم هماهنگ باشند مقدار ضریب همبستگی کندال برابر است با ۱.

اگر همه زوجها ناهماهنگ باشند ضریب همبستگی کندال برابر است با ۱-.

اگر X و Y مستقل باشند، انتظار داریم که ضریب همبستگی کندال نیز برابر با 0 باشد.

خلاصه

در جداول زیر آزمون‌های پارامتری و ناپارامتری برای یک متغیر، دو متغیر و سه متغیر و بیشتر طی جداول ۴، ۵، ۶ و ۷ خلاصه شده‌اند.

جدول ۴: آزمون آماری برای یک متغیر

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

جدول ۵: 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 واریانسANOVA	آزمون خی - دو

جدول ۶: 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	

جدول ۷: آزمون‌های آماری برای انواع داده

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

حل مثال ازآزمون‌های ناپارامتری در نرم افزار R

آزمون علامت ویلکاکسون

بیست دانشجو سال اول را به ده گروه دوتایی تقسیم می‌کنند به طوری که دو دانشجوی هر گروه از نظر بهره هوشی تقریبا یکسان باشند. به یک دانشجوی هر گروه با تلویزیون و به دانشجوی دیگر در کلاس، ریاضی عمومی را تدریس می‌کنند. در پایان نیمسال به همه آن‌ها یک نوع سؤال امتحانه می‌دهند. نمره‌های امتحان در جدول زیر داده شده‌اند. با استفاده از آزمون رتبه‌ای –نشانه‌ای ویلکاکسون، با میزان پنج درصد بیازمایید توزیع نمره تلویزیونی و توزیع نمره کلاسی اختلافی ندارند.

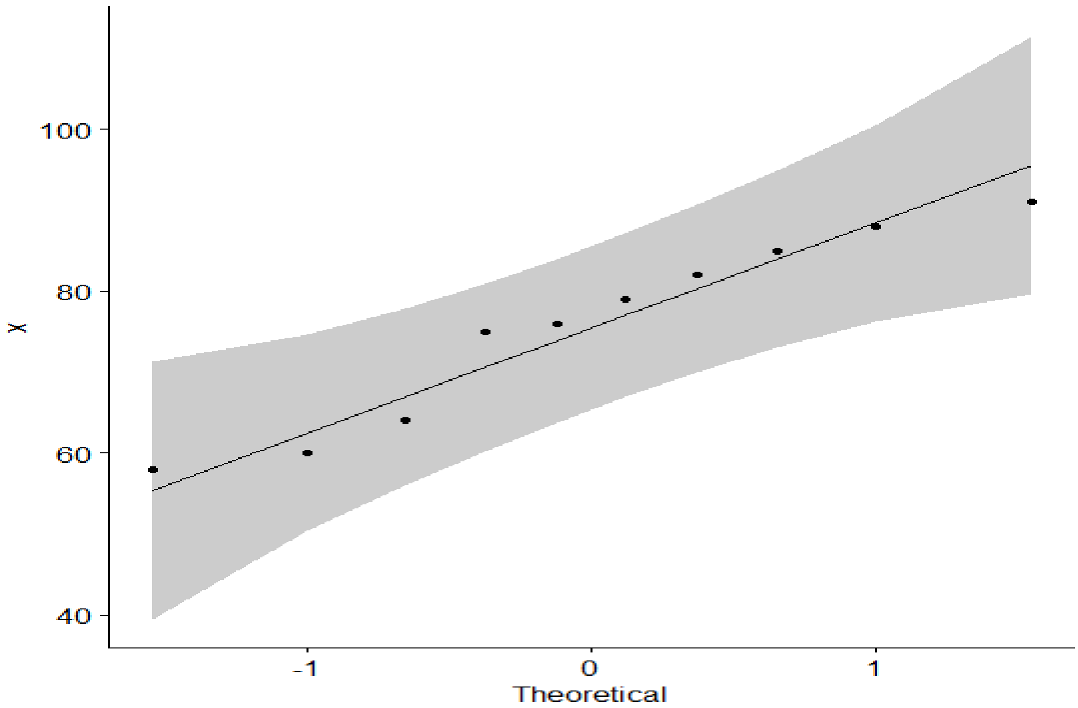
جدول ۸: داده‌های مثال آزمون علامت

گروه	نمره- تدریس با تلویزیون	نمره- تدریس در کلاس
۱	۷۶	۸۱
۲	۶۰	۵۲
۳	۸۵	۸۷
۴	۵۸	۷۰
۵	۹۱	۸۶
۶	۸۲	۹۰
۷	۶۴	۸۳
۸	۷۹	۸۵
۹	۸۸	۸۳
۱۰	۷۵	۷۷

```
> x<-c(76,60,85,58,91,82,64,79,88,75)
> y<-c(81,52,87,70,86,90,83,85,83,77)
```

قبل از انجام آزمون، نرمال بودن داده‌ها باید بررسی گردند. برای رسم نمودار "qqplot" برای تشخیص نرمال بودن داده‌ها به صورت صوری از پکیج "ggpubr" استفاده می‌کنیم.

```
> library("ggpubr")
> ggqqplot(x, ylab = "x")
```



شکل ۱۳: نمودار "qqplot" برای نمره- تدریس با تلویزیون

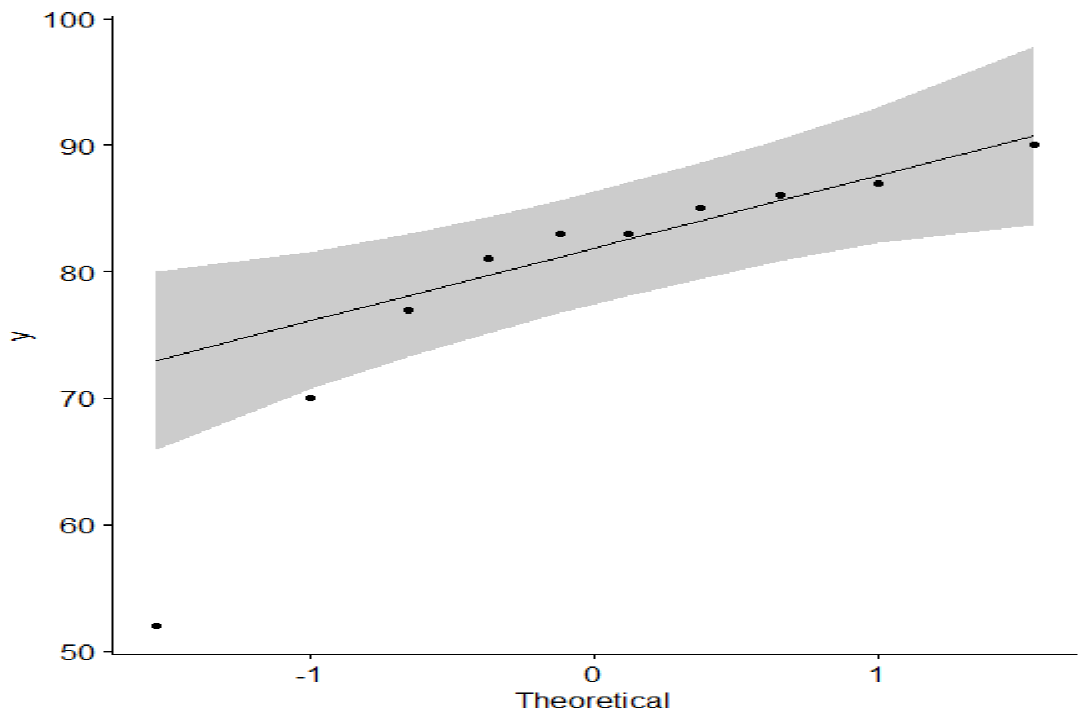
شکل۱۳ برای تشخیص نرمال بودن داده‌های نمره- تدریس با تلویزیون رسم شده است. با توجه به شکل نقاط تقریبا روی خط قرار می‌گیرند پس فرض نرمال بودن داده‌ها پذیرفته می‌شود.

برای بررسی دقیق‌تر نرمال بودن داده‌ها از آزمون "shapiro.test" استفاده می‌شود.

```
> shapiro.test(x)
	Shapiro-Wilk normality test
data:  x
W = 0.92705, p-value = 0.4195
```

در بالا آزمون نرمال بودن برای داده‌های نمره- تدریس با تلویزیون صورت گرفته است با توجه به اینکه مقدار p-value بزرگتر از ۰.۰۵ است فرض نرمال بودن داده‌ها پذیرفته می‌شود.

```
> ggqqplot(y, ylab = "y")
```



شکل ۱۴: نمودار "qqplot" برای نمره- تدریس در کلاس

شکل۱۴ برای تشخیص نرمال بودن داده‌های نمره- تدریس در کلاس رسم شده است. با توجه به شکل نقاط روی خط قرار نمی‌گیرند پس فرض نرمال بودن داده‌ها رد می‌شود.

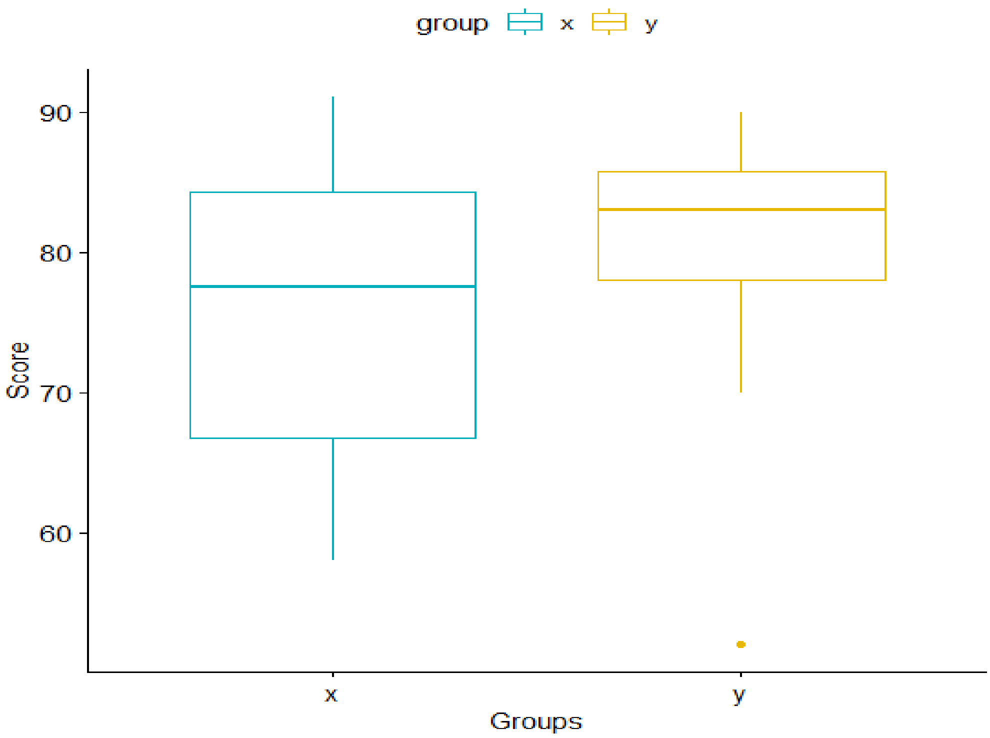
```
> shapiro.test(y)

	Shapiro-Wilk normality test

data:  y
W = 0.79477, p-value = 0.01252
```

در بالا آزمون نرمال بودن برای داده‌های نمره- تدریس در کلاس صورت گرفته است با توجه به اینکه مقدار p-value کمتر از ۰.۰۵ است فرض نرمال بودن داده‌ها رد می‌شود.

در زیر نمودار جعبه‌ای داده‌های میزان بارندگی و آلودگی هوا برای تشخیص نرمال بودن داده‌ها با پکیج "ggpubr" رسم شده است:

```
> library("ggpubr")
> my_data <- data.frame(
```

```
+             group = rep(c("x", "y"), each = 10),
+ Score = c(x,y)
+             )
> ggboxplot(my_data, x = "group", y = "Score",
+         color = "group", palette = c("#00AFBB", "#E7B800"),
+         order = c("x", "y"),
+         ylab = "Score", xlab = "Groups")
```



شکل۱۵: نمودار جعبه‌ای برای دو متغیر نمره- تدریس با تلویزیون و نمره- تدریس در کلاس

با توجه به شکل۱۵ داده‌های نمره تدریس با تلویزیون تقریبا نرمال است اما داده‌های نمره تدریس در کلاس نرمال به نظر نمی‌رسد.

```
> wilcox.test(x,y, paired=TRUE)
      Wilcoxon signed rank test with continuity correction
data:  x and y
V = 15.5, p-value = 0.2393
alternative hypothesis: true location shift is not equal to 0
```

با توجه به مقدار p-value که مقداری بزرگتر از ۰.۰۵ است فرض صفر پذیرفته می‌شود. یعنی توزیع نمره تلویزیونی و توزیع نمره کلاسی اختلافی ندارند.

آزمون من- ویتنی

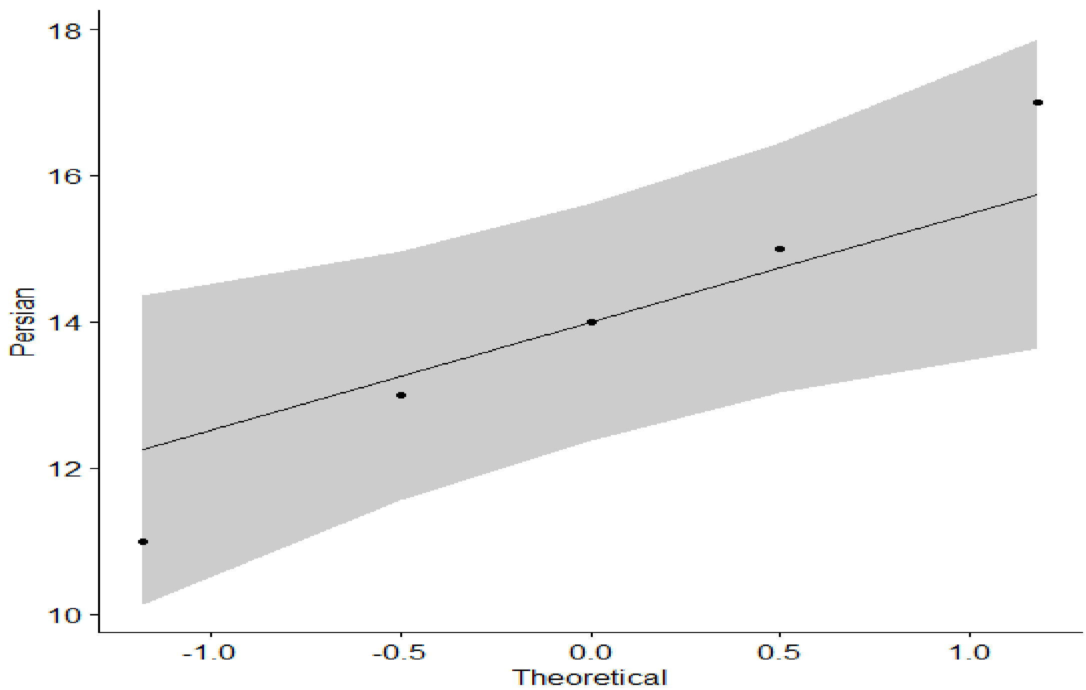
یک استاد ریاضی مدعی است که کتاب فارسی برای ریاضیات عمومی بهتر از کتاب انگلیسی می‌باشد. نمره‌های ۹ دانشجوی سال اول با این دو نوع کتاب در جدول زیر داده شده‌اند.

جدول ۹: نمرات ۹ دانشجو

کتاب فارسی	14	15	13	17	11
کتاب انگلیسی	12	10	16	9	

آیا ادعا را با میزان ده درصد می‌پذیرد؟ (آزمون من-ویتنی به کار ببرید)

```
> Persian<-c(14,15,13,17,11)

> English<-c(12,10,16,9)

> library("ggpubr")

> ggqqplot(Persian, ylab = "Persian")
```



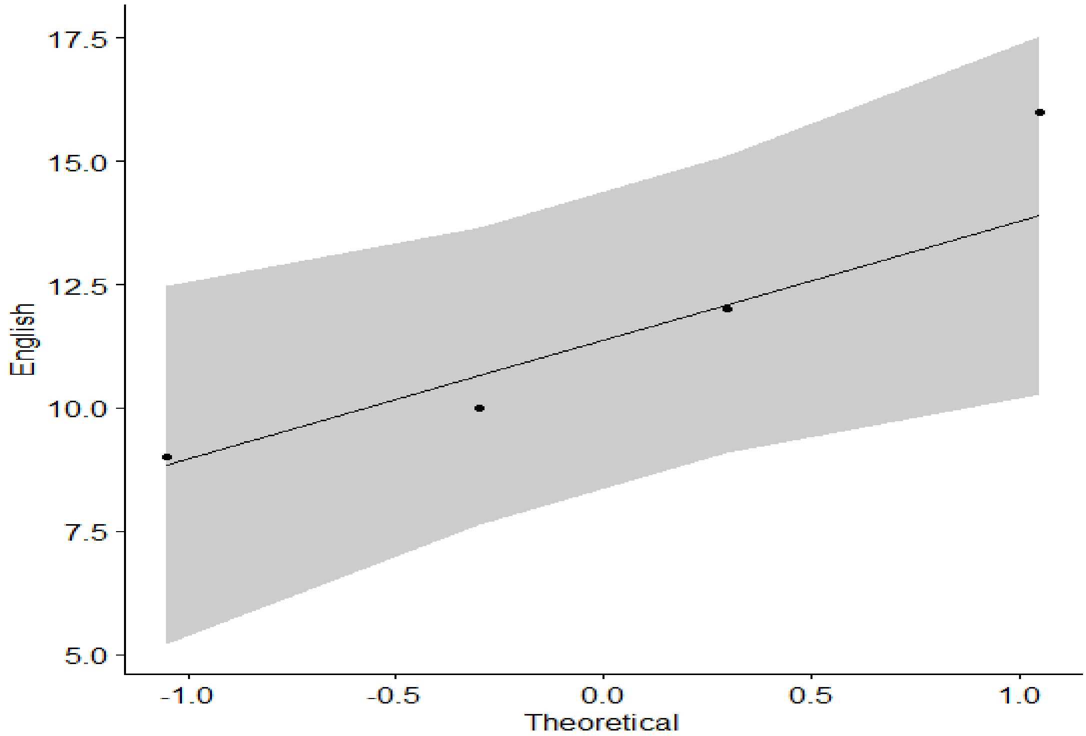
شکل۱۶: نمودار "qqplot" برای نمره کتاب فارسی

شکل ۱۶ برای تشخیص نرمال بودن داده‌های نمره کتاب فارسی رسم شده است. با توجه به شکل نقاط تقریبا روی خط قرار می‌گیرند پس فرض نرمال بودن داده‌ها پذیرفته می‌شود.

```
> shapiro.test(Persian)

	Shapiro-Wilk normality test

data:  Persian
W = 0.99929, p-value = 0.9998
```

در بالا آزمون نرمال بودن برای داده‌های نمره کتاب فارسی صورت گرفته است با توجه به اینکه مقدار p-value بزرگتر از ۰.۰۵ است فرض نرمال بودن داده‌ها پذیرفته می‌شود.

```
> ggqqplot(English, ylab = "English")
```



شکل۱۷: نمودار "qqplot" برای نمره کتاب فارسی

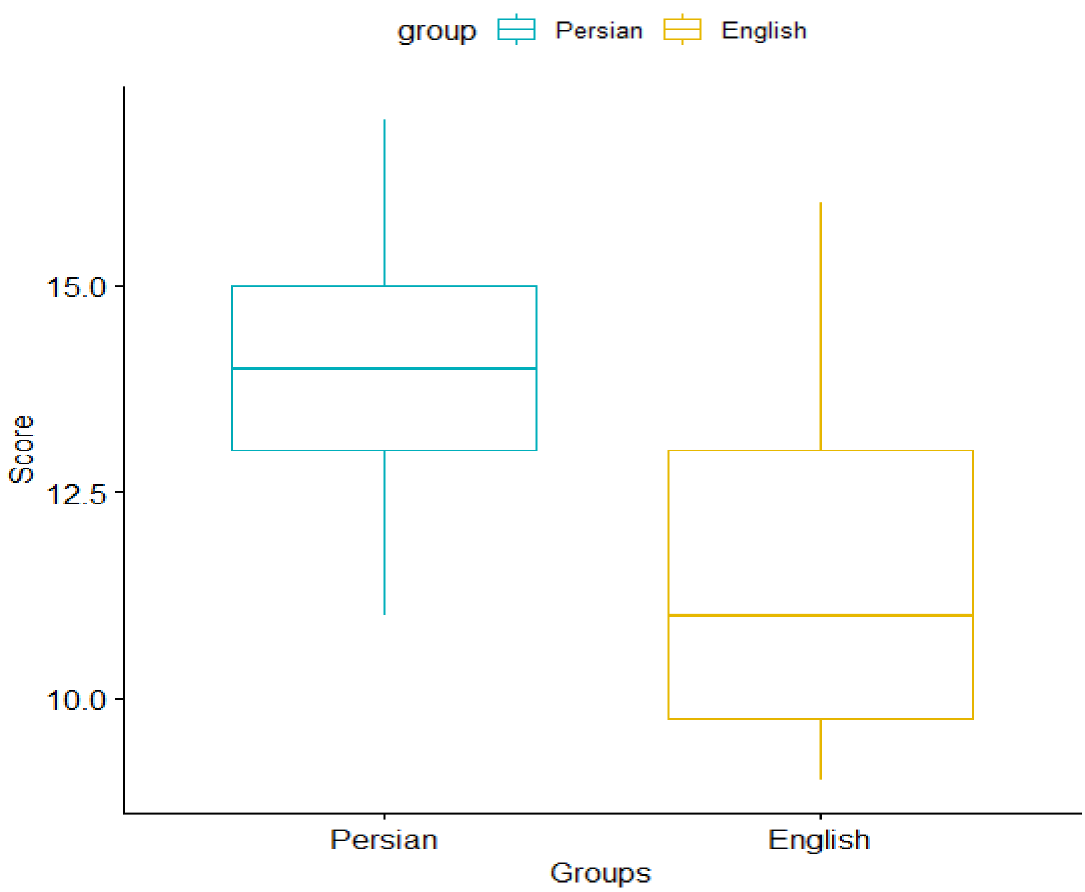
شکل۱۷ برای تشخیص نرمال بودن داده‌های نمره کتاب انگلیسی رسم شده است. با توجه به شکل نقاط تقریبا روی خط قرار می‌گیرند پس فرض نرمال بودن داده‌ها پذیرفته می‌شود.

```
> shapiro.test(English)

	Shapiro-Wilk normality test

data:  English
```

W = 0.9202, p-value = 0.5381

در بالا آزمون نرمال بودن برای داده‌های نمره کتاب انگلیسی صورت گرفته است با توجه به اینکه مقدار p-value بزرگتر از ۰.۰۵ است فرض نرمال بودن داده‌ها پذیرفته می‌شود.

```
> my_data <- data.frame(
+           group = rep(c("Persian", "English"), c(5,4)),
+           Score = c(Persian,English)
+           )
> ggboxplot(my_data, x = "group", y = "Score",
+       color = "group", palette = c("#00AFBB", "#E7B800"),
+       order = c("Persian", "English"),
+       ylab = "Score", xlab = "Groups")
```



شکل۱۸: نمودار جعبه‌ای برای دو متغیر نمره کتاب فارسی و نمره کتاب انگلیسی

با توجه به شکل ۱۸ داده‌های نمره کتاب فارسی و نمره کتاب انگلیسی هر دو نرمال به نظر می‌رسند.

```
> wilcox.test(Persian,English,alternative = "greater",conf.level = 0.9)

	Wilcoxon rank sum test

data:  Persian and English

W = 15, p-value = 0.1429
```

با توجه به مقدار p-value که مقداری بزرگتر از ۰.۰۵ است فرض صفر پذیرفته می‌شود به این معنا که کتاب فارسی برای ریاضیات عمومی بهتر از کتاب انگلیسی است.

آزمون مک نمار

با انتخاب یک نمونه ۶۰ نفری از دانشجویان، به طور تصادفی نظر آنان را در مورد تاثیر یک شیوه آموزشی جویا شده‌ایم. ۳۷ نفر معتقدند که این شیوه آموزشی موثر و ۲۲ نفر معتقدند غیر موثر است. سپس از مزایای این شیوه برای این دانشجویان گفته شد و مجدداً نظر آنان را جویا شدیم .این بار تعداد موافقان به ۲۲ و مخالفان به ۳۱ تغییر کرد. می‌خواهیم بدانیم آیا مزایای ذکر شده در نقطه نظر دانشجویان تفاوت معنی‌داری ایجاد کرده است یا نه؟

```
> colnames(mat) <- rownames(mat) <- c("No", "Yes")

> names(dimnames(mat)) = c("Before", "After")

> mat

     After

Before No Yes

  No  22  38

  Yes 31  29

> margin.table(mat, 1)

Before

 No Yes

 60  60

> margin.table(mat, 2)

After

 No Yes
```

53 67

```
> sum(mat)
[1] 120
> mcnemar.test(mat, correct=FALSE)

	McNemar's Chi-squared test

data:  mat
McNemar's chi-squared = 0.71014, df = 1, p-value = 0.3994
```

به دلیل اینکه مقدار p-value بزرگتر از ۰.۰۵ پس فرض صفر پذیرفته می‌شود به این معنا که بیان مزایا تاثیر معنا داری بر نظر دانشجویان نداشته است.

آزمون کراسکال-والیس

دانشکده‌ای دارای شش رشته تحصیلی مختلف با ۵۰ استاد است. برای ارزیابی روش تدریس اساتید، به تمامی دانشجویان پرسشنامه‌ای داده شده تا به ایشان رتبه دهند. نتیجه رتبه‌بندی اساتید بر اساس نمره های داده شده در جدول زیر در دسترس است .آیا می‌توان گفت که این ۶ رشته از لحاظ رتبه‌بندی استادان یکسانند؟

جدول ۱۰: نتیجه رتبه بندی اساتید

F	E	D	C	B	A
20	3	41	24	2	12
40	15	35	28	21	13
34	22	25	31	27	8
18	23	46	38	39	29
	11	42	49	48	4
	7	43	47	6	14
	1	9	44	26	5
		50	19	45	10
			36	17	16
				37	30
					32
					33

```
> A<-c(12,13,8,29,4,14,5,10,16,30,32,33)
> B<-c(2,21,27,39,48,6,26,45,17,37)
> C<-c(24,28,31,38,49,47,44,19,36)
> D<-c(41,35,25,46,42,43,9,50)
```

```
> E<-c(3,15,22,23,11,7,1)
> F<-c(20,40,34,18)
```

با توجه به اینکه داده‌ها رتبه‌بندی شده‌اند از آزمون ناپارامتری استفاده می‌کنیم:

```
> kruskal.test(list(A,B,C,D,E,F))

	Kruskal-Wallis rank sum test

data:  list(A, B, C, D, E, F)
Kruskal-Wallis chi-squared = 18.744, df = 5, p-value = 0.002145
```

به دلیل اینکه مقدار p-value کوچکتر از ۰.۰۵ پس فرض صفر رد می‌شود یعنی گروه‌ها از لحاظ روش تدریس متفاوت هستند.

آزمون فریدمن

در هشت گروه کاری طی آموزش‌های مکرر سعی شده است به طور یکسان از لحاظ ارائه خدمت مستقیم (X1)، امور اداری (X2)، آموزش (X3) و نظارت (X4) صورت گیرد .حال طی آزمونی به هر یک از این گروه‌ها در چهار زمینه گفته شده نمره داده شده و سپس نمرهء آن‌ها به رتبه تبدیل شده است .می‌خواهیم بررسی کنیم که آیا تفاوت معنی داری میان گروه‌ها از لحاظ ارائه خدمات گوناگون وجود دارد یا نه؟

جدول ۱۱: نمرات آزمون

	X1	X2	X3	X4
1	1	3	2	4
2	2	3	4	1
3	1	3	4	2
4	1	2	3	4
5	4	3	2	1
6	2	3	4	1
7	1	2	3	4
8	2	4	3	1

```
> x1<-c(1,2,1,1,4,2,1,2)
> x2<-c(3,3,3,2,3,3,2,4)
> x3<-c(2,4,4,3,2,4,3,3)
> x4<-c(4,1,2,4,1,1,4,1)
> x <-cbind(x1,x2,x3,x4)
```

با توجه به اینکه داده‌ها رتبه‌بندی شده‌اند از آزمون ناپارامتری استفاده می‌کنیم:

```
> friedman.test(x)
```

```
	Friedman rank sum test

data:  x
Friedman chi-squared = 5.55, df = 3, p-value = 0.1357
```

به دلیل اینکه مقدار p-value بزرگتر از ۰.۰۵ پس فرض صفر پذیرفته می‌شود به این معنا که تفاوت معنی داری میان گروه‌ها از لحاظ ارائه خدمات گوناگون وجود ندارد.

آزمون کوکران

فرض کنید که از شش دانشجوی رشته مدیریت خواسته شود که تصور کنند در پنج موقعیت متفاوت قرار گرفته‌اند و باید بین دو رفتار A (کد ۰) و B (کد ۱) یکی را انتخاب کنند. نتایج می تواند شبیه جدول زیر شود با بررسی جدول به نظر می‌رسد که در بعضی از موقعیت‌ها رفتار B (خان هایی که کد ۱ دارند) بیشتر از سایر موقعی ها انتخاب شده اند. این فرضیه را که تفاوت در رفتارهای انتخاب شده توسط دانشجویان معنی دار است یا خیر را آزمون کنید.

جدول ۱۲: نتایج تصور موقعیت‌ها

	موقعیت1	موقعیت2	موقعیت3	موقعیت4	موقعیت5
دانشجو1	0	0	1	1	1
دانشجو2	0	1	0	1	1
دانشجو3	1	1	1	1	1
دانشجو4	0	0	0	1	1
دانشجو5	0	0	0	0	0
دانشجو6	0	0	0	1	1

```
> library(nonpar)
> s1<-c(0,0,1,0,0,0)
> s2<-c(0,1,1,0,0,0)
> s3<-c(1,0,1,0,0,0)
> s4<-c(1,1,1,1,0,1)
> s5<-c(1,1,1,1,0,1)
> S<-cbind(s1,s2,s3,s4,s5)
> cochrans.q(S, alpha=0.01)

 Cochran's Q Test 

 H0: There is no difference in the effectiveness of treatments. 
 HA: There is a difference in the effectiveness of treatments.
```

Q = 11.6666666666667

Degrees of Freedom = 4

Significance Level = 0.01
The p-value is 0.0200100479145909

به دلیل اینکه مقدار p-value کوچکتر از ۰.۰۵ پس فرض صفر رد می‌شود یعنی تفاوت در رفتارهای انتخاب شده توسط دانشجویان، معنی دار است.

ضریب همبستگی

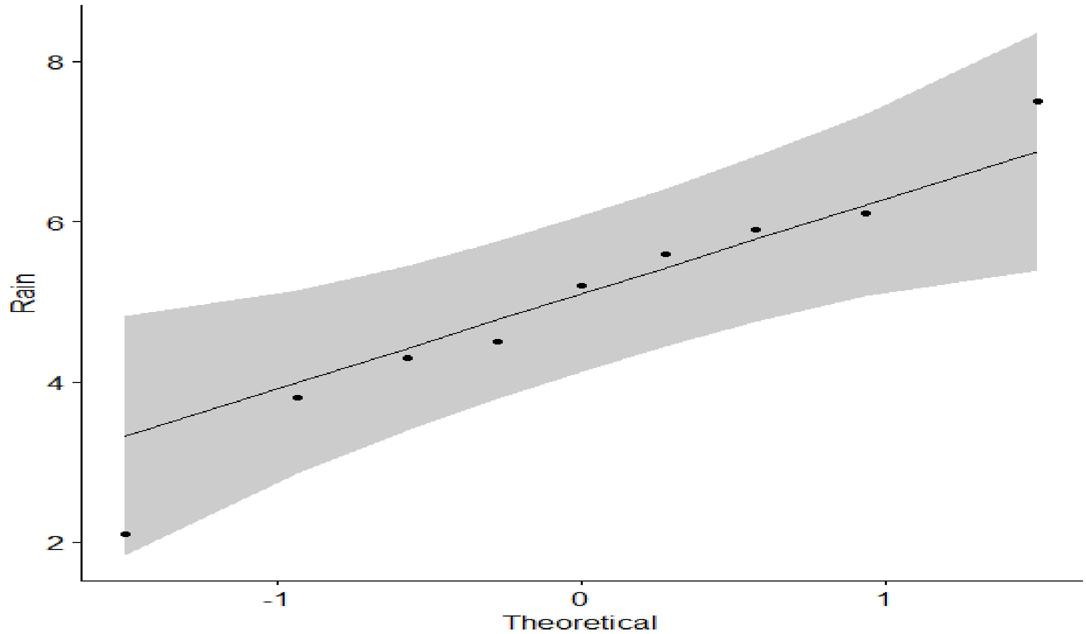
فرض کنید X میزان بارندگی روزانه (واحد ۰.۰۱ سانتی متر) و Y میزان آلودگی هوا (واحد میکروگرم در هر متر مکعب هوا) باشد. با استفاده از داده‌های زیر ضریب همبستگی پیرسون، اسپیرمن، کندال را پیدا کنید.

جدول ۱۳: داده‌های بارندگی

میزان بارندگی	4.3	4.5	5.9	5.6	6.1	5.2	3.8	2.1	7.5
آلودگی هوا	126	121	116	118	114	118	132	141	108

قبل از انجام آزمون، نرمال بودن داده‌ها باید بررسی گردند. برای رسم نمودار "qqplot" برای تشخیص نرمال بودن داده‌ها به صورت صوری از پکیج "ggpubr" استفاده می‌کنیم.

```
> library("ggpubr")
> Rain<-c(4.3,4.5,5.9,5.6,6.1,5.2,3.8,2.1,7.5)
> Pollution<-c(126,121,116,118,114,118,132,141,108)
> data<-data.frame(Rain,Pollution)
> ggqqplot(Rain, ylab = "Rain")
```



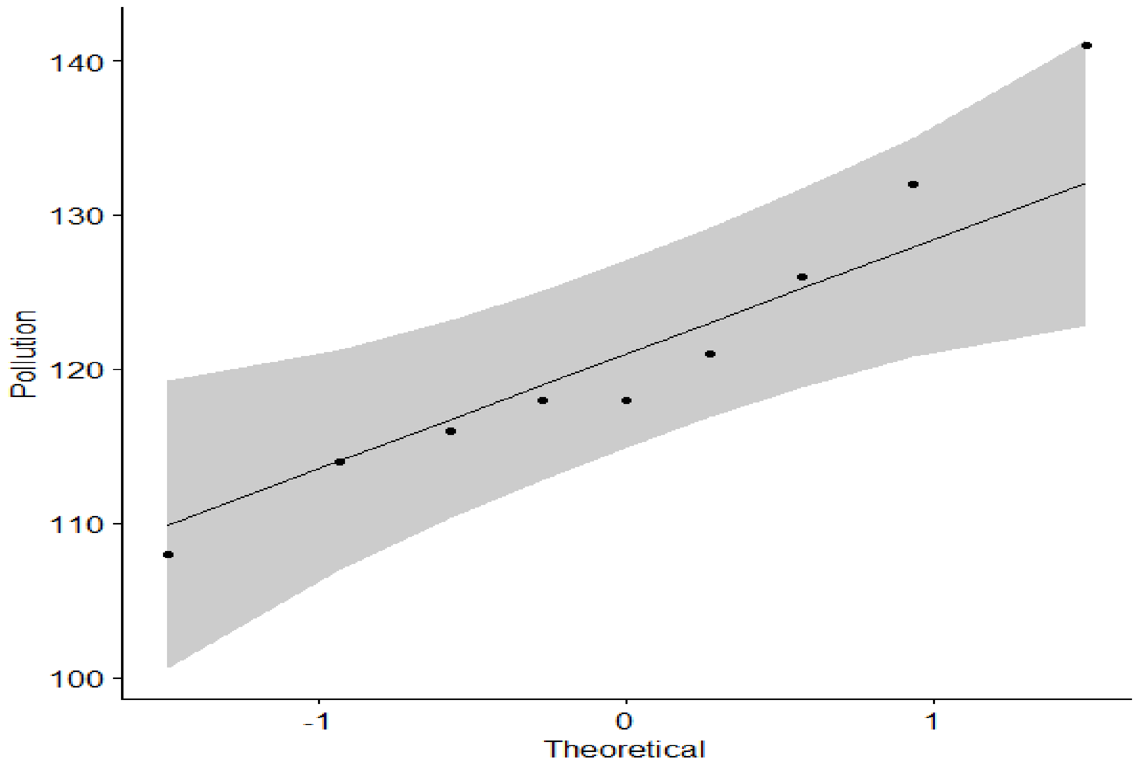
شکل۱۹: نمودار "qqplot" برای میزان بارندگی

شکل ۱۹ برای تشخیص نرمال بودن داده‌های میزان بارندگی رسم شده است. با توجه به شکل نقاط تقریبا روی خط قرار می‌گیرند پس فرض نرمال بودن داده‌ها پذیرفته می‌شود.

برای بررسی دقیق‌تر نرمال بودن داده‌ها از آزمون "shapiro.test" استفاده می‌شود.

```
> shapiro.test(Rain)

	Shapiro-Wilk normality test

data:  Rain

W = 0.98082, p-value = 0.9683
```

در بالا آزمون نرمال بودن برای داده‌های میزان بارندگی صورت گرفته است با توجه به اینکه مقدار p-value بزرگتر از ۰.۰۵ است فرض نرمال بودن داده‌ها پذیرفته می‌شود اما به دلیل اینکه داده‌ها کم هستند می‌توان از آزمون ناپارامتری استفاده کرد.



شکل۲۰: نمودار "qqplot" برای آلودگی هوا

شکل ۲۰ برای تشخیص نرمال بودن داده‌های آلودگی هوا رسم شده است. با توجه به شکل نقاط تقریبا روی خط قرار می‌گیرند پس فرض نرمال بودن داده‌ها پذیرفته می‌شود.

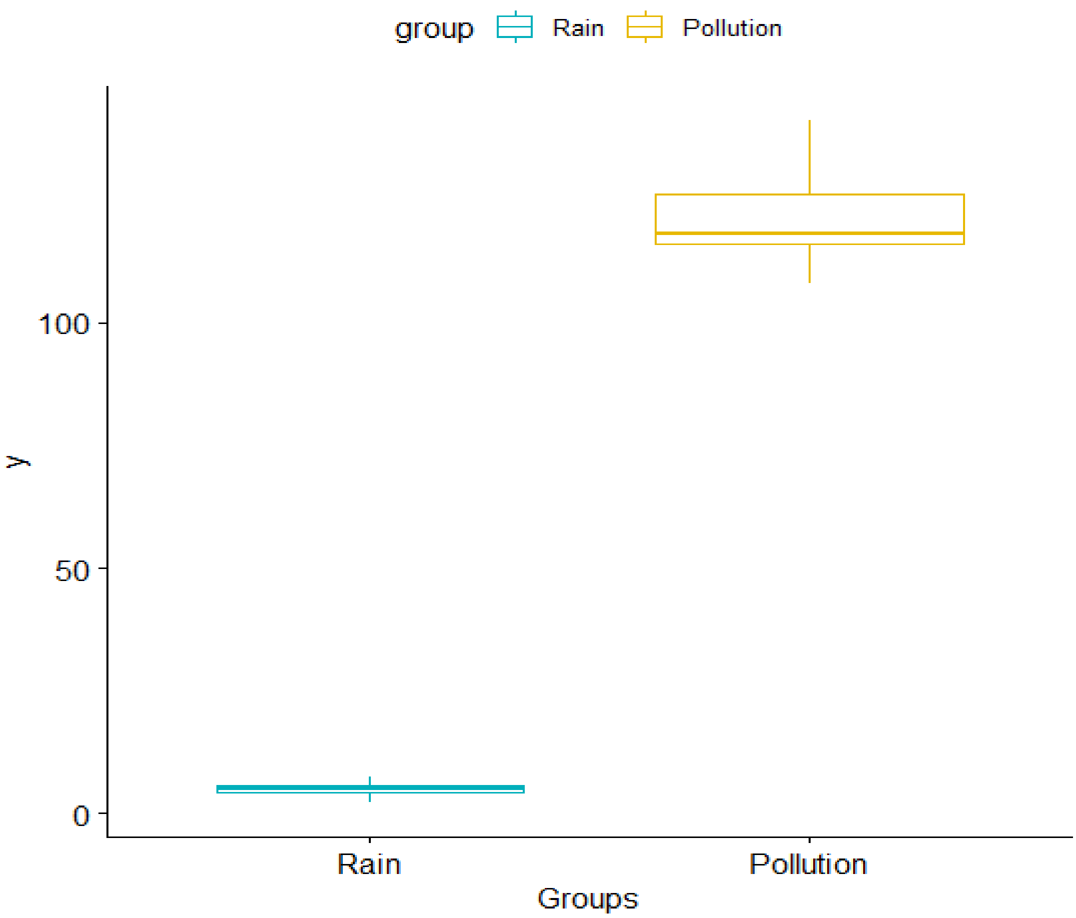
```
> shapiro.test(Pollution)

	Shapiro-Wilk normality test

data:  Pollution
W = 0.9424, p-value = 0.6074
```

در بالا آزمون نرمال بودن برای داده‌های آلودگی هوا صورت گرفته است با توجه به اینکه مقدار p-value بزرگتر از ۰.۰۵ است فرض نرمال بودن داده‌ها پذیرفته می‌شود.

در زیر نمودار جعبه‌ای داده‌های میزان بارندگی و آلودگی هوا برای تشخیص نرمال بودن داده‌ها با پکیج "ggpubr" رسم شده است:

```
> my_data <- data.frame(
+            group = rep(c("Rain", "Pollution"), each = 9),
```

```
+             y = c(Rain,Pollution)
+             )
> library("ggpubr")
> ggboxplot(my_data, x = "group", y = "y",
+         color = "group", palette = c("#00AFBB", "#E7B800"),
+         order = c("Rain", "Pollution"),
+         ylab = "y", xlab = "Groups")
```



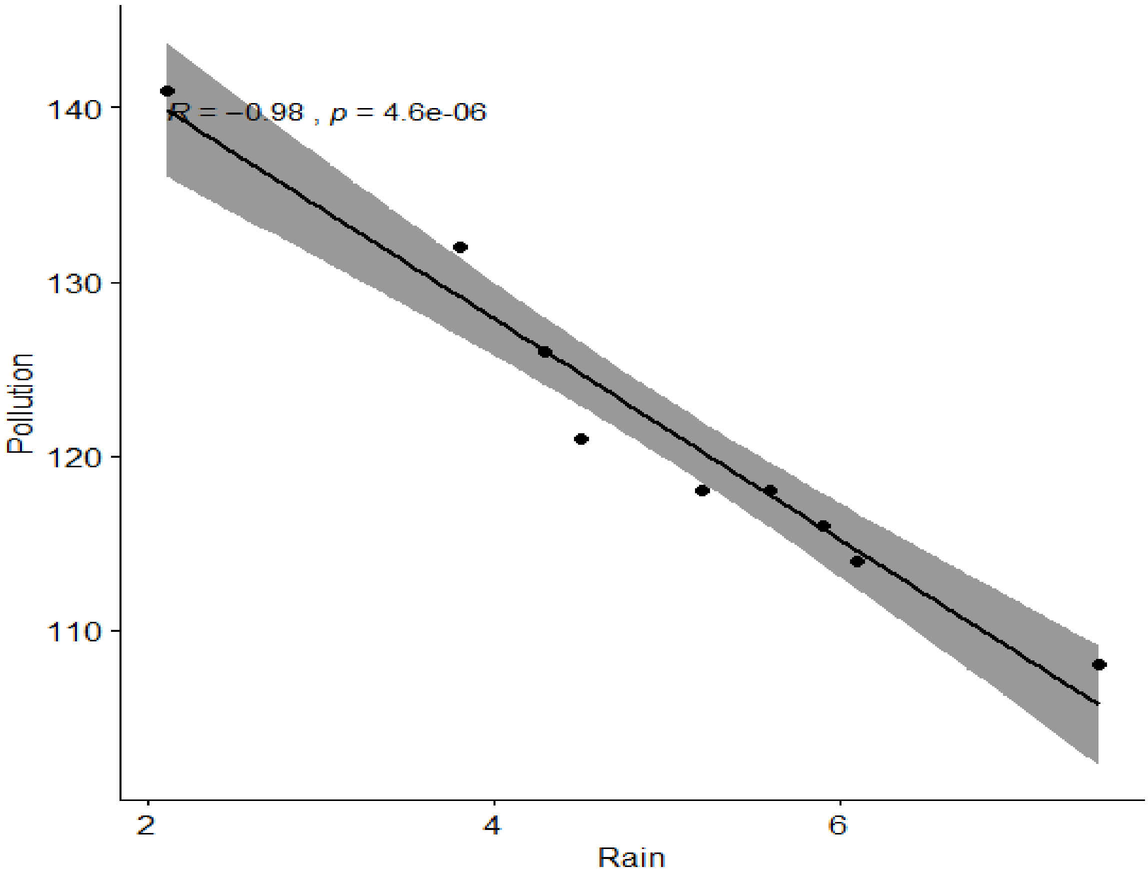
شکل۲۱: نمودار جعبه‌ای برای دو متغیر میزان بارندگی و آلودگی هوا

با توجه به شکل ۲۱ داده‌های میزان بارندگی و هم داده‌های آلودگی هوا تقریبا نرمال هستند.

ضریب همبستگی پیرسون

برای رسم نمودار همبستگی بین دو متغیر به میزان بارندگی و آلودگی هوا از پکیج "ggpubr" استفاده می‌شود.

```
> library("ggpubr")
> ggscatter(data, x = "Rain", y = "Pollution",
+         add = "reg.line", conf.int = TRUE,
+         cor.coef = TRUE, cor.method = "pearson",
+         xlab = "Rain", ylab = "Pollution")
>
```



شکل ۲۲: نمودار همبستگی بین میزان بارندگی و آلودگی هوا

با توجه به شکل ۲۲ بین میزان بارندگی و آلودگی هوا همبستگی بسیاری اما معکوس وجود دارد.

```
> res <- cor.test(Rain, Pollution,
+               method = "pearson")
```

```
> res

	Pearson's product-moment correlation

data:  Rain and Pollution
t = -12.6, df = 7, p-value = 4.579e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9956555 -0.8985433
sample estimates:
       cor 
-0.9786584
```

با توجه به مقدار p-value که مقداری کمتر از ۰.۰۵ است فرض صفر رد می‌شود به این معنا که مقدار همبستگی صفر نیست و دو متغیر میزان بارندگی و میزان آلودگی هوا با یکدیگر رابطه دارند و ضریب همبستگی پیرسون برابر 0.9786584- است.

ضریب همبستگی رتبه‌ای اسپیرمن

```
> res2 <-cor.test(Rain,Pollution,  method = "spearman")
> res2

	Spearman's rank correlation rho

data:  Rain and Pollution
S = 239.5, p-value = 1.542e-08
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho 
-0.9958246
```

با توجه به مقدار p-value که مقداری کمتر از ۰.۰۵ است فرض صفر رد می‌شود به این معنا که مقدار همبستگی صفر نیست و دو متغیر میزان بارندگی و میزان آلودگی هوا با یکدیگر رابطه دارند و ضریب همبستگی اسپیرمن برابر -0.9958246 است.

ضریب همبستگی کندال

```
> res2 <- cor.test(Rain,Pollution,  method="kendall")

> res2

        Kendall's rank correlation tau

data:  Rain and Pollution

z = -3.669, p-value = 0.0002435

alternative hypothesis: true tau is not equal to 0

sample estimates:

      tau

-0.9860133
```

با توجه به مقدار p-value که مقداری کمتر از ۰.۰۵ است فرض صفر رد می‌شود به این معنا که مقدار همبستگی صفر نیست و دو متغیر میزان بارندگی و میزان آلودگی هوا با یکدیگر رابطه دارند و ضریب همبستگی کندال برابر -0.9860133 است.

منابع

[1] Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Probability and Distributions. Journal of Diabetes and Lipid Disorders 201312(2):101-17 [in Persian].

[2] Sheskin DJ. Handbook of parametric and nonparametric statistical procedures Florida: Chapman & HalVCRC; 2004. 11

[3] Box, G.E., and Cox, D.R. (1964), An Analysis of Transformed Data, J. Royal Statistical Society, Ser. B, 39,211-252.

[4] Asghari Jafarabadi M, Mohammadi SM. Statistical Series: An Introduction to Inferential Statistics (Point Estimation, Confidence Interval and HypothesisTesting). Journal of Diabetes and Metabolic Disorders 2013 12(3) 173-192 [In Persian].

[5] عمیدی، ع . نظریه نمونه گیری و کاربردهای آن، جلد اول؛ تهران؛ مرکز نشر دانشگاهی.

[6] John Arbuthnot (1710). "An argument for Divine Providence, taken from the constant regularity observed in the births of both sexes" . Philosophical Transactions of the Royal Society of London. 27 (325–336): 186–190.

[7] Hemelrijk, J. (1952) A theorem on the sign test when ties are present, Proceedings KNAW series a.,

[8] تحلیل داده‌های روانشناسی با برنامه‌ٔ اس‌پی‌اس‌اس، ویرایش‌های ۸، ۹ و ۱۰؛ نیکلا بریس، ریچارد کمپ و رزمری سنلگار، ترجمه‌ٔ خدیجه‌ٔ علی آبادی و سید علی صمدی، انتشارات نیل، تهران، ۱۳۸۲.

[9] Mann, H.B. and Whitney,D.R. (1947), On a test of whether one of two random variable is stochastically larger than other, J. Annals of Math. Stat., Vol 18, 50-60.

[10] سیدمقتدی هاشمی پرست،(۱۳۸۹) آمار ناپارامتری کاربردی، مرکز نشر دانشگاهی.

[11]; Wallis (1952). "Use of ranks in one-criterion variance analysis". Journal of the American Kruskal Statistical Association. 47 (260): 583–621.

[12] Siegel; Castellan (1988). Nonparametric Statistics for the Behavioral Sciences (Second ed.). New York: McGraw–Hill.

[13] سید یعقوب حسینی، آمار ناپارامتریک، انتشارات دانشگاه علامه طباطبایی.

[14] Pearson, Karl; Beeton, M., & Yule, G.U. (1900). "On the Correlation Between Duration of Life and the Number of Offspring," Proceedings of the Royal Society of London, Vol. LXVII, pp. 159–179.

[15] Spearman C (1904). "The proof and measurement of association between two things". American Journal of Psychology. 15 (1): 72–101.

[16] Kendall, M. (1938). "A New Measure of Rank Correlation". Biometrika. 30 (1–2): 81–89.

[17] دکتر خلیل میرزایی (۱۳۹۰)، پزوهش ، پژوهشگری و پژوهشنامه نویسی. چاپ دوم/ تهران.

[18] McNemar, Quinn (June 18, 1947). "Note on the sampling error of the difference between correlated proportions or percentages". Psychometrika. 12 (2): 153–157.