

سری آمار: روش‌های متداول ناپارامتری

محمد اصغری جعفرآبادی^{۱*}، سیده مومنه محمدی^۲

چکیده

در مطالعات علوم پزشکی، موقعیت‌های وجود دارند که امکان استفاده از روش‌های متداول مبتنی بر توزیع نرمال (روش‌های پارامتری) وجود ندارد. هدف از این مقاله ارائه روش‌های متداول ناپارامتری و انجام استنباط مبتنی بر آن‌ها می‌باشد. مبانی و نحوه انجام محاسبات و همچنین کدهای نرم‌افزاری مربوط به روش‌های متداول ناپارامتری و استنباط مبتنی بر آن‌ها برای مثال‌های مطالعات علوم پزشکی به همراه ملاحظات انتخاب آزمون‌های ناپارامتری و کارایی نسبی آن‌ها در مقایسه با آزمون‌های پارامتری متناظر ارائه گردید.

در صورت برقرار نبودن پیش‌فرض‌های آزمون‌های پارامتری، بدون نگرانی برای از دست رفتن کارایی و حتی با کارایی بالاتر، باید از آزمون‌های ناپارامتری استفاده کرد. برای مقایسه متغیر کمی غیرنرمال یا رتبه‌ای بین دو گروه مستقل آزمون من-ویتنی، برای مقایسه متغیر کمی غیرنرمال یا رتبه‌ای بین بیش از دو گروه مستقل آزمون کروسکال-والیس، برای مقایسه یک متغیر کمی غیرنرمال در دو وضعیت وابسته یا گروه‌های همسان شده آزمون رتبه علامت‌دار ویلکاکسون و برای مقایسه یک متغیر رتبه‌ای در دو وضعیت وابسته یا گروه‌های همسان شده آزمون علامت باید استفاده شود. برای هر یک از آزمون‌های فوق، نتایج حاصل از مثال‌های مطالعاتی به‌همراه نحوه انجام محاسبات ارائه شد. برای بررسی رابطه یا اختلاف در انواع مطالعات علوم پزشکی، آزمون‌های ارائه شده با توجه به موقعیت و هدف مطالعه توصیه می‌شود.

واژگان کلیدی: ناپارامتری، آزاد توزیع، کارایی نسبی، من-ویتنی، کروسکال-والیس، ویلکاکسون، علامت

۱- مرکز تحقیقات آموزش علوم پزشکی، گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، تبریز، ایران

۲- گروه علوم تشریحی، دانشکده پزشکی، دانشگاه علوم پزشکی تبریز، تبریز، ایران

*شانی: تبریز، خیابان گلگشت، خیابان عطار نیشابوری، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، گروه آمار و اپیدمیولوژی،

کدپستی: ۵۱۶۶۱۴۷۱۱، تلفن: ۰۲-۰۴۱۱۳۳۵۷۵۸۰، نمابر: ۰۴۱۱۳۳۴۰۶۳۴، پست الکترونیک: asgharimo@tbzmed.ac.ir

تاریخ پذیرش: ۱۳۹۳/۱۲/۲۴

تاریخ درخواست اصلاح: ۱۳۹۳/۱۱/۱۱

تاریخ دریافت: ۱۳۹۳/۱۰/۱۹

مقدمه

روش‌های آماری معرفی شده در مقاله‌های قبلی این سری بر پیش‌فرض نرمال بودن داده‌ها استوار بودند [۱-۴]، تقریب نرمال در آن‌ها استفاده گردید [۷-۵، ۲] و یا این که محاسبات مربوط به فاصله اطمینان در آن‌ها مبتنی بر توزیع نرمال بود. با توجه به خصوصیات این توزیع و خوش تعریف بودن آن از لحاظ ریاضی [۸]، توزیع نرمال مبنای بسیاری از روش‌های آماری می‌باشد. همان‌طور که در بحث آمار توصیفی نیز بدان اشاره شد، نرمال بودن توزیع داده‌ها برای متغیرهای کمی سبب انتخاب میانگین به‌عنوان پارامتر مورد بررسی در داده‌ها می‌شود [۹] که استفاده از میانگین سبب سهولت استنتاج‌های ریاضی در روش‌های آماری و خواص بهینه در آن‌ها نظیر افزایش حساسیت آزمون‌ها می‌گردد. به هر حال،

- همیشه نرمال بودن توزیع داده‌ها را نمی‌توان به‌عنوان 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 به‌صورت رتبه‌ای و اسمی ارزیابی می‌شوند و اگرچه در مورد برخی از این موارد روش‌هایی در سری‌های قبلی ارائه شد [۷، ۳] ولی 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 روش‌های اصطلاحاً روش‌های آزاد توزیع (Distribution-Free) و یا ناپارامتری (Nonparametric) گفته می‌شود.

آزمون‌های 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 آزمون‌های پارامتری گفته می‌شود. در مقابل

آزمون‌های که نیاز به مشخص بودن توزیع احتمالی داده‌ها ندارند به‌دلیل در تقابل قرار گرفتن با آزمون‌های پارامتری به آن‌ها آزمون‌های ناپارامتری گفته می‌شود. عبارت‌های ناپارامتری و آزاد توزیع (که به‌طور معادل برای این شکل از آزمون‌ها به‌کار می‌روند) یکسان نیستند. آزمون ناپارامتری براساس نام آن، قاعداً 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 داده‌ها فرضی در نظر نمی‌گیرد و ممکن است روی پارامتر مشخص‌شده آن توزیع، آزمون را انجام دهد. به هر حال برای متغیرهای کمی، با وجود این تفاوت در نام، آزمون‌های ناپارامتری یا آزاد توزیع، زمانی استفاده می‌شوند که پیش‌فرض اساسی نرمالیتی توزیع داده‌ها به هر دلیلی (نظیر ماهیت داده‌ها، طراحی پیچیده مطالعه یا طرح نمونه‌گیری نامناسب و ..)

برقرار نباشد. کارایی نسبی آزمون‌های ناپارامتری در مقایسه با آزمون‌های معادل پارامتری آزمون‌های پارامتری با در نظر گرفتن پیش‌فرض نرمالیتی داده‌ها انجام می‌شوند و چون با در نظر گرفتن این پیش‌فرض آزمون متمرکز روی پارامتری خاص نظیر میانگین می‌گردد، به آن عنوان پارامتری داده می‌شود. معادل ناپارامتری این آزمون‌ها یعنی آزمون‌هایی که بدون در نظر گرفتن پیش‌فرض در نرمالیتی داده‌ها انجام می‌شوند. زمانی که شرایط آزمون‌های پارامتری برقرار باشند، آزمون‌های پارامتری به دلیل استفاده از اطلاع توزیع (نرمال) در داده‌ها در مقایسه با آزمون‌های ناپارامتری توان بیشتری دارند ولی زمانی که شرایط و یا پیش‌فرض‌های آزمون‌های پارامتری برقرار نباشد، توان و کارایی آزمون‌های ناپارامتری بیشتر هم باشد [۱۰].

آزمون‌های ناپارامتری از لحاظ اجرا و تفسیر ساده‌تر از آزمون‌های پارامتری معادل‌شان نیستند حتی یکی از نقدهایی که به آزمون‌های پارامتری وارد می‌شود این است که این آزمون‌ها جامعه یا فرآیند مشاهده شده را بیش از حد ساده می‌کنند و در واقع آزمون‌های پارامتری نه به دلیل این که بهینه‌تر هستند بلکه به‌دلیل سهولت به‌کارگیری آن‌ها به صورت متداول‌تری استفاده می‌شوند. یکی از مشکلاتی که در استفاده از آزمون‌های ناپارامتری وجود دارد، ذهنیت

منفی کاربران غیر متخصص می‌باشد. تصور عامه بر این است که آزمون‌های ناپارامتری در مقایسه با آزمون‌های پارامتری متناظر آن‌ها خیلی ضعیف‌تر عمل می‌کنند و تلاش آن‌ها بر این است که به هر ترتیب ممکن از آزمون‌های پارامتری استفاده نمایند. البته لازم به تذکر است در صورتی که پس از انجام تبدیل روی داده‌های کمی نرمالیتی آن‌ها برقرار شد باید آزمون‌های پارامتری را به کار برد. ولی به هر حال موقعیت‌های وجود دارند که ممکن است تبدیل روی داده‌ها سبب نرمالیتی آن‌ها نگردد. در چنین موقعیتی باید از آزمون ناپارامتری مناسب استفاده شود.

مسئله ضعیف‌تر عمل نمودن آزمون ناپارامتری به حساسیت یا توان آزمون آن‌ها بر می‌گردد و اغلب یک آزمون ناپارامتری به جای اصل داده‌ها از رتبه‌های متناظر با آن‌ها استفاده می‌کند و این موضوع سبب کاهش اطلاع در داده‌ها و در نتیجه کاهش توان آزمون می‌گردد. حال مسئله اساسی این است که این موضوع تا چه اندازه می‌تواند موثر باشد و آیا در همه شرایط موضوع کاهش اطلاع رخ می‌دهد و یا این که در شرایط خاص باید این موضوع را مد نظر قرار داد. زمانی که شرایط یک آزمون پارامتری برقرار باشد و به اشتباه از یک آزمون ناپارامتری معادل استفاده گردد، این موضوع سبب کاهش توان تصمیم‌گیری می‌شود. به عبارت دیگر، در زمان برقراری نرمالیتی داده‌ها، یک آزمون ناپارامتری در مقایسه با آزمون‌های پارامتری توان کمتری دارند.

مقایسه دو برآورد (نظیر میانگین و میانه) یا دو آزمون فرض (نظیر پارامتری و ناپارامتری) در مقوله استنباط آماری توسط شاخصی به نام کارایی نسبی (Relative Efficiency) سنجیده می‌شود. کارایی نسبی دو برآورد بر اساس نسبت واریانس دو برآورد به ازای حجم نمونه یکسان به دست می‌آید (مثلاً اگر واریانس میانگین در یک نمونه ثابت برابر ۲ و واریانس میانه برابر ۳ باشد، در این صورت کارایی نسبی میانگین نسبت به میانه برابر ۲ به ۳ خواهد بود «واریانس میانه تقسیم بر واریانس میانگین»). کارایی نسبی دو آزمون فرض بر اساس نسبت توان آزمون در دو آزمون فرض به ازای حجم نمونه یکسان

به دست می‌آید.

حتی زمانی که شرایط آزمون‌های پارامتری کاملاً برقرار باشد، آزمون‌های ناپارامتری در مقایسه با آزمون‌های پارامتری معادل آن‌ها، تنها مقدار اندکی توان کمتری دارند. در مقابل اگر شرایط آزمون‌های پارامتری برقرار نباشند فقط آزمون‌های ناپارامتری اعتبار لازم برای بررسی فرضیه‌های پژوهشی را دارند و در چنین موقعیت‌های ممکن است کارایی بیشتری هم نسبت به آزمون‌های پارامتری داشته باشند و نتیجه حاصل از آزمون‌های پارامتری ممکن است گمراه کننده باشد.

نمونه‌ای از کارایی نسبی آزمون‌های ناپارامتری در مقایسه با معادل‌های پارامتری آن‌ها در جدول ۱ آورده شده است. مثلاً کارایی نسبی آزمون من-ویتنی در مقایسه با آزمون t مستقل برای زمانی که شرایط آزمون پارامتری t مستقل برقرار باشد برابر ۹۵/۵٪ است. یعنی زمانی که توزیع داده‌ها نرمال باشد، اگر برای تایید یک اختلاف در آزمون t به ۹۶ نمونه نیاز باشد در همین شرایط برای آزمون من-ویتنی ۱۰۰ نمونه نیاز است.

جدول ۱-آزمون‌های پارامتری و معادل ناپارامتری آن‌ها و مقایسه کارایی نسبی آن‌ها

هدف	آزمون پارامتری و هدف مورد بررسی	آزمون ناپارامتری و هدف مورد بررسی	کارایی نسبی (توزیع داده‌ها نرمال) \$
مقایسه بین دو گروه مستقل	T مستقل: مقایسه میانگین متغیر	من-وینتنی یا جمع رتبه‌های ویلکاکسون: مقایسه یک متغیر کمی	۰/۹۵۵
مقایسه دو وضعیت وابسته نظیر قبل و بعد یا گروه‌های همسان شده	T زوجی: مقایسه میانگین متغیر	رتبه علامت‌دار ویلکاکسون: مقایسه یک متغیر کمی غیرنرمال	۰/۹۵۵
مقایسه بیش از دو گروه مستقل	تحلیل واریانس: مقایسه میانگین متغیر	کروسکال-والیس یا میانه: مقایسه یک متغیر کمی غیرنرمال یا رتبه‌ای	۰/۸۶۴
مقایسه بیش از دو وضعیت وابسته یا بیش از دو گروه همسان شده	تحلیل واریانس با اندازه‌گیری‌های تکراری: مقایسه میانگین متغیر	فریدمن، W کندال: مقایسه یک متغیر کمی غیرنرمال یا رتبه‌ای	$۰/۹۵۵ \times k/(k+1)$ (مثلاً برای $k=3$ برابر ۰/۷۲ می‌باشد).
رابطه بین دو متغیر	همبستگی پیرسون: رابطه بین دو متغیر کمی	همبستگی اسپیرمن یا τ کندال: رابطه بین دو متغیر کمی غیرنرمال یا رتبه‌ای	۰/۹۱
مقایسه دو وضعیت وابسته نظیر قبل و بعد یا گروه‌های همسان‌شده	مقایسه بیش از دو وضعیت وابسته نظیر یا بیش از دو گروه همسان شده	مک نماز: مقایسه یک متغیر اسمی دو حالتی	
بررسی نرمالیتی توزیع داده‌ها		Q ککران: مقایسه یک متغیر اسمی دو حالتی کولموگروف اسمیرنف، شاپیرو-ویلک، شاپیرو-فرانسیا	

منبع: [۱۰]

\$: شایان ذکر است موارد کارایی نسبی آزمون‌های ناپارامتری ذکر شده در مقایسه با معادل‌های پارامتری آن‌ها برای مواقعی است که شرایط نرمالیتی برقرار باشد و زمانی که شرایط نرمالیتی آزمون‌های پارامتری برقرار نباشد، کارایی نسبی آزمون‌های ناپارامتری در مقایسه با معادل‌های پارامتری آن‌ها می‌تواند بزرگتر از ۱ نیز باشد.

ملاحظات دیگر در تفاوت آزمون‌های پارامتری و

ناپارامتری

۱) بحث نرمال بودن توزیع داده‌ها صرفاً برای متغیرهای کمی (فاصله‌ای و نسبتی) مطرح است و در نتیجه آزمون‌های پارامتری که به نرمالیتی توزیع داده‌ها نیاز دارند محدود به این نوع از متغیرهای هستند.

درحالی‌که علاوه بر این نوع از متغیرها، متغیرهای اسمی و رتبه‌ای هم وجود دارند که در این موارد نیز نیاز است روش‌های آماری ناپارامتری به کار رود. به عبارت دقیق‌تر، آزمون‌های ناپارامتری شامل آزمون‌هایی می‌شود که برای متغیرهای کمی غیرنرمال، رتبه‌ای و اسمی به کار می‌رود و پیش فرضی را در مورد توزیع زیربنایی داده‌ها در نظر نمی‌گیرد.

آزمون‌های t یک نمونه‌ای، دونمونه‌ای مستقل و وابسته ممکن می‌باشد.

در ادامه این مقاله، آزمون‌های ناپارامتری معادل متداول‌ترین آزمون‌های پارامتری یعنی آزمون مقایسه میانگین در دو گروه مستقل (آزمون t مستقل) آزمون مقایسه میانگین در دو وضعیت وابسته (آزمون t زوجی) و آزمون مقایسه میانگین‌ها در بیش از دو جامعه مستقل (تحلیل واریانس) ارائه می‌گردد که به ترتیب شامل آزمون من-ویتنی آزمون رتبه- علامت‌دار ویلکاکسون (و علامت) و آزمون کروسکال-والیس می‌باشد (جدول ۱).

آزمون مقایسه دو گروه مستقل (من-ویتنی)

این آزمون در انواع مطالعات علوم پزشکی، برای بررسی اهداف اختصاصی زیر به کار می‌رود:

- ۱) مقایسه یک متغیر کمی غیرنرمال بین دو گروه مستقل
 - ۲) مقایسه یک متغیر رتبه‌ای بین دو گروه مستقل
- از پیش فرض‌های این آزمون، تصادفی بودن نمونه‌ها، مستقل بودن گروه‌ها از یکدیگر و حداقل رتبه‌ای بودن مقیاس اندازه‌گیری متغیر مطالعه است.

مثال ۱: (مقایسه یک متغیر کمی غیرنرمال بین دو گروه مستقل)

در مطالعه‌ای برای بررسی تاثیر مداخله مکمل‌یاری روی بر شاخص‌های التهابی، متغیر hsCRP (High sensitive C-) Reactive Protein) به‌عنوان یکی از متغیرهای اصلی مطالعه بود. در این مورد برای آشنایی با انجام محاسبات بخشی از داده‌ها انتخاب گردیدند که در جدول ۲ ارائه شده است (شایان ذکر است در اصل مطالعه این متغیر به صورت کمی نرمال استفاده شده است [۱۱]). ولی در این مقاله برای استفاده در این آزمون این بخش از داده‌های انتخاب شده پراکندگی زیادی داشته و به‌صورت متغیر کمی غیرنرمال استفاده شده است):

۲) مشاهده می‌شود برخی از کاربران به اشتباه آزمون‌های ناپارامتری را صرفاً مخصوص داده‌های با مقیاس اسمی یا رسته‌ای می‌دانند 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 به‌کارگیری است و از طرف دیگر ملاک استفاده از این آزمون‌ها حجم نمونه نیست و پیش فرض‌های زیربنایی برای انتخاب آزمون تعیین کننده هستند. به عبارت دیگر، اگر حجم نمونه کوچک باشد ولی توزیع داده‌ها برای یک متغیر کمی نرمال باشد، باز هم می‌توان از آزمون‌ها پارامتری برای بررسی رابطه یا اختلاف استفاده نمود.

۳) همچنین ملاحظه می‌شود برخی از کاربران چون حجم نمونه در مطالعه آن‌ها بیشتر از عدد ثابتی (معمولاً ۳۰) می‌باشد به اشتباه با استناد به قضیه حد مرکزی [۲]. آزمون‌های پارامتری را به جای آزمون‌های ناپارامتری برای متغیرهای کمی غیرنرمال به کار می‌برند. در حالی که در قضیه حد مرکزی بیان می‌شود که در صورت داشتن نمونه بزرگ، توزیع میانگین به نرمال میل می‌کند که در این مورد باید به دو نکته دقت کرد؛ اول این که عدد مشخص کننده نمونه بزرگ همیشه عدد ۳۰ نیست و در صورت کمی بودن متغیر و پراکندگی کم و نبودن داده‌های پرت در اصل داده‌ها عدد ۳۰ می‌تواند 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 روی میانگین‌ها تعریف شده باشد (نظیر آزمون‌های t) و برای آزمون‌هایی نظیر تحلیل واریانس و همبستگی قابل به‌کارگیری نیست. بنابراین در شرایط حجم نمونه بالا نمی‌توان معادل پارامتری را برای همه آزمون‌های ناپارامتری به کار 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، به ترتیب استفاده از

از نصف مقدار میانگین بزرگتر است). همچنین بررسی شاخص‌های دیگر ارزیابی نرمالیتی در داده‌ها [۹]: نشان داد که توزیع داده‌ها برای این متغیر در گروه‌های مطالعه نرمال نبودند. بنابراین در این داده‌ها پیش‌فرض نرمالیتی برقرار نبود و در این مورد معادل ناپارامتری این آزمون یعنی آزمون من-ویتنی استفاده خواهد شد. برای ارائه داده‌ها از شاخص‌های میانه (صدک ۲۵ و صدک ۷۵) استفاده می‌شود که نتایج آن در جدول ۴ ارائه شده است:

جدول ۴- میانه و صدک‌های ۲۵ و ۷۵ مقادیر hsCRP در دو گروه مداخله (n=۱۱) و کنترل (n=۱۰)

گروه	میانه	(صدک ۲۵ ، صدک ۷۵)
مداخله	۵/۵	(۴/۹ ، ۱۵/۵)
کنترل	۶/۳	(۵/۴ ، ۱۶/۷)

نتایج آزمون

با توجه به این که انجام محاسبات برای این آزمون وقت‌گیر و خارج از حوصله خوانندگان می‌باشد، بخشی از محاسبات آن به منظور درک عمیق‌تر این آزمون، در ضمیمه ۱ نشان داده شده است و در این مورد به محاسبات نرم‌افزاری برای دستیابی به نتایج اکتفا می‌گردد. نتایج این آزمون نشان داد که بین دو گروه اختلاف معنی‌داری وجود نداشت ($P > 0/05$) (ضمیمه ۱).

نکته:

۱) برای انجام این آزمون از دستور ranksum در نرم‌افزار STATA12 انجام شد، ranksum hsCRP, (by (group)).

۲) در این مثال، اگر آزمون t مستقل به جای رتبه روی اصل متغیر hsCRP انجام شود (ttest hscrp, (by (group))), مقدار Pvalue دو دنباله‌ای آزمون برابر ۰/۶۹۸ (در مقایسه با ۰/۱۹۱ در آزمون من ویتنی) به دست می‌آید که با مقایسه دو مقدار PValue آزمون‌های t مستقل و من-ویتنی، نتیجه می‌شود که توان آزمون من-ویتنی روی این داده‌ها از آزمون t بالاتر بوده است.

جدول ۲- بخشی از داده‌های متغیر hsCRP در دو گروه مداخله (کد ۱) و کنترل (کد ۲)

گروه	hsCRP
۱	۴/۶
۱	۴/۷
۱	۴/۹
۱	۵/۱
۱	۵/۲
۱	۵/۵
۱	۵/۸
۱	۱۵/۱
۱	۱۵/۵
۱	۱۶/۵
۱	۱۷/۲
۲	۵/۲
۲	۵/۳
۲	۵/۴
۲	۵/۶
۲	۶/۲
۲	۶/۳
۲	۱۵/۸
۲	۱۶/۷
۲	۱۷
۲	۱۷/۲

میانگین و انحراف معیار این متغیر در دو گروه به صورت زیر به دست آمد (جدول ۳):

جدول ۳- میانگین و انحراف معیار مقادیر hsCRP در دو گروه مداخله (n=۱۱) و کنترل (n=۱۰)

گروه	میانگین	انحراف معیار
مداخله	۹/۱	۵/۶
کنترل	۱۰/۱	۵/۷

برای بررسی هدف پژوهش مبنی بر مقایسه متغیر hsCRP در دو گروه، به نظر می‌رسد آزمون t مستقل مناسب باشد [۱]. اما با توجه به پراکندگی زیاد داده‌ها در هر یک از گروه‌های مورد بررسی، انحراف معیار داده‌ها نسبت به میانگین آن‌ها بزرگ به دست آمده است (انحراف معیار

مقایسه یک متغیر رتبه‌ای بین دو گروه مستقل

در این مورد دقیقاً مشابه با بخش قبلی (مقایسه متغیر کمی غیر نرمال بین دو گروه) عمل می‌گردد.

مثال ۲:

در مطالعه‌ای به منظور بررسی تاثیر تغییر وضعیت بیماران قلبی بعد از لوله‌گذاری بر روی برخی از پیامدهای آنها به منظور کاهش برخی از عوارض در این بیماران، دو گروه از افراد به ترتیب آزمایشی که برای آنها تغییر وضعیت و استفاده از یک بالش مخصوص استفاده شده بود و گروه کنترل که مراقبت روتین دریافت می‌کردند، یکی از اهداف اختصاصی، مقایسه نمره میزان رضایت آنها (در دامنه ۰ تا ۱۰)، بین گروه‌های آزمایشی و کنترل بود. با توجه به ماهیت رتبه‌ای ارزیابی انجام شده برای مقایسه آنها بین دو گروه از آزمون من-ویتنی استفاده شد [۱۲].. نتایج بعد از انجام مداخله، اختلاف معنی‌دار بین گروه‌های مداخله و کنترل را نشان داد ($P < 0/001$) (جزئیات نتایج نشان داده نشده است).

آزمون مقایسه دو وضعیت وابسته یا گروه‌های

همسان‌شده (رتبه علامت‌دار ویلکاکسون)

این آزمون در انواع مطالعات علوم پزشکی، برای بررسی اهداف اختصاصی زیر به کار می‌رود:

۱) مقایسه یک متغیر کمی غیرنرمال بین دو وضعیت وابسته

۲) مقایسه یک متغیر کمی غیر نرمال بین دو گروه

همسان شده انفرادی (Paired match)

از پیش فرض‌های این آزمون، تصادفی بودن نمونه‌ها و حداقل فاصله‌ای بودن مقیاس اندازه‌گیری متغیر مورد مطالعه است.

زمانی که علاوه بر دو پیش فرض فوق، نرمال بودن توزیع داده‌ها برقرار باشد، آزمون t زوجی برای این منظور به کار می‌رود. حال زمانی که پیش فرض سوم برقرار نباشد معادل ناپارامتری آزمون t زوجی یعنی آزمون رتبه-علامت‌دار ویلکاکسون استفاده خواهد شد.

مثال ۳: (مقایسه یک متغیر کمی غیر نرمال بین دو

وضعیت وابسته قبل و بعد مداخله)

در مطالعه برای بررسی تاثیر مداخله مکمل‌یاری آمگا ۳ بر شاخص‌های التهابی (مثال ۱)، متغیر hsCRP به عنوان یکی از متغیرهای اصلی مطالعه بود. در این مطالعه در دو گروه مداخله و شاهد، در دو مقطع زمانی قبل و بعد از انجام مداخله این متغیر اندازه‌گیری شد [۱۱].. برای مقایسه قبل و بعد مداخله در هر یک از این گروه‌ها نیاز به استفاده از آزمون زوجی است. با توجه به ماهیت غیرنرمال این متغیر، شرایط آزمون t زوجی برای انجام این مقایسه‌ها برقرار نیست و در نتیجه معادل ناپارامتری این آزمون یعنی آزمون رتبه علامت‌دار ویلکاکسون استفاده خواهد شد.

در این مورد برای آشنایی با انجام محاسبات، بخشی از داده‌ها و فقط در گروه مداخله انتخاب گردیدند که در جدول ۵ ارائه شده است:

جدول ۵- بخشی از داده‌های متغیر hsCRP در دو وضعیت

قبل و بعد مداخله در گروه مداخله (n=۱۱)

hsCRP قبل از مداخله	hsCRP بعد از مداخله
۴/۶	۳/۱
۴/۷	۴/۱
۴/۹	۳/۹
۵/۱	۳/۵
۵/۲	۵/۲
۵/۵	۴/۳
۵/۸	۵/۷
۱۵/۱	۱۳/۳
۱۵/۵	۱۵/۲
۱۶/۵	۱۵/۱
۱۷/۲	۱۵/۸

برای ارائه داده‌ها از شاخص‌های میانه (صدک ۲۵ و صدک ۷۵) استفاده می‌شود که نتایج آن در جدول ۶ ارائه شده است:

جدول ۶- میانه و صدک های ۲۵ و ۷۵ مقادیر hsCRP در دو

وضعیت قبل و بعد از مداخله در گروه مداخله (n=۱۱)

وضعیت	میانه (صدک ۲۵ ، صدک ۷۵)
قبل از مداخله	(۴/۹ ، ۱۵/۵)
بعد از مداخله	(۴/۲ ، ۱۴/۵)

نتایج آزمون

با توجه به این که انجام محاسبات برای این آزمون وقت‌گیر و خارج از حوصله خوانندگان می‌باشد، بخشی از محاسبات آن به‌منظور درک عمیق‌تر این آزمون، در ضمیمه ۲ نشان داده شده است و در این مورد به محاسبات نرم‌افزاری برای دستیابی به نتایج اکتفا می‌گردد. نتایج این آزمون نشان داد که بین دو گروه اختلاف معنی‌داری وجود داشت ($P < 0/05$) (ضمیمه ۲).

نکته:

۱) برای انجام این آزمون از دستور signrank در نرم‌افزار STATA12 انجام شد
(`signrank hscrbefore= hscrpafter`).

۲) در این مورد اگر آزمون T زوجی روی متغیر hsCRP انجام شود (`Ttest hscrbefore= hscrpafter`)، در این صورت مقدار Pvalue دو دنباله‌ای آزمون برابر $0/011$ (در مقایسه با $0/004$ در آزمون رتبه علامت‌دار ویلکاکسون) به دست می‌آید که در این مورد با مقایسه دو مقدار PValue آزمون‌های t زوجی و رتبه علامت‌دار ویلکاکسون، نتیجه می‌شود که توان آزمون رتبه علامت‌دار ویلکاکسون روی این داده‌ها از آزمون t زوجی بالاتر بوده است.

۳) برای آزمون رتبه-علامت‌دار ویلکاکسون، حداقل فاصله‌ای بودن مقیاس اندازه‌گیری یک شرط اساسی است و زمانی که این پیش‌فرض نیز برقرار نباشد، آزمون علامت به‌کار خواهد رفت.

آزمون مقایسه یک متغیر رتبه‌ای در دو وضعیت وابسته (آزمون علامت)

این آزمون در انواع مطالعات علوم پزشکی، برای بررسی اهداف اختصاصی زیر به‌کار می‌رود:

- ۱) مقایسه یک متغیر رتبه‌ای بین دو وضعیت وابسته
- ۲) مقایسه یک متغیر کمی رتبه‌ای بین دو گروه همسان شده انفرادی (Paired match)

از پیش‌فرض‌های این آزمون، تصادفی بودن نمونه‌ها و رتبه‌ای بودن مقیاس اندازه‌گیری متغیر مورد مطالعه است.

مثال ۴: (مقایسه یک متغیر کمی رتبه‌ای بین دو وضعیت

وابسته قبل و بعد از مداخله)

در مطالعه برای بررسی تاثیر مداخله مکمل‌یاری آمگا ۳، متغیر BMI (Body Mass Index) (با سطوح کم وزن، نرمال، اضافه وزن و چاق) به‌عنوان یکی از متغیرهای اصلی مطالعه بود. در این مطالعه در دو گروه مداخله و شاهد، در دو مقطع زمانی قبل و بعد از انجام مداخله، این متغیر اندازه‌گیری شد [۱۱]. برای مقایسه قبل و بعد از مداخله در هر یک از این گروه‌ها آزمون علامت با توجه به رتبه‌ای بودن این متغیر استفاده خواهد شد.

در این مورد برای آشنایی با انجام محاسبات بخشی از داده‌ها و فقط در گروه مداخله انتخاب گردیدند که در جدول ۷ ارائه شده است (شایان ذکر است در اصل مطالعه این متغیر به‌صورت کمی استفاده شده است. ولی در این مقاله برای استفاده در این آزمون سطوح BMI استفاده شده است):

جدول ۷- بخشی از داده‌های متغیر سطوح BMI در دو وضعیت قبل و بعد از مداخله در گروه مداخله (n=۱۱)

سطوح BMI قبل از مداخله	سطوح BMI بعد از مداخله
۱	۱
۲	۱
۳	۲
۱	۱
۱	۱
۳	۴
۴	۲
۱	۱
۴	۳
۳	۲
۳	۴

برای ارائه داده‌ها از شاخص‌های فراوانی و درصد استفاده می‌شود که نتایج آن در جدول ۸ ارائه شده است:

جدول ۸- توزیع فراوانی (درصد) سطوح BMI در دو وضعیت قبل و بعد از مداخله در گروه مداخله (n=۱۱)

گروه	کم وزن	نرمال	اضافه وزن	چاق
قبل از مداخله	۴ (۳۶/۴)	۱ (۹/۰)	۳ (۲۷/۳)	۳ (۲۷/۳)
بعد از مداخله	۴ (۳۶/۴)	۲ (۱۸/۲)	۵ (۴۵/۴)	۰ (۰/۰)

نتایج آزمون

با توجه به این که انجام محاسبات برای این آزمون وقت گیر و خارج از حوصله خوانندگان می باشد، بخشی از محاسبات آن به منظور درک عمیق تر این آزمون، در ضمیمه ۳ نشان داده شده است و در این مورد به محاسبات نرم افزاری برای دستیابی به نتایج اکتفا می گردد. نتایج این آزمون نشان داد که بین دو گروه اختلاف معنی داری وجود نداشت ($P = ۰/۱۲۵$) (ضمیمه ۳).

نکته:

برای انجام این آزمون از دستور signtest در نرم افزار STATA12 استفاده شد ($\text{signtest hscrpafter= bmibefore}$).

آزمون مقایسه یک متغیر اسمی در دو وضعیت وابسته (آزمون مک-نمار)

برای مقایسه یک متغیر اسمی دو حالتی در دو وضعیت وابسته، آزمون مک-نمار استفاده می شود که این آزمون به طور مفصل در سری های قبلی ارائه شده است [۷].

آزمون مقایسه بیش از دو گروه مستقل (کروسکال-والیس)

برای مقایسه میانگین یک متغیر کمی نرمال در بیش از دو گروه، از تحلیل واریانس استفاده شد، به عبارت دیگر برای مقایسه میانگین زمانی تحلیل واریانس قابل استفاده بود که پیش فرض های تصادفی بودن نمونه در گروه ها، حداقل فاصله ای بودن متغیر وابسته، نرمال بودن توزیع داده ها در هر یک از گروه ها و همگنی واریانس داده ها در گروه ها برقرار باشد [۱].

در موقعیت های عملی ممکن است که انحراف جدی در پیش فرض های مورد بررسی وجود داشته باشد. در این موارد نیاز است که راه حل های جایگزین برای حل مسئله را مورد استفاده قرار داد. یکی از پیشنهاد های اساسی که در

بحث تحلیل واریانس استفاده از تبدیل داده ها (برای متغیر وابسته کمی) است. ولی استفاده از تبدیل نیز نرمال شدن توزیع داده ها و یا همگنی واریانس آن ها را تضمین نمی کند. همچنین ممکن است متغیر وابسته به صورت رتبه ای ارزیابی شده باشد. در این صورت برای مقایسه پیش از دو جامعه از معادل ناپارامتری تحلیل واریانس، یعنی آزمون کروسکال-والیس استفاده خواهد شد. این آزمون، تعمیمی از آزمون من-ویتنی برای مقایسه دو گروه مستقل است و تقریباً ایده یکسانی برای انجام آزمون در آن وجود دارد. این آزمون در انواع مطالعات علوم پزشکی، برای بررسی اهداف اختصاصی زیر به کار می رود:

۱) مقایسه یک متغیر کمی غیرنرمال بین بیش از دو گروه مستقل

۲) مقایسه یک متغیر رتبه ای بین بیش از دو گروه مستقل از پیش فرض های این آزمون، تصادفی بودن نمونه ها، مستقل بودن گروه ها از یکدیگر و حداقل رتبه ای بودن مقیاس اندازه گیری متغیر مطالعه است.

مثال ۵:

برای بررسی تاثیر مداخله تغییر وضعیت روی میزان درد بعد از لوله گذاری قلبی، دو گروه مداخله فعال (B و C) و یک گروه شاهد (A) در نظر گرفته شد که گروه های فوق به ترتیب $n_a = 7$ و $n_b = 6$ و $n_c = 6$ نفر و در مجموع $n = 21$ حضور داشتند. در ۶ زمان مختلف در گروه های مداخله میزان درد با استفاده VAS ارزیابی شد که در این مقیاس نمرات در دامنه (۰-۱۰) تعریف می شوند [۱۳]. در مقاله حاضر، میانگین ارزیابی های انجام شده درد در ۳ زمان به دست آمد و به عنوان ملاک ارزیابی قرار گرفت. بخشی از داده های این مطالعه در جدول ۹ خلاصه شده است:

جدول ۹-بخشی از داده‌های متغیر درد (pain) در گروه‌های شاهد و مداخله

	(n=۸)A	(n=۷)B	(n=۶)C
	۶/۴	۲/۵	۱/۳
	۶/۸	۳/۷	۴/۱
	۷/۲	۴/۹	۴/۹
	۸/۳	۵/۴	۵/۲
	۸/۴	۵/۹	۵/۵
	۹/۱	۸/۱	۸/۲
	۹/۴	۸/۲	
	۹/۷		

برای ارائه داده‌ها از شاخص‌های میانه (صدک ۲۵ - صدک ۷۵) استفاده می‌شود که نتایج آن در جدول ۱۰ ارائه شده است:

جدول ۱۰- توزیع میانه (صدک ۲۵ - صدک ۷۵) متغیر درد در گروه‌های شاهد و مداخله

گروه	میانه (صدک ۲۵ - صدک ۷۵) متغیر درد
(n=۶)A	۸/۴ (۷/۰ - ۹/۳)
(n=۷)B	۵/۴ (۳/۷ - ۸/۱)
(n=۸)C	۵/۱ (۴/۱ - ۵/۵)

نتایج آزمون

با توجه به این که انجام محاسبات برای این آزمون وقت‌گیر و خارج از حوصله خوانندگان می‌باشد، بخشی از محاسبات آن به‌منظور درک عمیق‌تر این آزمون، در ضمیمه ۴ نشان داده شده است و در این مورد به محاسبات نرم‌افزاری برای دستیابی به نتایج اکتفا می‌گردد. نتایج این آزمون نشان داد که بین دو گروه اختلاف معنی‌داری وجود داشت ($P = ۰/۰۰۷$) (ضمیمه ۴).

در این مورد مشابه با تحلیل واریانس در صورت معنی‌دار شدن آزمون کلی مقایسه بین سه گروه، این نتیجه حاصل می‌شود که حداقل یکی از گروه‌ها با سایر گروه‌ها اختلاف معنی‌داری دارند و برای تعیین عامل اختلاف نیاز است آزمون تعقیبی انجام شود. نکته حائز اهمیت مشابه تحلیل واریانس این است که در این مقایسه‌های دو به دو باید نرخ خطا تصحیح گردد تا پدیده چند آزمونی (Multiple Testing) رخ ندهد. برای داده‌های این مثال،

آزمون تعقیبی با اعمال تصحیح بونفرونی انجام شد و نتایج نشان داد که بین گروه‌های A و B ($P = ۰/۰۲۲$) و بین گروه‌های A و C ($P = ۰/۰۱۴$) اختلاف معنی‌داری مشاهده شد ولی بین گروه‌های B و C ($P = ۱/۰۰۰$) اختلاف معنی‌داری مشاهده نشد.

نکته:

(۱) برای انجام این آزمون از دستور `kwallis` در نرم‌افزار STATA12 استفاده شد (`kwallis pain, by(group)`).

(۲) در این مورد اگر آزمون تحلیل واریانس روی متغیر درد انجام شود (`anova pain group`)، در این صورت مقدار Pvalue آزمون برابر $۰/۰۰۸$ (در مقایسه با $۰/۰۰۷$ در آزمون کروسکال والیس) به دست می‌آید که در این مورد با مقایسه دو مقدار PValue تحلیل واریانس و کروسکال والیس، نتیجه می‌شود که توان آزمون کروسکال والیس روی این داده‌ها اندکی از تحلیل واریانس بالاتر بوده است.

(۳) در مثال فوق، آزمون کروسکال والیس برای مقایسه یک متغیر رتبه‌ای بین بیش از دو گروه مستقل به کار رفت، این آزمون با همین منطق برای مقایسه یک متغیر کمی غیرنرمال بین بیش از دو گروه مستقل نیز به کار می‌رود.

بررسی رابطه بین دو متغیر کمی غیرنرمال و یا رتبه‌ای (همبستگی رتبه‌ای اسپیرمن)

آزمون همبستگی پیرسون برای بررسی رابطه بین دو متغیر کمی نرمال به کار می‌رود. در مقابل اگر مقیاس اندازه‌گیری متغیر مورد بررسی رتبه‌ای باشد و یا اینکه پیش‌فرض نرمالیتی داده‌ها برقرار نباشد، از معادل ناپارامتری همبستگی پیرسون یعنی ضریب همبستگی رتبه‌ای اسپیرمن استفاده خواهد شد که در سری‌های قبلی به‌طور مفصل توضیح داده شده است [۳]. در این مورد به‌طور معادل می‌توان ضریب توافق کندال را نیز به کار برد.

نتیجه‌گیری نهایی

در انواع مطالعات علوم پزشکی، در صورت برقرار نبودن پیش‌فرض‌های آزمون‌های پارامتری، بدون نگرانی برای از

جدول ۱۱- بخشی از داده‌های متغیر درد در دو گروه کنترل (n=۱۰) و مداخله (n=۱۱)

hsCRP	گروه‌های مطالعه
۴/۶	۱
۴/۷	۱
۴/۹	۱
۵/۱	۱
۵/۲	۱
۵/۲	۲
۵/۳	۲
۵/۴	۲
۵/۵	۱
۵/۶	۲
۵/۸	۱
۶/۲	۲
۶/۳	۲
۱۵/۱	۱
۱۵/۵	۱
۱۵/۸	۲
۱۶/۵	۱
۱۶/۷	۲
۱۷	۲
۱۷/۲	۱
۱۷/۲	۲

دست رفتن کارایی، باید از آزمون‌های ناپارامتری (حتی با کارایی بالاتر) استفاده کرد. برای مقایسه متغیر کمی غیرنرمال یا رتبه‌ای بین دو گروه مستقل آزمون من-ویتنی، برای مقایسه متغیر کمی غیرنرمال یا رتبه‌ای بین بیش از دو گروه مستقل آزمون کروسکال والیس، برای مقایسه یک متغیر کمی غیرنرمال در دو وضعیت وابسته یا گروه‌های همسان شده آزمون رتبه علامت‌دار ویلکاکسون و برای مقایسه یک متغیر رتبه‌ای در دو وضعیت وابسته یا گروه‌های همسان شده آزمون علامت، با توجه به موقعیت و هدف مطالعه توصیه می‌شود.

سپاسگزاری

از همکاری صمیمانه دانشکده بهداشت دانشگاه علوم پزشکی تبریز سپاسگزاریم. این مقاله منبع تأمین مالی نداشته است.

ضمیمه‌ها

ضمیمه ۱

برای انجام محاسبات، روی بخشی از داده‌ها (جدول ۲) انتخاب شده‌اند. برای انجا آزمون من-ویتنی دو منطبق و در نتیجه دو روش متفاوت وجود دارد که در ادامه روش متداول‌تر مبتنی بر گام‌های محاسباتی انجام آزمون ارائه خواهد شد.

گام اول: مشاهدات در دو گروه روی هم ریخته می‌شوند و به ترتیب از کوچک به بزرگ مرتب می‌شوند (جدول ۱۱).

گام دوم: به مشاهدات به ترتیب از کوچک به بزرگ رتبه اختصاص داده می‌شود و زمانی که دو مشاهده مشابه به هم وجود داشته باشد به آن‌ها میانگین رتبه‌های آن‌ها اختصاص داده می‌شود که به این اعداد هم‌رتبه یا گره (tie) گفته می‌شود (مثلاً: برای دو عدد ۱۷/۲ که در رتبه‌های ۲۰ و ۲۱ قرار گرفته‌اند، به هر دو عدد میانگین ۲۰ و ۲۱ یعنی ۲۰/۵ اختصاص داده شده است) (جدول ۱۲).

جدول ۱۲- بخشی از داده‌های متغیر درد در دو گروه کنترل (n=۱۰) و مداخله (n=۱۱) به همراه رتبه متناظر

رتبه	hsCRP	گروه
۱	۴/۶	۱
۲	۴/۷	۱
۳	۴/۹	۱
۴	۵/۱	۱
۵	۵/۲	۱
۶	۵/۲	۲
۷	۵/۳	۲
۸	۵/۴	۲
۹	۵/۵	۱
۱۰	۵/۶	۲
۱۱	۵/۸	۱
۱۲	۶/۲	۲
۱۳	۶/۳	۲
۱۴	۱۵/۱	۱
۱۵	۱۵/۵	۱
۱۶	۱۵/۸	۲
۱۷	۱۶/۵	۱
۱۸	۱۶/۷	۲
۱۹	۱۷	۲
۲۰/۵	۱۷/۲	۱
۲۰/۵	۱۷/۲	۲

گام سوم: مشاهدات بر حسب گروه مرتب می‌شوند (جدول ۱۳):

گام چهارم: مجموع و میانگین رتبه‌ها در هر گروه به دست می‌آید (جدول ۱۳):

جدول ۱۳- بخشی از داده‌های متغیر درد در دو گروه کنترل (n=۱۰) و مداخله (n=۱۱) به همراه رتبه متناظر و مجموع و میانگین رتبه‌ها

گروه	hsCRP	rank		
۱	۴/۶	۱		
۱	۴/۷	۲		
۱	۴/۹	۳		
۱	۵/۱	۴		
۱	۵/۲	۵		
۱	۵/۵	۹		
۱	۵/۸	۱۱		
۱	۱۵/۱	۱۴		
۱	۱۵/۵	۱۵		
۱	۱۶/۵	۱۷		
۱	۱۷/۲	۲۰/۵	مجموع	۱۰۱/۵۰
			میانگین	۹/۲۳
۲	۵/۲	۵		
۲	۵/۳	۷		
۲	۵/۴	۸		
۲	۵/۶	۱۰		
۲	۶/۲	۱۲		
۲	۶/۳	۱۳		
۲	۱۵/۸	۱۶		
۲	۱۶/۷	۱۸		
۲	۱۷	۱۹		
۲	۱۷/۲	۲۰/۵	مجموع	۱۲۸/۵۰
			میانگین	۱۲/۸۵

```
.ranksum hscrp, by(group)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
expected ranksum obs group
121 102 11 1
110 129 10 2
231 231 21 combined
unadjusted variance 201.67
adjustment for ties -0.26
adjusted variance 201.40
Ho: hscrp(group==1) = hscrp(group==2)
z = -1.339
Prob > |z| = 0.1806
```

گام پنجم: انجام آزمون t مستقل روی رتبه‌ها به جای داده‌های اصلی (مقاله ۴) (نتایج با اجرای دستور ttest در نرم افزار STATA12 به صورت زیر به دست می‌آید) (ttest (rank, by(group) که در این دستور به جای اصل داده‌ها از رتبه متناظر با آن‌ها (rank به جای hsCRP) استفاده شده است. نتایج نشان داد که بین دو گروه اختلاف معنی‌داری وجود نداشت (P = ۰/۰۵ > ۰/۰۵).
نتایج انجام محاسبات با نرم افزار STATA12 به صورت زیر می‌باشد:

ضمیمه ۲

می‌گردد.

۲) چون جهت یا علامت اختلاف‌ها ربطی به اندازه یا بزرگی اختلاف ندارد در این مرحله علامت‌ها نادیده گرفته می‌شود، یعنی در این مرحله مقادیر قدر مطلق اختلاف‌ها مورد استفاده قرار می‌گیرند.

گام سوم: به اندازه مطلق اختلاف رتبه داده می‌شود، یعنی به کوچک‌ترین اختلاف مطلق عدد ۱ و به بزرگ‌ترین آن‌ها رتبه n (در این مثال ۱۰) داده می‌شود. به علاوه به اختلاف‌های یکسان، میانگین رتبه (هم رتبه) اختصاص داده می‌شود.

گام چهارم: به رتبه‌ها علامت داده می‌شود یعنی رتبه‌هایی که اختلاف متناظر با آن‌ها مثبت بود علامت مثبت و رتبه‌هایی که اختلاف متناظر با آن‌ها منفی بود علامت منفی می‌گیرند (آزمون رتبه علامت‌دار).

گام پنجم: مجموع رتبه‌ها به دست می‌آید (جدول ۱۴).

برای انجام محاسبات، روی بخشی از داده‌ها (جدول ۵) انتخاب شده‌اند. برای انجام آزمون رتبه علامت‌دار ویلکاکسون، گام‌های محاسباتی انجام آزمون به صورت زیر ارائه خواهد شد:

گام اول: اختلاف بین قبل و بعد مداخله به دست می‌آید (جدول ۱۴).

گام دوم: قدر مطلق اختلاف‌های محاسبه شده به دست می‌آید (جدول ۱۴).

نکته:

۱) چون هدف بررسی تغییرات است زوج مشاهداتی که اختلاف آن‌ها صفر می‌شود، چون اطلاع مفیدی برای آزمون فراهم نمی‌کنند کنار گذاشته می‌شوند. در واقع اختلاف صفر نتیجه‌ای است که تحت فرض صفر عدم اختلاف مطرح می‌گردد و اختلاف از این فرضیه در هر یک از دو جهت مثبت یا منفی سبب انحراف از این فرض

جدول ۱۴- بخشی از داده‌های متغیر hsCRP برای قبل و بعد از مداخله به همراه اختلاف و رتبه متناظر و مجموع رتبه‌ها

قبل از مداخله	بعد از مداخله	تفاضل (گام اول)	قدر مطلق تفاضل (گام دوم)	رتبه (گام سوم)	رتبه (گام چهارم)
۴/۶	۳/۱	-۱/۵	۱/۵	۸	-۸
۴/۷	۴/۱	-۰/۶	۰/۶	۳	-۳
۴/۹	۳/۹	-۱	۱	۴	-۴
۵/۱	۳/۵	-۱/۶	۱/۶	۹	-۹
۵/۲	۵/۲	۰	۰	-----	-----
۵/۵	۴/۳	-۱/۲	۱/۲	۵	-۵
۵/۸	۵/۷	-۰/۱	۰/۱	۱	-۱
۱۵/۱	۱۳/۳	-۱/۸	۱/۸	۱۰	-۱۰
۱۵/۵	۱۵/۲	-۰/۳	۰/۳	۲	-۲
۱۶/۵	۱۵/۱	-۱/۴	۱/۴	۷	-۷
۱۷/۲	۱۵/۸	-۱/۴	۱/۴	۶	-۶
				گام (مجموع W =	-۵۵
				n=10, پنجم	

در این مثال می‌تواند اختیار نماید برابر ۵۵- است و آن زمانی است همه اختلاف‌ها منفی باشند (مشابه این مثال). همان‌طور که ملاحظه می‌شود این دامنه یعنی (۵۵ و -۵۵) یک دامنه متقارن حول صفر است و زمانی که مشابه با مثال فوق بار علامت‌های منفی بیشتر باشد علامت W منفی خواهد شد و در مثال حاضر، به این معنی است که امتیازهای حاصل پس از مداخله کوچکتر هستند. در مورد زمانی که تعداد علامت‌های مثبت بیشتر باشند علامت W مثبت خواهد شد. بر اساس فرض صفر نیز W به سمت مقدار صفر نزدیک می‌شود (البته در دامنه تغییرات تصادفی). و در این حالت بین علامت‌های مثبت و

گام ششم: محاسبه شاخص آزمون و تصمیم‌گیری

جمع رتبه‌های علامت‌دار کمیته است که با W نشان داده می‌شود که در این مثال برابر ۵۵- به دست آمد. از طرفی با توجه به اینکه برای دو مشاهده اختلاف صفر برابر به دست آمد، دو مشاهده از تحلیل کنار گذاشته شدند و تعداد مشاهدات نیز برابر ۱۰ شد. مجموع n رتبه متوالی برابر است با:

$$۵۵ = ۲ \div (۱۰ + ۱) \times ۱۰ = n \times (n + ۱) \div ۲ = \text{مجموع}$$

بنابراین ماکسیمم مقدار مثبتی که شاخص W در این مثال می‌تواند اختیار نماید برابر ۵۵ می‌باشد و آن زمانی است که همه اختلاف‌ها مثبت باشند و ماکسیمم مقدار منفی که شاخص W

32.5	65	10	negative
1	1	1	zero
66	66	11	all
unadjusted variance	126.50		
adjustment for ties	-0.13		
adjustment for zeros	-0.25		

adjusted variance	126.13		
Ho: hsrpbefore = hsrpafter			
z =	2.894		
Prob > z =	0.0038		

نکته:

شایان ذکر است فرمول‌های فوق در حجم نمونه‌های بزرگ (بزرگتر از ۱۰) مناسب است حالا اگر حجم نمونه کوچکتر از ۱۰ باشد، پیشنهاد می‌شود که از توزیع دقیق نمونه‌گیری W استفاده شود که مقادیر بحرانی آن در جدول ۱۵ ارائه شده است:

منفی تعادل برقرار می‌شود. با افزایش حجم نمونه، توزیع نمونه‌گیری W به نرمال نزدیک می‌شود، به طوری که برای nهای بزرگتر از ۱۰، تقریب توزیع نرمال آنقدر مناسب است که بتوان شاخص Z نرمال را به صورت زیر تشکیل داد.

$$Z = (W - 0.5) / \sqrt{\{n(n+1)(2n+1)\} / 6}$$

$$= (-55 - 0.5) / \sqrt{\{10 \times (10+1) \times (20+1)\} / 6} = -2/83$$

در مثال فوق، چون به ازای سطح معنی‌داری ۰/۰۵ قدر مطلق Z محاسبه شده از ۱/۹۶ بزرگتر است، در نتیجه بین دو وضعیت قبل و بعد از انجام مداخله تفاوت معنی‌داری وجود داشت (P = ۰/۰۰۴ < ۰/۰۵).

نتایج انجام محاسبات با نرم افزار STATA12 به صورت زیر می‌باشد:

```
. signrank hsrpbefore= hsrpafter
Wilcoxon signed-rank test
expected    sum ranks    obs    sign
32.5         0             0      positive
```

جدول ۱۵- مقادیر بحرانی آزمون رتبه علامت‌دار ویلکاکسون برای نمونه‌های کوچک

	۰/۰۵	۰/۰۲۵	۰/۰۱	۰/۰۰۵	سطح معنی‌داری یک طرفه <
	۰/۱۰	۰/۰۵	۰/۰۲	۰/۰۱	سطح معنی‌داری دو طرفه <
n					
۵	۱۵	--	--	--	
۶	۱۷	۲۱	--	--	
۷	۲۲	۲۴	۲۸	--	
۸	۲۶	۳۰	۳۴	۳۶	
۹	۲۹	۳۵	۳۹	۴۳	

می‌آید (جدول ۱۶).

گام دوم: علامت مثبت (+) و منفی (-) به ترتیب برای اختلاف‌های محاسبه شده مثبت و منفی در نظر گرفته می‌شود (جدول ۱۶) چون هدف بررسی تغییرات است، زوج مشاهداتی که اختلاف آن‌ها صفر می‌شود، چون اطلاع مفیدی برای آزمون فراهم نمی‌کنند کنار گذاشته می‌شوند.
گام سوم: مجموع علامت‌های مثبت محاسبه می‌شود (جدول ۱۶):

به عنوان مثال برای یک آزمون دو دنباله‌ای و برای حجم نمونه ۷، اگر مقدار W بزرگتر از ۲۴ باشد در این صورت اختلاف بین دو وضعیت معنی‌دار خواهد بود.

ضمیمه ۳

برای انجام محاسبات، روی بخشی از داده‌ها (جدول ۷) انتخاب شده‌اند. برای انجام آزمون علامت، گام‌های محاسباتی انجام‌آزمون به صورت زیر ارائه خواهد شد:
گام اول: اختلاف بین قبل و بعد مداخله به دست

جدول ۱۶- بخشی از داده‌های متغیر BMI برای قبل و بعد از مداخله به همراه اختلاف و علامت‌های متناظر

علامت	تفاضل	بعد از مداخله	قبل از مداخله
----	۰	۱	۱
-	-۱	۱	۲
-	-۱	۲	۳
----	۰	۱	۱
----	۰	۱	۱
+	+۱	۴	۳
-	-۲	۲	۴
----	۰	۱	۱
-	-۱	۳	۴
-	-۱	۲	۳
-	-۱	۳	۴
۱	مجموع علامت‌های +		

نمونه گیری X به نرمال نزدیک می‌شود، به طوری که برای nهای بزرگتر از ۱۰، تقریب توزیع نرمال آن قدر مناسب است که بتوان شاخص Z نرمال را به صورت زیر تشکیل داد:

$$Z = (x - 0.5 \times n) \div \sqrt{(n \times 0.5 \times 0.5)} =$$

$$(1 - 3/5) \div \sqrt{(7 \times 0.5 \times 0.5)} = -1.589$$

در مثال فوق، چون به ازای سطح معنی‌داری ۰/۰۵ قدر مطلق Z محاسبه شده از ۱/۹۶ بزرگتر نیست، در نتیجه بین دو وضعیت قبل و بعد از انجام مداخله تفاوت معنی‌داری وجود نداشت (P = ۰/۰۶۷ > ۰/۰۵).

نتایج انجام محاسبات با نرم افزار STATA12 به صورت زیر می‌باشد:

```
. signstest hscrapafter= bmibefore
Sign test
```

expected	observed	sign
3.5	1	positive
3.5	6	negative
4	4	zero
11	11	all

One-sided tests:

Ho: median of bmiafter - bmibefore = 0 vs.

Ha: median of bmiafter - bmibefore > 0

Pr(#positive >= 1) =

Binomial(n = 7, x >= 1, p = 0.5) = 0.9922

Ho: median of bmiafter - bmibefore = 0 vs.

Ha: median of bmiafter - bmibefore < 0

Pr(#negative >= 6) =

Binomial(n = 7, x >= 6, p = 0.5) = 0.0625

Two-sided test:

Ho: median of bmiafter - bmibefore = 0 vs.

Ha: median of bmiafter - bmibefore != 0

Pr(#positive >= 6 or #negative >= 6) =

min(1, 2*Binomial(n = 7, x >= 6, p = 0.5)) = 0.125

گام چهارم: محاسبه شاخص آزمون و تصمیم گیری

جمع‌علامت‌های + کمیتی است که با X نشان داده می‌شود که در این مثال برابر ۱ به دست آمد. از طرفی با توجه به اینکه برای ۴ مشاهده اختلاف صفر برابر به دست آمد دو مشاهده از تحلیل کنار گذاشته شدند و تعداد مشاهدات نیز برابر ۷ شد. با توجه به این که مشاهده تعداد مساوی علامت‌های مثبت و منفی نشان دهنده برابری اختلاف‌های افزایشی و کاهش‌ی بین قبل و بعد از مداخله است، هر چه از این حالت تساوی فاصله ایجاد شود، نشان دهنده اختلاف بین قبل و بعد از مداخله خواهد بود. در این مثال، از تعداد ۷ مشاهده که تغییر در آن‌ها صورت گرفته بود، تعداد ۱ مثبت و ۶ منفی مشاهده شد که برای بررسی معنی‌دار بودن این اختلاف نیاز است احتمال مشاهده ۱ مثبت و یا کمتر و همچنین تعداد ۶ مثبت یا بیشتر، تحت فرض صفر برابری این دو علامت به دست آید که به صورت زیر خواهد بود:

$$P(x \leq 1) = P(x=1) + P(x=0) = ({}^7_1) \times (0.5)^1 \times (0.5)^6$$

$$+ ({}^7_0) \times (0.5)^0 \times (0.5)^7 = 0.055 + 0.0075 = 0.0625$$

$$P(x \geq 6) = P(x=6) + P(x=7) = ({}^7_6) \times (0.5)^6 \times (0.5)^1$$

$$+ ({}^7_7) \times (0.5)^7 \times (0.5)^0 = 0.0075 + 0.005 = 0.0625$$

$$P(x \leq 1) + P(x \geq 6) = 0.0625 + 0.0625 = 0.125$$

در نتیجه بین قبل و بعد از مداخله اختلاف معنی‌داری وجود نداشت (P = ۰/۱۲۵ > ۰/۰۵).

شایان ذکر است فرمول فوق احتمال دقیق بر اساس توزیع دو جمله‌ای را نتیجه می‌دهد. با افزایش حجم نمونه، توزیع

ضمیمه ۴

جدول ۱۸- بخشی از داده‌های متغیر درد برای گروه‌های
مداخله و کنترل (A، B و C) به همراه رتبه‌های متناظر و
مجموع و میانگین رتبه‌ها

گروه	میزان درد	رتبه		
A	۶/۴	۱۱		
A	۶/۸	۱۲		
A	۷/۲	۱۳		
		۳۳		
A	۸/۳	۱۷		
A	۸/۴	۱۸		
A	۹/۱	۱۹		
A	۹/۴	۲۰		
A	۹/۷	۲۱	n	مجموع
			رتبه‌ها	میانگین
			۸	۱۳۱
				۱۶/۴
B	۲/۵	۲		
B	۳/۷	۳		
B	۴/۹	۵/۵		
B	۵/۴	۸		
B	۵/۹	۱۰		
B	۸/۱	۱۴		
B	۸/۲	۱۵/۵	n	مجموع
			رتبه‌ها	میانگین
			۷	۵۸
				۸/۳
C	۱/۳	۱		
C	۴/۱	۴		
C	۴/۹	۵/۵		
C	۵/۲	۷		
C	۵/۵	۹		
C	۸/۲	۱۵/۵	n	مجموع
			رتبه‌ها	میانگین
			۶	۴۲
				۷/۰

گام چهارم: محاسبه تفاضل بین گروهی

آزمون کروسکال-والیس شبیه به تحلیل واریانس SS_{BG} یا مجموع توان دوم انحراف‌های بین گروهی نسبت به میانگین کل محاسبه می‌شود [۱]، اما با این تفاوت که به جای اصل داده از رتبه‌های متناظر با آن‌ها استفاده می‌شود:

$$SS_{BG} = (۱۳۱)^2 \div ۸ + (۵۸)^2 \div ۷ - (۲۳۱)^2 \div ۲۱ = ۳۷۸/۷$$

برای انجام محاسبات، روی بخشی از داده‌ها (جدول ۹) انتخاب شده‌اند. برای انجام آزمون علامت، گام‌های محاسباتی انجام آزمون به صورت زیر ارائه خواهد شد:

گام اول: ابتدا داده‌های صورت گرفته در هر ۳ گروه روی هم ریخته می‌شود و مجموع‌های شامل $n=۲۱$ ارزیابی به وجود می‌آید (جدول ۱۷).

گام دوم: این ۲۱ مشاهده به ترتیب از کوچک‌ترین تا بزرگ‌ترین با اعداد ۲۱۱ رتبه دهی می‌شود و در صورت وجود اعداد مشابه میانگین رتبه‌های متناظر به آن‌ها اختصاص داده شده است (مشاهدات ۴/۹ و ۸/۲) (جدول ۱۷):

جدول ۱۷- بخشی از داده‌های متغیر درد برای گروه‌های
مداخله و کنترل (A، B و C) به همراه رتبه‌های متناظر

گروه	میزان درد	رتبه
C	۱/۳	۱
B	۲/۵	۲
B	۳/۷	۳
C	۱.۴	۴
B	۴/۹	۵/۵
C	۴/۹	۵/۵
C	۵/۲	۷
B	۵/۴	۸
C	۵/۵	۹
B	۵/۹	۱۰
A	۶/۴	۱۱
A	۶/۸	۱۲
A	۷/۲	۱۳
B	۸/۱	۱۴
B	۸/۲	۱۵/۵
C	۸/۲	۱۵/۵
A	۸/۳	۱۷
A	۸/۴	۱۸
A	۹/۱	۱۹
A	۹/۴	۲۰
A	۹/۷	۲۱

گام سوم: نمونه‌ها به همراه رتبه‌های متناظر با آن‌ها از یکدیگر جدا می‌شوند و مجموع و میانگین رتبه‌های مربوط به هر گروه به‌طور جداگانه محاسبه می‌گردد (جدول ۱۸):

نتایج انجام محاسبات با نرم افزار STATA12 به صورت زیر می باشد:

. kwallis pain, by(group)

Kruskal-Wallis equality-of-populations rank test

Rank Sum	Obs	group
131.00	8	A
58.00	7	B
42.00	6	C

chi-squared = 9.836 with 2 d.f.

probability = 0.0073

chi-squared with ties = 9.849 with 2 d.f.

probability = 0.0073

.kwallis pain if group=="A" | group == "B",
by(group)

Kruskal-Wallis equality-of-populations rank test

Rank Sum	Obs	group
86.00	8	A
34.00	7	B

chi-squared = 6.482 with 1 d.f.

probability = 0.0109

chi-squared with ties = 6.482 with 1 d.f.

probability = 0.0109

.kwallis pain if group=="A" | group == "C",
by(group)

Kruskal-Wallis equality-of-populations rank test

Rank Sum	Obs	group
81.00	8	A
24.00	6	C

chi-squared = 7.350 with 1 d.f.

probability = 0.0067

chi-squared with ties = 7.350 with 1 d.f.

probability = 0.0067

.kwallis pain if group=="B" | group == "C",
by(group)

Kruskal-Wallis equality-of-populations rank test

Rank Sum	Obs	group
52.00	7	B
39.00	6	C

chi-squared = 0.184 with 1 d.f.

probability = 0.6682

chi-squared with ties = 0.185 with 1 d.f.

probability = 0.6674

که در آن ۲۳۱ جمع کل رتبه‌ها و ۲۱ تعداد کل نمونه است.

گام پنجم: محاسبه شاخص آزمون (H) و تصمیم‌گیری

شاخص آزمون با فرمول زیر به دست می‌آید:

$$H = SS_{BG} \div (n \times (n + 1) \div 12) =$$

$$378/7 \div (21 \times 22 \div 12) = 9/84$$

به عبارت دیگر، برای محاسبه شاخص آزمون کروسکال-والیس (SS_{BG}(R) به میانگین (مورد انتظار) توزیع نمونه‌گیری خود یعنی (n(n+1)/۱۲) تقسیم می‌شود.

زمانی که تعداد نمونه‌ها بزرگ باشد یعنی تعداد مشاهدات در هر گروه بزرگتر از ۵ باشد در این صورت توزیع نمونه‌گیری شاخص آزمون به توزیع کای دو خیلی نزدیک می‌شود. درجه آزادی توزیع کای دو برابر تعداد گروه‌ها منهای یک (k-1) می‌باشد. حتی زمانی که در هر یک از گروه‌ها تعداد ۳ یا بیشتر مشاهده وجود داشته باشد، این تقریب مناسب خواهد بود. برای تصمیم‌گیری در مورد وجود یا عدم اختلاف بین گروه‌ها باید مقدار H با مقدار جدولی توزیع کای دو با یک اطمینان مشخص نظیر ۰/۹۵ مقایسه گردد.

در این مثال، به‌ازای ۲ (=۳-۱) درجه آزادی، مقدار جدول کای دو برای ۵ درصد خطا (۰/۹۵ اطمینان) برابر ۰/۹۹ می‌باشد که براین اساس فرضیه عدد متفاوت بین گروه‌های مورد بررسی در سطح خطای ۵ رد می‌شود که مشابه با تحلیل واریانس فرض مقابل بیان می‌کند که حداقل یکی از گروه‌ها با سایرین اختلاف معنی‌داری وجود داشت. بنابراین نیاز است آزمون تعقیبی انجام شود که در این مورد هم می‌توان آزمون تعقیبی LSD را با انجام تصحیح بونفرونی روی رتبه‌ها به کار برد [۱]. نتایج نشان داد که با اعمال تصحیح بونفرونی، بین گروه‌های A و B و بین گروه‌های A و C اختلاف معنی‌داری مشاهده شد ولی بین گروه‌های B و C اختلاف معنی‌داری مشاهده نشد.

مآخذ

1. Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Tests for Comparing the Means. *Journal of Diabetes and Metabolic Disorders* 2013;12(4):265-91 [In Persian].
2. Asghari Jafarabadi M, Mohammadi SM. Statistical Series: An Introduction to Inferential Statistics (Point Estimation, Confidence Interval and Hypothesis Testing). *Journal of Diabetes and Metabolic Disorders* 2013; 12(3) 173-192 [In Persian].
3. Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Correlation and Regression. *Journal of Diabetes and Metabolic Disorders* 2013;12(6):479-506 [In Persian].
4. Snedecor GW, Cochran WG. *Statistical Methods*. 8th ed: Iowa State University Press; 1989.
5. Asghari Jafarabadi M, Soltani A, Mohammadi SM. Statistical Series: Analysis of Contingency Tables 2 (Measures of Association). *Journal of Diabetes and Metabolic Disorders* 2014 Under Press [In Persian].
6. Asghari Jafarabadi M, Soltani A, Mohammadi SM. The Analysis of Contingency Tables 3 (Measures of Effect and Agreement). *Journal of Diabetes and Metabolic Disorders* 2014 Under Press [In Persian].
7. Asghari Jafarabadi M, Soltani A, Mohammadi SM. Statistical Series: Analysis of Contingency Tables 1 (Chi Squared Tests). *Journal of Diabetes and Metabolic Disorders* 2014; 13(2):83-101 [In Persian].
8. Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Probability and Distributions. *Journal of Diabetes and Lipid Disorders* 2013;12(2):101-17 [in Persian].
9. Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Summarizing and Displaying Data. *Journal of Diabetes and Lipid Disorders* 2013; 12(2):83-100 [in Persian].
10. Sheskin DJ. *Handbook of parametric and nonparametric statistical procedures* Florida: Chapman & Hall/CRC; 2004.
11. Mohammadi E, Rafraf M, Farzadi L, Asghari-Jafarabadi M, Sabour S. Effects of omega-3 fatty acids supplementation on serum adiponectin levels and some metabolic risk factors in women with polycystic ovary syndrome. *Asian Pacific Journal of Clinical Nutrition* 2012;21(4):511-8.
12. Rezaei Adaryani M, Ahmadi F, Mohamadi E, Asghari Jafarabadi M. The effect of changing position and early ambulation after cardiac catheterization on patients' outcomes- A single-blind randomized controlled trial. *International Journal of Nursing Studies* 2009; 46(8):1035-42.
13. Rezaei-Adaryani R, A Hmadi F, Mohamadi E, Asghari-Jafarabadi M. The effect of three positioning methods on patient outcomes after cardiac catheterization. *Journal of Advanced Nursing* 2009; 65(2):417-24.

STATISTICAL SERIES: COMMON NON-PARAMETRIC METHODS

Mohammad Asghari Jafarabadi*¹, Seyedeh Momeneh Mohammadi²

1. *Medical Education Research Center, Department of Statistics and Epidemiology, Faculty of Health, Tabriz University of Medical Sciences, Tabriz, Iran*
2. *Department of Anatomical Sciences, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran*

ABSTRACT

There are situations in medical studies, wherein it is impossible to use the methods based on normal distribution (parametric methods). This paper objects to introduce common nonparametric methods and the inferences based on the methods in medical studies. Principles and method of calculations along with the software codes for common nonparametric methods and inference based on them were presented taking into account the considerations relevant to choose the nonparametric methods and their relative efficiency with examples in medical studies. In the situation where the assumptions are not satisfied, the nonparametric methods should be used without caution to lose the efficiency or even with higher efficiency of these methods. To compare a non-normal or ordinal variable between two groups Mann-Whitney test, to compare a non-normal or ordinal variable among more than two groups Kruskal-Wallis test, to compare a non-normal variable between two related situations or matched groups Wilcoxon test and to compare an ordinal variable between two related situations or matched groups Sign test should be used. In each of these tests the results of research based examples were presented along with the methods of their calculations. To assess the relation or difference in all types of medical studies, these tests are recommended considering the situation and purpose of study.

Keywords: Nonparametric, Distribution Free, Relative Efficiency, Mann-Whitney, Kruskal-Wallis, Wilcoxon, Sign

* Department of Statistics and Epidemiology, Tabriz University of Medical Sciences, Golgasht St. Attar e Neshabouri St. Faculty of Health, Tabriz University of Medical Sciences, Tabriz, Iran. Postal Code: 5166614711 Tel: 0098411-3357580-2, Fax: 0098411-3340634, Mobile: 0098912-4390958. Email: asgharimo@tbzmed.ac.ir