

سری آمار: آزمون‌های مقایسه میانگین‌ها

محمد اصغری جعفرآبادی^{۱*}، اکبر سلطانی^۲، سیده مومنه محمدی^۳

چکیده

در بسیاری از مطالعات علوم پزشکی، هدف بررسی تاثیر یک مداخله یا ارزیابی اختلاف بین گروه‌های مطالعه برای یک متغیر کمی است. هدف از این مقاله، معرفی مفاهیم و روش‌های ساده و کاربردی روش‌های آماری بررسی اختلاف در این مطالعات می‌باشد.

مبانی و نحوه انجام محاسبات آزمون فرض‌های بررسی اختلاف، نحوه محاسبه فاصله اطمینان و نحوه گزارش کردن یافته‌ها برای مقایسه میانگین با یک عدد ثابت، مقایسه میانگین در دو وضعیت وابسته، مقایسه میانگین در دو و بیش از دو گروه مستقل ارائه گردید.

برای متغیر کمی نرمال، آزمون t یک نمونه‌ای جهت مقایسه میانگین با یک عدد ملاک، آزمون t زوجی برای مقایسه میانگین‌ها در دو وضعیت وابسته (زوجی)، آزمون t مستقل برای مقایسه میانگین در دو گروه مستقل و تحلیل واریانس و آزمون‌های تعقیبی برای مقایسه میانگین در بیش از دو گروه مستقل به کار می‌روند. در هر یک از وضعیت‌های فوق علاوه بر P -Value برای تصمیم‌گیری، اندازه اثر مناسب اختلاف و فاصله اطمینان آن باید گزارش شود. بر اساس مثال‌های مطالعاتی، محاسبات آزمون فرض‌ها و فاصله اطمینان‌های مزبور انجام و نتایج آنها ارائه گردید.

برای بررسی تاثیر مداخله بر، یا بررسی رابطه عوامل خطر با یک متغیر کمی، آزمون‌ها و فواصل اطمینان معرفی شده با توجه به موقعیت و هدف مطالعه توصیه می‌شود.

واژگان کلیدی: مقایسه میانگین‌ها، آزمون t ، تحلیل واریانس

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

۲- مرکز تحقیقات غدد و متابولیسم، پژوهشکده علوم بالینی غدد و متابولیسم، دانشگاه علوم پزشکی تهران، تهران، ایران

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

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

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

تاریخ پذیرش: ۱۳۹۲/۰۲/۲۸

تاریخ درخواست اصلاح: ۱۳۹۲/۰۲/۲۵

تاریخ دریافت: ۱۳۹۲/۰۲/۱۷

مقدمه

در انواع مطالعات علوم پزشکی، موقعیت‌هایی وجود دارند که پیامدهای اصلی مطالعه در قالب متغیرهای عددی کمی (فاصله‌ای و نسبتی) اندازه‌گیری می‌شوند. اهداف اختصاصی مطروحه معمولاً تعیین وضعیت متغیرهای مورد بررسی و ارزیابی رابطه این متغیرها با عوامل خطر (در مطالعات مشاهده‌ای) یا بررسی اختلاف آنها بین گروه‌های مداخله و گواه (در مطالعات مداخله‌ای) است. بنابراین از نظر آماری، بررسی رابطه یا اختلاف مورد نظر است که در این مورد نیاز به انجام آزمون فرض آماری می‌باشد [۱].

در انتخاب آزمون‌های آماری، برقراری برخی از پیش‌فرض‌های آماری در مورد داده‌های مورد بررسی تعیین‌کننده است. یکی از مهمترین آنها برقراری توزیع نرمال (یا عدم آن) در مورد متغیرهای کمی است [۲]. در این مقاله با قبول برقراری پیش‌فرض نرمال درباره داده‌ها، آزمون‌های مزبور ارائه خواهند شد. در این آزمون‌ها، که در آنها با پذیرش برقراری توزیع نرمال درباره داده‌ها، پارامتر مشخصی از جامعه نظیر میانگین مقایسه می‌شود اصطلاحاً **آزمون‌های پارامتری** نامیده می‌شوند. در سری‌های آتی روش‌هایی ارائه خواهد شد که در صورت عدم برقراری توزیع نرمال برای انجام مقایسه‌ها در داده‌ها قابل به کارگیری هستند که اصطلاحاً (به دلیل تقابل در آزمون‌های پارامتری) **آزمون‌های ناپارامتری** نامیده می‌شوند.

برای خلاصه‌سازی متغیرهای کمی نرمال از شاخص‌های میانگین (انحراف معیار) استفاده می‌شود [۳]، به عبارت دیگر متغیری که توزیع نرمال داشته باشد با دو پارامتر میانگین و انحراف معیار مشخص می‌شود [۲]. بنابراین مسئله آماری مورد بررسی در این موقعیت‌های مطالعاتی با اهداف تعیین و بررسی رابطه، به دست آوردن برآورد عددی و فاصله‌ای میانگین (که در این موارد به [۱] مراجعه شود) و انجام آزمون فرض درباره میانگین‌هاست. با توجه به نوع مطالعه و اهداف و فرضیات مورد بررسی در آنها، ممکن است یکی از آزمون فرض‌های زیر مورد استفاده قرار گیرد:

- الف- مقایسه میانگین متغیر کمی نرمال با یک استاندارد (یک عدد ثابت از پیش تعیین شده)
 ب- مقایسه میانگین متغیر کمی نرمال بین دو وضعیت وابسته (زوج مشاهدات)
 ج- مقایسه میانگین متغیر کمی نرمال بین دو گروه مستقل
 د- مقایسه میانگین متغیر کمی نرمال میان بیش از دو گروه مستقل

الف- مقایسه میانگین متغیر کمی نرمال با یک استاندارد (آزمون t یک نمونه‌ای)

بررسی همسانی با استاندارد، هدف برخی از موقعیت‌های عملی در مطالعات علوم پزشکی است، بنابراین زمانی که متغیر مورد بررسی کمی باشد، هدف مقایسه‌ی میانگین مقادیر یک متغیر کمی نرمال با یک عدد معین است. این عدد معین ممکن است در زمینه اختصاصی مورد بررسی به عنوان یک استاندارد وجود داشته باشد و یا ممکن است زمینه تخصصی ایده‌ای درباره این عدد معین ارائه نماید؛ مثلاً مقایسه سلامت عمومی عده‌ای از بیماران با عدد ۲۳ بر اساس پرسشنامه سلامت عمومی. برای مقایسه میانگین مقادیر یک متغیر کمی نرمال با یک عدد معین از آزمون **t** یک نمونه‌ای استفاده می‌شود.

مثال ۱: در مطالعه‌ای با هدف بررسی وضعیت آگاهی، نگرش و عملکرد افراد دیابتی نوع ۲ [۴]، تعداد ۲۰۰ نفر بیمار دیابتی مراجعه کننده به انجمن دیابتی شهر کرج، به تصادف انتخاب شده بودند. برای بررسی آگاهی، تعداد ۱۴ سوال با پاسخ صحیح و غلط در پرسشنامه طراحی شده بود و برای پاسخ‌های صحیح نمره ۱ و برای پاسخ‌های غلط نمره صفر اختصاص داده شده بود بنابراین نمره کل آگاهی که بر حسب مجموع ۱۴ سوال پرسشنامه به دست آمد در دامنه صفر تا ۱۴ قرار می‌گیرد. بنابراین در صورتی که فردی از این تعداد سوال، ۷ مورد را صحیح جواب دهد آگاهی او درباره دیابت در حد متوسط است. این موضوع را می‌توان توسط یک آزمون آماری بررسی نمود که آیا میانگین کل افراد شرکت کننده با حد متوسط اختلاف دارد یا خیر یا به عبارت دیگر آیا میانگین نمره آگاهی در افراد این جامعه با عدد ۷ (عدد ملاک)، اختلاف دارد یا خیر؟

آگاهی افراد در جامعه آماری مورد بررسی به طور معنی داری با ۷ (حد متوسط) اختلاف داشت.

نکته:

- ۱- اگر پراکندگی یا واریانس جامعه مشخص باشد در این صورت می توان به جای t مقدار استاندارد Z را به دست آورد و بر اساس توزیع نرمال، احتمالات مورد نظر را حساب کرد و به سوالات استنباط آماری میانگین پاسخ گفت.
- ۲- اگر فرضیه مطرح شده به این ترتیب بود که «میانگین آگاهی افراد در جامعه آماری مورد بررسی از ۷ (حد متوسط) بزرگتر است»، در این صورت کافی است که تنها سطح زیر منحنی سمت راست $7/81$ را به دست آورد (که تقریباً برابر صفر بود). بنابراین مقدار P -Value برای این آزمون تقریباً برابر صفر به دست می آید و در نتیجه می توان این گونه تصمیم گرفت که میانگین آگاهی افراد در جامعه آماری مورد بررسی به طور معنی داری از ۷ (حد متوسط) بزرگتر بود.

نکته:

برای مثال ۱، فاصله اطمینان ۹۵ درصدی برای میانگین جامعه بر اساس فرمول

$$((\text{خطای معیار}) \times 2 \pm \text{میانگین}) = \text{فاصله اطمینان}$$

$$\text{میانگین [۱]}$$

به صورت

$$8/25 \pm 2 \times (0/16)$$

$$(8/57 \text{ تا } 7/93)$$

بنابراین فاصله اطمینان ۹۵ درصدی برای اختلاف میانگین با عدد ۷ به صورت زیر به دست می آید:

$$(1/57 \text{ تا } 0/93)$$

چون این فاصله اطمینان عدد تحت فرض صفر برای اختلاف یعنی عدد صفر را در بر ندارد بنابراین می توان این گونه تصمیم گرفت که میانگین آگاهی افراد در جامعه

بر اساس نمونه‌ی تصادفی ۲۰۰ نفری از این جامعه، میانگین نمره آگاهی برابر ۸/۲۵ (انحراف معیار ۲/۲۶)، به دست آمد بنابراین مسئله این است، احتمال این که میانگین نمونه با عدد ۷ اختلاف داشته باشد، چقدر است؟ به عبارت دیگر هدف بررسی فرضیه پژوهشی زیر است:

میانگین آگاهی افراد در جامعه آماری مورد بررسی با ۷ (حد متوسط) اختلاف دارد.

اگر به صورت شهودی به این مسئله نگاه کنیم می توان این گونه نتیجه گرفت که عدد ۸/۲۵ از ۷ اختلاف دارد ولی نیاز است این انحراف از لحاظ آماری معنی دار گردد [۱]. خطای معیار این متغیر برابر $0/16$ (۲/۲۶ تقسیم بر رادیکال ۲۰۰) به دست می آید و در یک بررسی ساده معلوم می شود که عدد $1/25$ (یعنی تفاضل ۸/۲۵ از ۷) بیش از سه برابر انحراف معیار میانگین (خطای معیار) از میانگین فاصله گرفته و در نتیجه باید احتمال آن کوچک باشد. حال برای این که این مقدار دقیقاً مشخص گردد نیاز است ابتدا اختلاف استاندارد شده در قالب توزیع t حساب شود که مقدار آن با فرمول زیر به دست می آید:

انحراف معیار توزیع نمونه گیری مربوط \div ((میانگین توزیع نمونه گیری مربوط) - (مقدار مشاهده شده)) = مقدار استاندارد شده (t)

در مورد مثال حاضر:

$$t = (8/25 - 7) \div 0/16 = 7/81$$

بر اساس درجه آزادی ۱۹۹ (۲۰۰ - ۱) و مراجعه به جدول توزیع t (جدول ۱۵)، ملاحظه می شود که مقدار عددی $7/81$ نقطه‌ای را نشان می دهد که در سمت راست آن سطح زیر منحنی تقریباً برابر صفر است، به عبارت دیگر بر اساس یک نمونه تصادفی با تعداد ۲۰۰ مورد، احتمال اینکه میانگین نمونه‌ای بزرگتر از ۷ باشد تقریباً برابر صفر است.

به همین ترتیب سطح زیر منحنی در سمت چپ $7/81$ نیز تقریباً برابر صفر است و در نتیجه، احتمال اینکه میانگین نمونه‌ای بزرگتر از ۷ و یا کوچکتر از آن باشد (آزمون دو دنباله‌ای) نیز تقریباً برابر صفر است. در نتیجه مقدار P -Value برای این آزمون تقریباً برابر صفر به دست می آید و در نتیجه می توان این گونه تصمیم گرفت که میانگین

آماري مورد بررسی به طور معنی‌داری با ۷ (حد متوسط) یافته‌های این بررسی را می‌توان به صورت جدول ۱ ارائه اختلاف داشت.

نحوه ارائه یافته‌ها برای این مطالعه:

جدول ۱- نتایج مقایسه میانگین آگاهی افراد با عدد ۷ (حد متوسط) (n=۲۰۰)

متغیر	میانگین (انحراف معیار)	اختلاف میانگین با عدد ۷ (فاصله اطمینان ۹۵٪)	P-Value
آگاهی	۸/۲۵ (۲/۲۶)	* ۱/۲۵ (۱/۵۷ تا ۰/۹۳)	< ۰/۰۰۱

آزمون t یک نمونه‌ای

*: P < ۰/۰۰۵

دامنه ممکن برای آگاهی: صفر تا ۱۴

دو عددی که برای هر مشاهده وجود داشت یک عدد حاصل می‌شود و در نتیجه مسئله مورد بررسی شبیه به آزمون t یک نمونه‌ای می‌شود. اگر هدف بررسی مساوی بودن میانگین‌های قبل و بعد باشد یعنی ادعای مطرح شده به این صورت باشد که میانگین بعد نسبت به قبل از اجرای مداخله تغییر نموده است (کاهش و یا افزایش یافته است) در این صورت فرض صفر آزمون این است که تفاضل مشاهدات برابر صفر است و فرض مقابل در این آزمون بیان می‌کند که این تفاضل برابر صفر نیست (کوچکتر و یا بزرگتر از صفر است). به علاوه پیش‌فرض‌هایی که در مورد آزمون t یک نمونه‌ای بیان شد در این جا نیز باید برقرار باشند به عبارت دیگر متغیر مورد بررسی باید کمی پیوسته باشد و تفاضل‌های زوج مشاهدات از توزیع نرمال تبعیت کنند.

مثال ۲:

در یک مطالعه مداخله‌ای با هدف بررسی آموزش از راه دور روی دانش، آگاهی و نگرش افراد، اثر این مداخله روی پروفایل‌های لیبیدی، به عنوان متغیرهای ثانوی در مطالعه نیز بررسی گردید. جدول ۲، مقادیر اندازه‌گیری‌های قبل و بعد از مداخله را برای متغیر کلسترول در تعداد ۴۳ نفر از بیماران نشان می‌دهد.

ب- مقایسه میانگین متغیر کمی نرمال بین دو وضعیت وابسته (زوج مشاهدات) (آزمون t زوجی)
آزمون‌های زوجی زمانی استفاده می‌شوند که داده‌ها بر اساس یک نمونه در دو وضعیت وابسته (زوجی) حاصل شده باشند نظیر:

- آزمون‌های قبل و بعد (برای بررسی اختلاف بین قبل و بعد در هر یک از گروه‌ها و مخصوصاً در مطالعات تک گروهی)
- دو نمونه همسان شده (کارآزمایی‌های بالینی یا مطالعات مورد-شاهدی همسان‌شده فرد به فرد و مطالعه در زمینه اندام‌های زوجی)
- مطالعات کارآزمایی متقاطع
- مقایسه دو مفهوم نظیر وضعیت موجود و وضعیت مطلوب که توسط مجموعه یکسانی از افراد ارزیابی می‌شوند.

زمانی که متغیر مورد بررسی کمی نرمال باشد برای مقایسه متغیر مورد بررسی در دو وضعیت زوجی یا وابسته آزمون t زوجی استفاده می‌شود. در این روش، از تفاضل‌های بین زوج مشاهدات استفاده شده و درباره تفاضل میانگین‌ها در جامعه استنباط انجام می‌شود. زمانی که تفاضل بین مقادیر قبل و بعد محاسبه می‌شود، به ازای

جدول ۲- اندازه‌گیری‌های قبل و بعد از مداخله برای متغیر کلسترول (n=۴۳)

شماره بیمار	کلسترول قبل از مداخله	کلسترول بعد از مداخله	تفاضل مشاهدات (بعد منهای قبل)	شماره بیمار	کلسترول قبل از مداخله	کلسترول بعد از مداخله	تفاضل مشاهدات (بعد منهای قبل)
۱	۲۵۰	۲۴۱	-۹	۲۲	۱۸۰	۱۶۸	-۱۲
۲	۲۴۰	۱۷۳	-۶۷	۲۳	۱۷۰	۱۵۰	-۲۰
۳	۱۳۵	۱۷۹	۴۴	۲۴	۱۶۵	۱۵۲	-۱۳
۴	۱۲۰	۱۲۰	۰	۲۵	۱۵۴	۱۰۴	-۵۰
۵	۱۶۹	۱۵۰	-۱۹	۲۶	۱۷۱	۱۵۰	-۲۱
۶	۱۶۸	۱۵۲	-۱۶	۲۷	۱۴۵	۱۳۵	-۱۰
۷	۱۷۸	۱۷۵	-۳	۲۸	۱۸۰	۱۳۲	-۴۸
۸	۱۸۸	۲۰۹	۲۱	۲۹	۲۱۲	۱۹۲	-۲۰
۹	۲۳۸	۱۸۶	-۵۲	۳۰	۱۶۵	۱۳۷	-۲۸
۱۰	۲۵۱	۱۷۹	-۷۲	۳۱	۱۴۲	۱۵۸	۱۶
۱۱	۱۸۲	۱۶۱	-۲۱	۳۲	۱۲۲	۱۱۹	-۳
۱۲	۱۶۱	۱۵۲	-۹	۳۳	۲۱۰	۱۱۷	-۹۳
۱۳	۱۳۰	۲۱۰	۸۰	۳۴	۲۵۴	۱۵۸	-۹۶
۱۴	۲۰۷	۱۹۰	-۱۷	۳۵	۱۹۱	۲۲۹	۳۸
۱۵	۱۴۴	۱۴۰	-۴	۳۶	۱۴۶	۱۳۲	-۱۴
۱۶	۱۲۲	۱۱۸	-۴	۳۷	۱۵۷	۱۱۶	-۴۱
۱۷	۱۵۰	۱۱۵	-۳۵	۳۸	۱۹۸	۱۹۰	-۸
۱۸	۲۴۹	۲۴۵	-۴	۳۹	۲۰۲	۱۸۹	-۱۳
۱۹	۱۶۶	۱۵۰	-۱۶	۴۰	۱۵۰	۱۳۷	-۱۳
۲۰	۲۸۴	۱۸۰	-۱۰۴	۴۱	۱۷۳	۱۶۰	-۱۳
۲۱	۲۴۵	۲۳۸	-۷	۴۲	۱۸۳	۱۹۲	۹

در ستون آخر تفاضل مشاهدات (بعد منهای قبل) ارائه شده است. در واقع در آزمون t زوجی، هدف مقایسه مشاهدات ستون آخر با عددی ثابت نظیر صفر است که در این صورت فرضیه مورد بررسی را می‌توان به شکل زیر بیان کرد:

برای میانگین کلسترول، بین قبل و بعد از انجام مداخله اختلاف وجود دارد.

یا به عبارت دیگر:

تفاضل مشاهدات کلسترول (بین قبل و بعد از انجام مداخله)، با صفراختلاف دارند.

به دلیل این که محاسبات مربوط فقط روی این تفاضل‌ها صورت می‌گیرد، بنابراین پس از محاسبه تفاضل‌ها سایر

محاسبات دقیقاً مشابه آزمون t یک نمونه‌ای است. مقادیر میانگین، انحراف معیار و خطای معیار به ترتیب برابر ۱۴/۹۳، ۳۹/۹۱ و ۶/۰۹ به دست می‌آید. در مرحله بعدی باید شاخص آزمون و مقدار آن مشخص گردد که به صورت زیر می‌باشد:

$$\begin{aligned} & \text{(جذر تعداد نمونه} \div \text{انحراف معیار نمونه‌ای تفاضل} \\ & \text{مشاهدات)} \div \text{(میانگین نمونه‌ای تفاضل مشاهدات)} = t \\ & \text{(خطای معیار تفاضل مشاهدات)} \div \text{(میانگین نمونه‌ای} \\ & \text{برای تفاضل مشاهدات)} = \\ & = (6/09) \div (14/93) \\ & = 2/45 \end{aligned}$$

۲/۴۵- (تقریباً برابر ۰/۰۰۹) و در نتیجه، احتمال اینکه اختلاف میانگین‌های زوج مشاهدات کوچکتر از صفر از آن باشد (آزمون یک دنباله‌ای) تقریباً برابر ۰/۰۰۹ است. در نتیجه مقدار P-Value برای این آزمون تقریباً برابر ۰/۰۰۹ به دست می‌آید و در نتیجه می‌توان این گونه تصمیم گرفت که میانگین کلسترول، بعد از انجام مداخله به طور معنی‌داری نسبت به قبل آن کاهش داشت ($P < 0/05$).

نکته:

برای مثال ۲، فاصله اطمینان ۹۵ درصدی برای میانگین جامعه بر اساس فرمول

$$(\text{خطای معیار تفاضلها}) \times 2 \pm \text{میانگین تفاضل}$$

مشاهدات) = فاصله اطمینان میانگین اختلاف

به صورت

$$14/93 \pm 2 \times (6/09)$$

$$(27/11 \text{ تا } 2/75)$$

چون این فاصله اطمینان عدد تحت فرض صفر برای اختلاف یعنی عدد صفر را در بر ندارد، بنابراین می‌توان این گونه تصمیم گرفت که برای میانگین کلسترول، بین قبل و بعد از انجام مداخله اختلاف معنی‌داری وجود داشت.

نحوه‌ی ارائه یافته‌ها برای این مطالعه:

یافته‌های این بررسی را می‌توان به صورت جدول ۳ ارائه کرد:

جدول ۳- نتایج مقایسه کلسترول بین قبل و بعد از انجام مداخله ($n=43$)

متغیر	قبل از مداخله	بعد از مداخله	اختلاف میانگین‌ها (فاصله اطمینان ۹۵٪)	درصد تغییرات بعد نسبت به قبل #	P-Value
کلسترول (mg/dl)	۱۶۵/۹۵ (۳۸/۱۸)	۱۸۰/۸۸ (۴۱/۴۷)	* ۱۴/۹۳ (۲/۷۵ تا ۲۷/۱۱)	٪ ۹/۰۰	۰/۰۱۸

آزمون t زوجی

* $P < 0/05$

میانگین (انحراف معیار) گزارش شده است.

درصد تغییرات نسبت به قبل با فرمول (اختلاف میانگین‌ها + میانگین قبل از مداخله $\times 100$) به دست می‌آید و تغییرات (اندازه اثر) نسبی (نسبت به قبل مداخله) نامیده می‌شود. در مقابل، اختلاف میانگین‌ها تغییرات (اندازه اثر) مطلق نامیده می‌شود.

درجه آزادی در این آزمون برابر ۴۲ (یعنی ۴۳ منهای ۱) است و همچنین این آزمون با توجه به نحوه بیان ادعا (به صورت بدون جهت)، دو دنباله‌ای است. با مراجعه به جدول توزیع t (جدول ۱۵)، ملاحظه می‌شود که مقدار عددی ۲/۴۵ نقطه‌ای را نشان می‌دهد که در سمت راست آن سطح زیر منحنی تقریباً برابر ۰/۰۰۹ است، به عبارت دیگر، بر اساس یک نمونه تصادفی ۴۳ تایی، احتمال اینکه اختلاف میانگین‌های زوج مشاهدات بزرگتر از صفر باشد تقریباً برابر ۰/۰۰۹ است. به همین ترتیب سطح زیر منحنی در سمت چپ ۲/۴۵ - نیز تقریباً برابر ۰/۰۰۹ است و در نتیجه، احتمال اینکه اختلاف میانگین‌های زوج مشاهدات کوچکتر از صفر و یا بزرگتر از آن باشد (آزمون دو دنباله‌ای) تقریباً برابر ۰/۰۱۸ است. در نتیجه مقدار P-Value برای این آزمون تقریباً برابر ۰/۰۱۸ به دست می‌آید و در نتیجه می‌توان این گونه تصمیم گرفت که برای میانگین کلسترول، بین قبل و بعد از انجام مداخله اختلاف معنی‌داری وجود داشت ($P < 0/05$).

نکته:

۱- اگر پراکندگی یا واریانس جامعه مشخص باشد در این صورت می‌توان به جای t مقدار استاندارد Z را به دست آورد و بر اساس توزیع نرمال، احتمالات مورد نظر را حساب کرد.

۲- اگر فرضیه مطرح شده به این ترتیب بود که «میانگین کلسترول، بعد از انجام مداخله نسبت به قبل آن کاهش می‌یابد»، در این صورت کافی است که تنها سطح زیر منحنی سمت چپ

بین میانگین گروه‌ها اختلاف وجود دارد (میانگین یک گروه در مقایسه با گروه دیگر بزرگتر یا کوچکتر است). در این صورت فرض صفر آزمون این است که اختلاف بین میانگین‌ها در دو جامعه برابر صفر است و فرض مقابل در این آزمون بیان می‌کند که این اختلاف برابر صفر نیست (کوچکتر و یا بزرگتر از صفر است). علاوه بر پیش‌فرض‌های تصادفی بودن نمونه‌ها، کمی و نرمال بودن متغیر مورد بررسی، که در مورد آزمون‌های t یک نمونه‌ای و t زوجی بیان شد، نیاز است دو گروه از هم مستقل نیز باشند. این موضوع ایده بررسی توزیع نمونه‌گیری اختلاف‌های نمونه‌ای را فراهم می‌کند. بنابراین برای انجام این آزمون نیاز است توزیع اختلاف یا تفاضل میانگین در دو گروه مشخص باشد [۱].

مثال ۳:

در یک مطالعه مداخله‌ای با هدف بررسی آموزش از راه دور روی دانش، آگاهی و نگرش افراد (مثال ۲)، اثر این مداخله روی پروفایل‌های لیپیدی، به عنوان متغیرهای ثانوی در مطالعه نیز بررسی گردید. جدول ۴، مقادیر اندازه‌گیری‌های قبل و بعد از مداخله را برای متغیر HDL-C در تعداد ۴۳ و ۳۸ نفر از بیماران به ترتیب در دو گروه مداخله و کنترل نشان می‌دهد.

جدول ۴- اندازه‌گیری‌های قبل و بعد از مداخله را برای متغیر HDL-C در دو گروه مداخله ($n=43$) و کنترل ($n=38$)

گروه	میانگین	انحراف معیار	خطای معیار	
HDL-C قبل از مداخله	مداخله	۴۳/۶۷	۱۰/۷۶	۱/۶۴
	کنترل	۴۹/۰۸	۱۲/۴۵	۲/۰۲
HDL-C بعد از مداخله	مداخله	۴۳/۷۲	۹/۲۰	۱/۴۰
	کنترل	۴۶/۲۶	۱۰/۱۷	۱/۶۵

صورت یک اختلاف استاندارد شده (شاخص آزمون) بر اساس Z ساخته می‌شود ولی به دلیل این که در اغلب موقعیت‌های کاربردی، انحراف معیارها (واریانس‌ها) جامعه معلوم نیستند، بنابراین از برآورد انحراف معیار نمونه استفاده می‌شود، که در این صورت شاخص استاندارد شده t خواهد بود [۱].

آزمون t در دو حالت واریانس‌های برابر و نابرابر انجام می‌شود. در ادامه فرض می‌شود که انحراف معیارها

ج- مقایسه میانگین متغیر کمی نرمال بین دو گروه مستقل (آزمون t مستقل):

آزمون‌های مستقل زمانی استفاده می‌شوند که هدف مقایسه‌ی دو گروه مستقل از هم باشند نظیر:

- آزمون‌های بررسی اختلاف بین گروه‌های مورد و شاهد (غیر همسان) در مطالعات مورد-شاهدی، همگروهی و یا کارآزمایی بالینی
- بررسی اختلاف بین گروه‌هایی که توسط متغیرهای دموگرافیک یا زمینه‌ای دوحالتی (نظیر جنسیت) تعریف می‌شود.
- بررسی اختلاف بین دو گروه در تحلیل‌های زیر گروهی (Subgroup Analyses)

زمانی که متغیر مورد بررسی کمی نرمال باشد، برای مقایسه متغیر مورد بررسی در دو گروه مستقل آزمون t مستقل استفاده می‌شود. لازم به ذکر است که این متغیر باید پیوسته باشد ولی در مورد متغیرهای کمی گسسته هم با در نظر گرفتن یک سری ملاحظات (مثلاً حجم نمونه بزرگ) قابل به کارگیری است. اگر هدف بررسی مساوی بودن میانگین‌های بین دو گروه باشد یعنی ادعای مطرح شده یا فرضیه پژوهشی به این صورت باشد که:

فرض کنید هدف مقایسه‌ی HDL-C قبل از مداخله در دو گروه مداخله و کنترل باشد. همان‌طور که اشاره شد، برای این منظور نیاز است توزیع اختلاف یا تفاضل میانگین در دو گروه مشخص باشد. زمانی که دو میانگین نرمال باشند، توزیع اختلاف آنها هم نرمال خواهد بود. تحت فرض صفر، اختلاف دو میانگین برابر صفر در نظر گرفته می‌شود، بنابراین میانگین این توزیع نرمال برابر صفر است. اگر انحراف معیارها (واریانس‌ها) جامعه معلوم باشند در این

واریانس مشترک، در واقع یک میانگین وزنی از واریانس‌ها در دو گروه می‌باشد؛ برای محاسبه مشترک ابتدا نیاز است واریانس در هر گروه در $n-1$ یا درجه آزادی [۱] آن گروه ضرب شود که به آن مجموع توان دوم‌ها گفته می‌شود (Sum of Squared یا SS)، در گام بعدی مجموع درجه آزادی‌ها در دو گروه و همچنین SS ها در دو گروه به دست می‌آید. نتایج محاسبات در جدول ۵ ارائه شده است:

جدول ۵- محاسبه واریانس مشترک برای متغیر HDL-C در گروه‌های مداخله و کنترل

گروه	تعداد نمونه	انحراف معیار	واریانس (توان دوم انحراف معیار)	$n-1$	SS (واریانس ضرب در $n-1$)
مداخله	۴۳	۱۰/۷۶	۱۱۵/۷۸	۴۲	۴۸۶۲/۶۶
کنترل	۳۸	۱۲/۴۵	۱۵۵/۰۰	۳۷	۵۷۳۵/۰۹
مداخله HDL-C قبل از مداخله		مجموع		۷۹	۱۰۵۹۷/۷۵

(واریانس‌ها)ی دو جامعه یکسان هستند و محاسبات بر اساس آن انجام خواهد شد. در این حالت، به دلیل برابر واریانس‌های دو جامعه، واریانس‌های نمونه‌ای محاسبه شده روی هم ریخته شده و واریانس مشترک و سپس انحراف معیار مشترکی برای هر دو گروه محاسبه می‌شود. این انحراف معیار در واقع همان انحراف معیار توزیع نمونه‌گیری اختلاف دو میانگین می‌باشد.

مراجعه به جدول توزیع t (جدول ۱۵)، ملاحظه می‌شود که مقدار عددی $۲/۱۰$ نقطه‌ای را نشان می‌دهد که در سمت راست آن سطح زیر منحنی تقریباً برابر $۰/۰۱۹۵$ است، به عبارت دیگر بر اساس درجه آزادی ۷۹ تایی، احتمال اینکه اختلاف میانگین‌های دو نمونه بزرگتر از صفر باشد تقریباً برابر $۰/۰۱۹۵$ است. به همین ترتیب سطح زیر منحنی در سمت چپ $۲/۱۰ -$ نیز تقریباً برابر $۰/۰۱۹۵$ است و در نتیجه، احتمال اینکه اختلاف میانگین‌های دو نمونه کوچکتر از صفر و یا بزرگتر از آن باشد (آزمون دو دنباله‌ای) تقریباً برابر $۰/۰۳۹$ است. در نتیجه مقدار P -Value برای این آزمون تقریباً برابر $۰/۰۳۹$ به دست می‌آید و در نتیجه می‌توان این گونه تصمیم گرفت که برای میانگین HDL-C قبل از انجام مداخله، بین دو گروه اختلاف معنی‌داری وجود داشت ($P < ۰/۰۵$).

نکته:

- ۱- اگر پراکندگی یا واریانس جامعه مشخص باشد در این صورت می‌توان به جای t مقدار استاندارد Z را به دست آورد و بر اساس توزیع نرمال، احتمالات مورد نظر را حساب کرد.
- ۲- اگر فرضیه مطرح شده به این ترتیب بود که «میانگین HDL-C قبل از انجام مداخله در گروه

مجموع درجه آزادی‌ها در دو گروه، درجه آزادی کل برای آزمون t را (در حالت واریانس‌های برابر) نشان می‌دهد. در نهایت برای محاسبه واریانس مشترک کفایت مجموع SS ها به مجموع درجات آزادی تقسیم شود:

$$۱۳۴/۱۵ = ۱۰۵۹۷/۷۵ \div ۷۹ = \text{مجموع درجات آزادی} \div$$

$$\text{مجموع SS ها} = \text{واریانس مشترک}$$

در نتیجه خطای معیار اختلاف میانگین‌ها با استفاده از فرمول [۱]

$$\text{جذر} ((\text{واریانس مشترک} \div n1) + (\text{واریانس مشترک} \div n2))$$

$$\text{جذر} ((۱۳۴/۱۵ \div ۳۸) + (۱۳۴/۱۵ \div ۴۳))$$

برابر با $۲/۵۸$ خواهد بود. در مرحله بعدی باید شاخص t و مقدار آن مشخص گردد که با فرمول زیر به دست می‌آید:

$$(\text{خطای معیار اختلاف میانگین‌ها}) \div (\text{اختلاف میانگین‌های}$$

$$\text{نمونه‌ای}) = t$$

$$= (۲/۵۸) \div (۵/۴۱)$$

$$= ۲/۱۰$$

حال بر اساس مقادیر به دست آمده می‌توان به این سوال پاسخ داد: احتمال این که اختلاف استاندارد شده میانگین‌ها در دو جامعه بزرگتر از $۲/۱۰$ باشد، چقدر است؟ همچنین این آزمون با توجه به نحوه بیان ادعا دو دنباله‌ای است. با

((خطای معیار اختلاف) $\times 2 \pm$ اختلاف میانگین) =

فاصله اطمینان اختلاف میانگین‌ها

به صورت

$$0/41 \pm 2 \times (2/58)$$

$$(0/25 \text{ تا } 10/57)$$

چون این فاصله اطمینان عدد تحت فرض صفر برای اختلاف یعنی عدد صفر را در بر ندارد، بنابراین می‌توان این گونه تصمیم گرفت که برای برای میانگین HDL-C قبل از انجام مداخله، بین دو گروه اختلاف معنی‌داری وجود داشت ($P < 0/05$).

نحوه ارائه یافته‌ها برای این مطالعه:

یافته‌های این بررسی را می‌توان به صورت جدول ۶ ارائه کرد:

کنترل نسبت به گروه مداخله بزرگتر است»، در این صورت کافی است که تنها سطح زیر منحنی سمت چپ $2/10$ (تقریباً برابر $0/019$) و در نتیجه، احتمال اینکه اختلاف میانگین‌های دو نمونه از صفر بزرگتر باشد (آزمون یک دنباله‌ای) تقریباً برابر $0/019$ است. در نتیجه مقدار P -Value برای این آزمون تقریباً برابر $0/019$ به دست می‌آید و در نتیجه می‌توان این گونه تصمیم گرفت که میانگین HDL-C قبل از انجام مداخله، در گروه کنترل نسبت به گروه مداخله بزرگتر است ($P < 0/05$).

نکته:

برای مثال ۳، فاصله اطمینان ۹۵ درصدی برای میانگین جامعه بر اساس فرمول:

جدول ۶- نتایج مقایسه HDL-C قبل از انجام مداخله، بین دو گروه مداخله ($n=43$) و کنترل ($n=38$)

متغیر	گروه مداخله	گروه کنترل	اختلاف میانگین‌ها (فاصله اطمینان ۹۵٪)	درصد تغییرات گروه مداخله نسبت به کنترل #	P-Value
HDL-C قبل از انجام مداخله (mg/dl)	۴۳/۶۷ (۱۰/۷۶)	۴۹/۰۸ (۱۲/۴۵)	* ۵/۴۱ تا ۱۰/۵۷ (۰/۲۵)	۱۱/۰۲٪	۰/۰۳۹

آزمون t مستقل

*: $P < 0/05$

میانگین (انحراف معیار) گزارش شده است.

درصد تغییرات نسبت به گروه کنترل با فرمول (اختلاف میانگین‌ها ÷ میانگین گروه کنترل $\times 100$) به دست می‌آید و تغییرات (اندازه اثر) نسبی (نسبت به گروه کنترل) نامیده می‌شود. در مقابل اختلاف میانگین‌ها، تغییرات (اندازه اثر) مطلق نامیده می‌شود.

نکته:

فرمول t محاسبه شده فوق، زمانی مناسب است که واریانس دو جامعه برابر باشد و این از آنجا ناشی می‌شود که واریانس ادغام شده یا مشترک برای دو جامعه بر اساس این برابری محاسبه می‌شود. زمانی که واریانس دو جامعه یکسان نباشد دیگر نمی‌توان از فرمول فوق برای محاسبه t استفاده کرد و نیاز است فرمول t مبتنی بر واریانس‌های نابرابر به کار برده شود و یا این که آزمون t مبتنی بر واریانس‌های برابر زمانی به کار رود که ابتدا روی داده‌ها تبدیلی (نظیر لگاریتمی) اعمال شده و از این موضوع اطمینان حاصل گردد. این موضوع در اغلب نرم‌افزارهای

آماري رعایت شده و نتایج این آزمون هم بر اساس برابری و هم بر اساس نابرابری واریانس‌ها ارائه می‌شود و برای آن که خواننده تصمیم بگیرد که از نتایج کدام آزمون باید استفاده کند آزمونی نیز برای مقایسه واریانس‌ها ارائه می‌دهد و زمانی که P-value آزمون مقاداری بزرگتر از $0/05$ اختیار کند در این صورت از نتایج برابری واریانس‌ها استفاده خواهد شد. اما آزمون‌هایی که برای برابری واریانس‌ها استفاده می‌شود شامل آزمون Bartlett و Levene هستند که از بین این دو آزمون، آزمون Levene به دلیل این که آزموی نیرومند می‌باشد، توصیه می‌گردد. چون برای آزمون F نیاز است که توزیع داده‌ها نرمال باشد.

کواریانس که ترکیبی از تحلیل واریانس و رگرسیون است، برای مقایسه متغیرهای کمی نرمال در دو یا چند گروه، در صورت وجود اختلافات در اندازه‌گیری‌های پایه (قبل از مداخله) و یا وجود رابطه بین آنها و اندازه‌گیری‌های بعد مداخله، می‌تواند آنها را در تحلیل وارد نماید و اثر آنها را تعدیل (Adjust) نماید. علاوه بر آن، در صورت وجود متغیرهای مخدوشگر، این تحلیل قادر است اثر متغیرهای مخدوشگر را نیز کنترل کند (تحلیل کواریانس یکی از راه‌های کنترل اثر متغیرهای مخدوشگر از طریق تعدیل اثر آنها حین تحلیل داده‌هاست)، ولی به عنوان یک راه حل میانی و ساده، می‌توان تغییرات قبل و بعد از انجام مداخله را در گروه‌های مطالعه به دست آورد و میانگین این تغییرات را با آزمون t مستقل در دو گروه با هم مقایسه کرد. به عبارت دیگر اندازه‌ای از تغییر که در گروه مداخله ایجاد شده است با اندازه‌ای از تغییر که در گروه شاهد ایجاد شده مقایسه می‌شود و زمانی مداخله موثر بوده است که این دو مقدار با هم تفاوت معنی‌داری داشته باشند. فرضیه مطرح شده به این صورت می‌باشد که:

بین میانگین تغییرات گروه‌ها اختلاف وجود دارد،

در این صورت فرض صفر آزمون این است که اختلاف بین میانگین‌ها تغییرات در دو جامعه برابر صفر است و فرض مقابل در این آزمون بیان می‌کند که این اختلاف برابر صفر نیست (کوچکتر و یا بزرگتر از صفر است). سایر ملاحظات مطرح شده برای آزمون t مستقل در مثال ۳، در این مورد هم صادق است.

مثال ۴:

در ادامه مثال ۳، برای محاسبه مقادیر تغییرات اندازه‌گیری‌های قبل و بعد از مداخله برای متغیر HDL-C در تعداد ۴۳ و ۳۸ نفر از بیماران به ترتیب در دو گروه مداخله و کنترل، ابتدا نیاز است اختلاف بین قبل و بعد مداخله محاسبه شود و در گام بعدی، میانگین، انحراف معیار و خطای معیار اختلاف‌ها برای هر یک از گروه‌ها محاسبه می‌شود (جدول ۷):

حال برای داده‌های مثال فوق فرض کنید هدف بررسی برابری (همگنی) واریانس‌ها با استفاده از آزمون F باشد در این صورت ابتدا شاخص F به صورت زیر محاسبه می‌شود:

$$F = 155/00 \div 115/78 = 1/34$$

همان‌طور که ملاحظه می‌شود شاخص F بر اساس تقسیم دو واریانس به دست آمده است که یکی در صورت کسر قرار دارد و دیگری در مخرج و در نتیجه این شاخص دو درجه آزادی دارد که یکی در صورت کسر قرار می‌گیرد و دیگری در مخرج. درجات آزادی F به این ترتیب محاسبه می‌شوند که تعداد نمونه برای محاسبه واریانس صورت و مخرج هر کدام منهای یک شده و به ترتیب درجات آزادی صورت و مخرج را تشکیل می‌دهند. بنابراین برای داده‌های فوق درجات آزادی صورت و مخرج به ترتیب برابر ۴۲ و ۳۷ می‌باشند. با مراجعه به جدول F (جدول ۱۶) با درجات آزادی ۴۲ و ۳۷ با توجه به این که مقدار محاسبه شده از ۱/۳۴ از مقدار جدولی ۱/۷۱ کوچکتر است، در نتیجه برابری واریانس‌ها در دو جامعه تایید شد ($P > 0/05$).

آزمون t مستقل برای بررسی اثر مداخله

اگر هدف بررسی تاثیر مداخله روی یک متغیر کمی باشد، در این صورت از نظر آماری هدف مقایسه میانگین بین گروه‌های مداخله و شاهد بعد از انجام مداخله می‌باشد. اگر این موضوع پذیرفته شود که هیچ متغیر مخدوشگری بر این تاثیرگذاری دخالت نمی‌کند، باز هم نیاز است که مقادیر میانگین گروه‌ها در اندازه‌گیری پایه یعنی قبل مداخله مساوی باشند تا اختلافی که بعد از مداخله مشاهده می‌شود ناشی از اثر مداخله باشد. حال اگر مقادیر پایه یکسان نباشند، در این صورت بخشی از اختلاف مشاهده شده مربوط به اختلاف اولیه گروه‌هاست که نیاز است به گونه‌ای در محاسبات لحاظ شود تا بتوان به اثر خالص مداخله دست یافت.

اگر چه راه حل اساسی برای این منظور استفاده از تحلیل‌هایی نظیر تحلیل کواریانس است. در تحلیل

جدول ۷- تغییرات بین قبل و بعد از مداخله برای متغیر HDL-C در دو گروه مداخله (n=۴۳) و کنترل (n=۳۸)

گروه	میانگین تغییرات	انحراف معیار تغییرات	خطای معیار تغییرات
مداخله	۰/۰۵	۶/۹۹	۱/۰۷
کنترل	-۲/۸۲	۷/۱۷	۱/۱۶

همان‌طور که ملاحظه می‌شود، در گروه مداخله ۰/۰۵ واحد افزایش در HDL-C و در گروه کنترل ۲/۸۲ واحد کاهش در این تغییر وجود داشته است. هدف مقایسه این دو تغییر

محاسبه واریانس مشترک:

جدول ۸- محاسبه واریانس مشترک برای تغییرات (بین قبل و بعد) HDL-C در گروه‌های مداخله و کنترل

گروه	تعداد نمونه	انحراف معیار	واریانس (توان دوم انحراف معیار)	SS (واریانس ضرب در n-1)
مداخله	۴۳	۶/۹۹	۴۸/۸۵	۲۰۵۱/۹۱
کنترل	۳۸	۷/۱۷	۵۱/۳۴	۱۸۹۹/۷۱
مداخله		مجموع		۳۹۵۱/۶۲

$$50/02 = 3951/62 + 79 = \text{مجموع درجات آزادی} \div \text{مجموع SS ها} = \text{واریانس مشترک}$$

در نتیجه خطای معیار اختلاف میانگین‌ها با استفاده از فرمول [۱]

$$\text{جزر} ((\text{واریانس مشترک} + n1) + (\text{واریانس مشترک} + n2))$$

$$\text{جزر} ((50/02 + 38) + (50/02 + 43))$$

است و در نتیجه، احتمال اینکه اختلاف میانگین‌های دو نمونه کوچکتر از صفر و یا بزرگتر از آن باشد (آزمون دو دنباله‌ای) تقریباً برابر ۰/۰۷۳ است. در نتیجه مقدار P-Value برای این آزمون تقریباً برابر ۰/۰۷۳ به دست می‌آید و در نتیجه می‌توان این گونه تصمیم گرفت که برای میانگین HDL-C قبل از انجام مداخله، بین دو گروه اختلاف معنی‌داری وجود داشت ($P < 0/05$).

نکته:

برای مثال ۳، فاصله اطمینان ۹۵ درصدی برای میانگین جامعه بر اساس فرمول

$$= (\text{خطای معیار اختلاف}) \times 2 \pm (\text{اختلاف میانگین})$$

فاصله اطمینان اختلاف میانگین‌ها

به صورت

$$2/86 \pm 2 \times (1/57)$$

$$(6/00 \text{ تا } 0/28)$$

چون این فاصله اطمینان عدد تحت فرض صفر برای اختلاف یعنی عدد صفر را در بر ندارد، بنابراین می‌توان

برابر با ۱/۵۷ خواهد بود. در مرحله بعدی باید شاخص t و مقدار آن مشخص گردد که با فرمول زیر به دست می‌آید:

(خطای معیار اختلاف میانگین‌ها) \div (اختلاف میانگین‌های

$$\text{نمونه‌ای}) = t$$

$$= (1/57) \div (2/86)$$

$$= 1/82$$

حال بر اساس مقادیر به دست آمده می‌توان به این سوال پاسخ داد: احتمال این که اختلاف استاندارد شده‌ی میانگین‌ها در دو جامعه بزرگتر از ۱/۸۲ باشد، چقدر است؟ همچنین این آزمون با توجه به نحوه بیان ادعا دو دنباله‌ای است. با مراجعه به جدول توزیع t (جدول ۱۵)، ملاحظه می‌شود که مقدار عددی ۱/۸۲ نقطه‌ای را نشان می‌دهد که در سمت راست آن سطح زیر منحنی تقریباً برابر ۰/۰۳۶۵ است، به عبارت دیگر بر اساس درجه آزادی ۷۹ تایی، احتمال اینکه اختلاف میانگین‌های دو نمونه بزرگتر از صفر باشد تقریباً برابر ۰/۰۳۶۵ است. به همین ترتیب سطح زیر منحنی در سمت چپ ۱/۸۲ - نیز تقریباً برابر ۰/۰۳۶۵

این گونه تصمیم گرفت که برای میانگین HDL-C قبل از انجام مداخله، بین دو گروه اختلاف معنی‌داری وجود داشت ($P < 0/05$).

نحوه ارائه یافته‌ها برای این مطالعه:
یافته‌های این بررسی را می‌توان به صورت جدول ۹ ارائه کرد:

جدول ۹: نتایج مقایسه تغییرات (بین قبل و بعد از انجام مداخله) HDL-C، بین دو گروه مداخله ($n=43$) و کنترل ($n=38$)

متغیر	گروه مداخله	گروه کنترل	اختلاف میانگین‌ها (فاصله اطمینان ۹۵٪)	درصد تغییرات بعد نسبت به قبل #	P-Value
HDL-C قبل از انجام مداخله (mg/dl)	۰/۰۵ (۶/۹۹)	۲/۸۲ (۷/۱۷)	* ۲/۸۶ (۰/۲۸ تا ۶/۰۰)	٪ ۱۰۱/۶۵	۰/۰۳۹

آزمون t مستقل

*: $P < 0/05$

میانگین (انحراف معیار) گزارش شده است.

درصد تغییرات نسبت به گروه کنترل با فرمول (اختلاف میانگین‌ها + میانگین گروه کنترل $\times 100$) به دست می‌آید و تغییرات (اندازه اثر) نسبی (نسبت به گروه کنترل) نامیده می‌شود. در مقابل اختلاف میانگین‌ها، تغییرات (اندازه اثر) مطلق نامیده می‌شود.

اول و سوم و در نهایت مقایسه دوم و سوم و با جمع‌بندی این نتایج، مسئله مقایسه سه گروه حل شود.

به دلیل اینکه در هر یک از سه مقایسه، میزان ۰/۰۵ خطا وجود دارد، این موضوع سبب مسئله چند آزمونی (multiple testing) می‌شود که میزان خطا خانواده مقایسه‌ها (Family Error Rate) یا خطای نهایی در سه مقایسه بیشتر از ۰/۰۵ می‌گردد ($1 - (0/95)^3 = 0/14$) اساس [۱]. مسئله دومی که در مقایسه دو به دو گروه‌ها وجود دارد این است که همزمانی مقایسه‌ها از بین می‌رود؛ مثلاً در مقایسه بین گروه‌های اول و دوم، اثری از گروه سوم دیده نمی‌شود ولی در واقع هدف مقایسه همزمان سه گروه بوده است. بنابر آنچه گفته شد، استفاده از چند آزمون t برای حل مسئله مقایسه بیش از دو گروه توصیه نمی‌شود و باید از تحلیل واریانس استفاده شود.

شاخص آزمون برای تصمیم‌گیری در آزمون t، یک اختلاف استاندارد شده است که در آن، اختلاف میانگین‌ها در صورت کسر و خطای معیار این اختلاف در مخرج کسر قرار می‌گرفت. اختلاف بین میانگین‌ها در واقع اختلاف بین گروهی و انحراف معیار این اختلاف نیز در واقع پراکندگی درونی مشاهدات در داخل هر گروه است که برای دو گروه به صورت مشترک محاسبه می‌شود و زمانی اختلاف‌ها معنی‌دار می‌شود که t از یک مقدار معین بزرگتر

د- مقایسه میانگین متغیر کمی نرمال میان بیش از دو گروه مستقل (تحلیل واریانس):

این تحلیل برای مقایسه میانگین‌ها برای بیش از دو گروه به کار می‌رود و تعمیمی از آزمون t مستقل برای بیش از دو گروه است. در انواع مطالعات مداخله‌ای و مشاهده‌ای برای مقایسه گروه‌های مطالعه به صورت زیر می‌توان از این تحلیل استفاده کرد:

- بررسی اختلاف بین گروه‌های مورد و بیش از

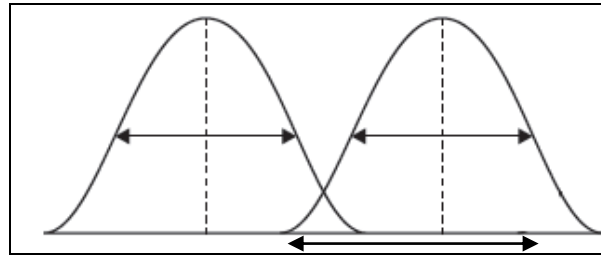
میانگین متغیر کمی نرمال میان بیش از دو گروه مستقل یک شاهد (غیر همسان) در مطالعات مورد-شاهدی، همگروهی و یا کارآزمایی بالینی با بیش از دو گروه

- بررسی اختلاف بین گروه‌هایی که توسط متغیرهای دموگرافیک یا زمینه‌ای بیش از دو حالتی (نظیر سطوح مختلف تحصیلات) تعریف می‌شود.

- بررسی اختلاف بین بیش از دو گروهی در تحلیل‌های زیر گروهی (Subgroup Analyses)

زمانی که هدف مقایسه میانگین متغیر کمی نرمال میان بیش از دو گروه مستقل باشد از تحلیل واریانس استفاده می‌شود. در ابتدا ممکن است این سوال پیش آید که آیا با استفاده از آزمون t نمی‌توان مثلاً سه گروه را به صورت دو به دو با هم مقایسه کرد: مقایسه گروه اول و دوم، مقایسه

شود یعنی اختلاف موجود در صورت کسر بزرگتر از پراکندگی موجود در مخرج کسر شود. در واقع اگر اختلاف بین میانگین‌ها، بزرگتر از پراکندگی درون گروه‌ها



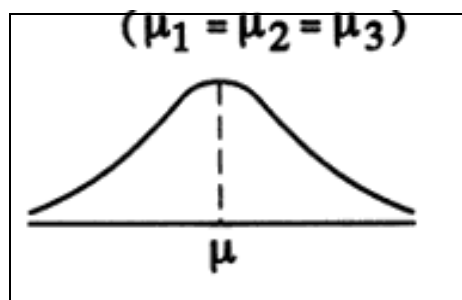
شکل ۱- اختلاف بین گروهی در مقایسه با اختلاف‌های درون گروهی

شاخص t به توان ۲ برسد تبدیل به شاخص F در تحلیل واریانس می‌شود.

در حالت کلی وضعیت‌های مختلف برای مقایسه‌ی ۳ میانگین را نیز می‌توان در شکل‌های ۲ تا ۵ مشاهده نمود:

(۱) وضعیتی که هر سه گروه از یک جامعه استخراج شوند که در این صورت میانگین گروه‌ها (و با فرض برابری واریانس گروه‌ها) یکسان خواهد بود (شکل ۲). در این شکل اختلاف بین گروهی برابر صفر است و پراکندگی درون گروهی وجود دارد، به عبارت دیگر نسبت واریانس بین گروهی به واریانس درون گروهی برابر صفر است و در نتیجه بین گروه‌ها اختلاف وجود ندارد.

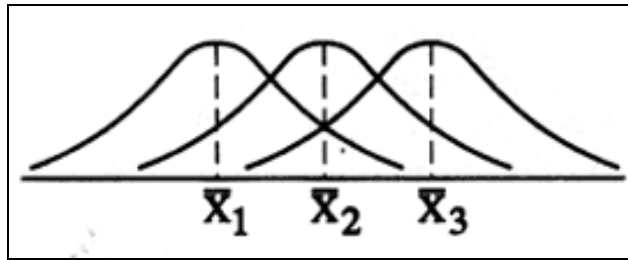
ایده فوق ایده‌ای است که بر اساس آن شاخص‌های مربوط به مقایسه میانگین‌ها در بیش از دو گروه بنا می‌شود. در واقع اصطلاح تحلیل واریانس (تجزیه واریانس)، که برای مقایسه میانگین در بیش از دو گروه به کار می‌رود، این ایده را برای ساختن شاخص آزمون به کار می‌برد؛ به طور خلاصه پراکندگی کل داده‌ها به پراکندگی بین گروه‌ها و پراکندگی درون گروه‌ها تجزیه (افراز) می‌شود و زمانی که واریانس بین گروهی بزرگتر از واریانس درون گروهی گردد، میان میانگین‌ها اختلاف وجود دارد. تحلیل واریانس در واقع تعمیمی از آزمون t مستقل برای بیش از دو گروه است. با این تفاوت که در تحلیل واریانس به جای انحراف معیار از خود واریانس استفاده می‌شود و اگر انحراف معیار به توان ۲ برسد برابر واریانس می‌شود و به همین دلیل اگر



شکل ۲: میانگین‌های یکسان

نسبت به پراکندگی درون گروهی کوچک است، به عبارت دیگر نسبت واریانس بین گروهی به واریانس درون گروهی کوچک است و در نتیجه بین گروه‌ها اختلاف وجود ندارد.

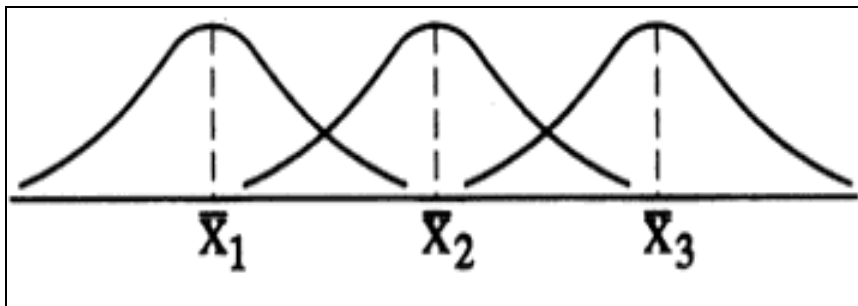
(۲) وضعیتی که اختلاف کوچکی بین سه گروه وجود داشته باشد و در این صورت احتمالاً بین میانگین گروه‌ها اختلاف معنی‌داری وجود ندارد (شکل ۳). در این شکل اختلاف بین گروهی



شکل ۳- اختلاف ناچیز بین سه گروه

درون گروهی بزرگ است، به عبارت دیگر، نسبت واریانس بین گروهی به واریانس درون گروهی قابل ملاحظه است و در نتیجه بین گروه‌ها اختلاف وجود دارد.

(۳) وضعیتی که اختلاف قابل ملاحظه‌ای بین سه گروه وجود داشته باشد و در این صورت بین میانگین حداقل یکی از گروه‌ها با سایر گروه‌ها اختلاف معنی‌داری وجود دارد (شکل ۴). در این شکل اختلاف بین گروهی نسبت به پراکندگی



شکل ۴- اختلاف قابل ملاحظه بین سه گروه

قرار گیرند. به زبان ساده، سه مورد اخیر مسیر آزمون‌ها را مشخص می‌کنند یعنی اگر این شرایط برقرار باشند می‌توان از تحلیل واریانس و در غیر این صورت نمی‌توان از آن استفاده نمود. بنابراین با برقراری پیش‌فرض‌ها و تحت فرض صفر میانگین گروه‌ها مورد بررسی با هم برابر هستند و در صورتی که این فرض رد شود می‌توان بیان نمود که حداقل یکی از گروه‌ها با سایرین اختلاف دارد. شایان ذکر است این موضوع که کدام گروه‌ها با هم اختلاف دارند توسط تحلیل واریانس مشخص نمی‌شود، یعنی تحلیل واریانس مشخص نمی‌کند که کدامیک از زوج گروه‌ها با هم اختلاف دارند و برای این منظور نیاز است آزمون‌های (مقایسه‌ای) تعقیبی (Post hoc) را انجام داد.

آزمون‌های تعقیبی

پس از آشکار شدن معنی‌داری اختلاف گروه‌ها در تحلیل واریانس، در آزمون‌های تعقیبی، میانگین گروه‌ها به صورت دو به دو با هم مقایسه شده تا دقیقاً معلوم شود که

همان‌طور که در شکل‌های ۲ تا ۵ ملاحظه می‌شود، اگر اختلاف بین میانگین گروه‌ها یعنی پراکندگی بین گروه‌ها باید بزرگتر از پراکندگی درونی آنها باشد تا این اختلاف از لحاظ آماری معنی‌دار شود، به عبارت دیگر، در این صورت منحنی‌های مربوط به گروه‌ها از یکدیگر متمایز می‌شود به گونه‌ای که منحنی یک گروه با منحنی گروه دیگر تداخل یا هم‌پوشانی جدی نخواهد داشت.

مشابه با آزمون t برای دو نمونه، فرض صفر تحلیل واریانس نیز بیان می‌کند که داده‌ها از جوامع یکسانی هستند و این موضوع ایجاد می‌کند که برای داده‌های با توزیع نرمال، میانگین و واریانس داده‌ها یکسان باشند. فرضیه‌های پژوهشی که در این زمینه مطرح می‌شوند بیان می‌دارند که بین میانگین گروه‌ها اختلاف وجود دارد. برای بررسی این موضوع، نرمال بودن توزیع داده‌ها، مستقل بودن گروه‌ها از یکدیگر و برابری واریانس (اگرچه به طور مستقیم در فرضیه‌ها و اهداف پژوهشی مطرح نمی‌شوند ولی باید) به عنوان پیش‌فرض‌های این آزمون مورد ارزیابی

۴- در صورتی که در میانگین گروه‌ها یک روند صعودی و نزولی وجود داشته باشد یعنی این که میانگین گروه‌ها به ترتیب از گروه‌های اول تا آخر بزرگ و یا کوچک شوند، در این صورت استفاده از آزمون روند توصیه می‌شود که در صورت مقایسه دو به دو با روش t معمولی نمی‌توان این روند را کشف کرد. در تحلیل واریانس متغیر کیفی (معمولاً اسمی) و متغیر وابسته یک متغیر کمی (پیوسته) است. اما اگر زمانی متغیر مستقل یک متغیر رتبه‌ای باشد مسئله‌ی روند ممکن است وجود داشته باشد که در این صورت می‌گویند اصطلاحاً رابطه‌ی دوز - پاسخ (Dose-Response) وجود دارد. برای بررسی این موضوع با متغیر مستقل مورد بررسی به صورت کمی برخورد شده و وجود روند خطی (و یا شکل‌های دیگری از آن مثل درجه دوم و ...) نیز بررسی می‌شود که این موضوع در ادبیات تحلیل واریانس تحت عنوان مقابله‌ها (Contrasts) بررسی می‌شود. برای محاسبات این بخش می‌توان از نرم‌افزارهای آماری استفاده کرد. به هر حال زمانی که مقابله‌های خطی معنی‌دار شوند (P-value ان کمتر از ۰/۰۵ شود)، بدین معنی است که میانگین‌های گروه‌ها دارای یک روند خطی (صعودی یا نزولی) هستند.

۵- ملاحظه شد که در آزمون t با انجام ۳ مقایسه، نرخ خطا از ۰/۰۵ به ۰/۱۴ افزایش یافت. اغلب آزمون‌های تعقیبی به گونه‌ای طراحی شده‌اند که با انجام تعدادی مقایسه نرخ خطا را در سطح ۰/۰۵ ثابت نگه دارند.

البته در این مورد می‌توان نرخ خطا را طوری کمتر در نظر گرفت که پس از انجام مقایسه‌ها نرخ نهایی خطا از ۰/۰۵ بیشتر نشود. بونفرونی نشان داد که اگر مثلاً ۳ مقایسه تعقیبی وجود دارد نرخ خطای اولیه ۰/۰۵ را بر ۳ تقسیم کنید و خطا را $0/017 = (0/05 \div 3)$ در نظر بگیرید در این صورت نرخ نهایی خطا از ۰/۰۵ بیشتر نخواهد بود. اما بازم با توجه به نکات ۲ و ۳ اشکالاتی به صرف مقایسه‌های دو به دو توسط آزمون T و

کدام یک از دو گروه‌ها هستند که با هم اختلاف دارند. ممکن است در این مورد نقدهایی به صورت زیر وجود داشته باشد:

۱- ممکن است نیاز به مقایسه همه زوج‌گروه‌ها نباشد، که در این مورد می‌توان به همان تعداد لازم مقایسه‌ها را انجام داد؛ ممکن است برخی از مقایسه‌ها از پیش تعیین شده باشند که به آنها مقایسه‌های پیش‌تجربی گفته می‌شود. در مقابل ممکن است هیچ مقایسه‌ای از پیش تعیین نشده باشد که در این صورت مقایسه‌های پس‌تجربی در پی کشف اختلاف دو به دوی میانگین‌ها می‌باشد.

۲- در این مقایسه‌های دو به دو نیاز است که اثر گروه‌های دیگر غیر از دو گروه مورد مقایسه نیز در نظر گرفته می‌شود که برای این منظور از آزمون t اصلاح شده‌ای استفاده می‌شود که در آن از واریانس درون‌گروهی مربوط به همه‌ی گروه‌ها جذر گرفته و از آن به عنوان انحراف معیار در مخرج کسر استفاده می‌شود و اختلاف بین دو گروه مورد بررسی در صورت کسر قرار می‌گیرد که بر این اساس شاخص t اصلاح شده در دو مورد با شاخص t معمولی متفاوت است:

- انحراف معیار در مخرج کسر
- درجه آزادی آزمون t اصلاح شده که برابر است با درجات آزادی درون‌گروهی

و این دو تفاوت نتیجه اعمال اثر گروه‌های دیگر غیر از دو گروه مورد مقایسه است که به این ترتیب اثر آنها تعدیل می‌شود.

۳- این مقایسه‌ها (یعنی مقایسه‌های تعقیبی) زمانی انجام می‌شود که این نتیجه حاصل شده باشد که بر اساس نتایج تحلیل واریانس میان گروه‌ها اختلافی وجود داشته باشد و این مقایسه‌ها تلاش در جهت یافتن اختلاف حاصل باشد.

بنابراین همان‌طور که اشاره شد، زمانی که نتیجه تحلیل واریانس معنی‌دار نباشد، منطقی نیست و توصیه نمی‌شود که اختلاف میان گروه‌ها از طریق مقایسه‌های تعقیبی بررسی شود.

از منابع استفاده از این آزمون را برای زمانی توصیه می‌کنند که تعداد مقایسه‌ها کم باشد (مثلاً ۳ مقایسه) و برای تعداد بیشتری از مقایسه‌ها استفاده از آزمون‌های تعقیبی دیگر توصیه می‌شود. همچنین زمانی که حجم نمونه در گروه‌ها کوچک باشد که در این صورت توان آزمون به دلیل کوچک بودن حجم نمونه پایین می‌آید که استفاده از این آزمون سبب افزایش توان آشکارسازی اختلاف بین گروه‌ها می‌شود که در این مورد افزایش توان آزمون می‌تواند یک مزیت باشد یعنی آزمون با نرخ خطای بالاتر ولی در عوض با توان بالاتری انجام می‌شود و بدین ترتیب توان کوچک آزمون با بالا رفتن نرخ خطا بالاتر رفته جبران می‌شود.

۲) آزمون تعقیبی بونفرونی (تصحیح بونفرونی روی آزمون LSD)

همان‌طور که اشاره شد آزمون LSD تفاوت‌های کوچک را آشکار می‌کند و این یک مزیت آن بود ولی نرخ خطا به ازای مقایسه‌های دو به دو افزایش می‌یابد که این عیب آن می‌باشد. برای تصحیح این اشکال بونفرونی روشی را پیشنهاد داده است که بر اساس آن نیاز است نرخ خطای اولیه متناسب با تعداد مقایسه‌ها کمتر از $0/05$ در نظر گرفته شود (برای سه مقایسه برابر $0/017 = (3 \div 0/05)$) [۱].

در زمینه‌ی آزمون‌های تعقیبی تعداد زیادی از آزمون‌ها معرفی شده‌اند که برای آشنایی با مفهوم آنها به دو مورد از آزمون‌های متداول در این زمینه اشاره شده است. با توجه به این که پرداختن به همه‌ی آنها در این مقال نمی‌گنجد تنها به ذکر نکاتی چند در باب استفاده‌ی بهینه از این آزمون‌ها بسنده خواهد شد.

۱) آزمون توکی، آزمونی است که به نوعی مصالحه بین بونفرونی و LSD است؛ از لحاظ تعدیل میزان خطا، مانند بونفرونی است و نرخ خطا را در $0/05$ ثابت نگه می‌دارد ولی مانند بونفرونی محافظه‌کار نیست ولی به اندازه LSD نیز آزادانه فرض صفر را رد نمی‌کند و به همین دلیل به تفاوت واقعی معنی‌داری یا (Honestly Significant Difference) HSD

با تعدیل بونفرونی وجود دارد. یعنی اثر گروه‌های دیگر و اختلاف همزمان گروه‌ها در نظر گرفته نشده است.

بنابراین در حالت کلی و به طور خلاصه مسئله مقایسه‌های تعقیبی باید پس از معنی‌داری تحلیل واریانس و مبتنی بر واریانس درون‌گروهی کل مطرح شود. این آزمون‌ها در پی این موضوع هستند که دو جنبه‌ی متناقض در مقایسه‌ها را اعمال نمایند:

الف- کنترل میزان خطا

ب- آشکار نمودن اختلاف موجود میان گروه‌ها و بر اساس این دو خصوصیت آزمون‌های مختلفی وجود دارند که در ادامه به برخی از متداولترین آنها پرداخته خواهد شد که در ادبیات علوم پزشکی بیشتر مورد استفاده قرار می‌گیرند.

۱) آزمون کمترین تفاوت معنی‌داری

(Least Significant Difference or LSD)

این آزمون که در واقع اصلاحی از آزمون t معمولی است برای مقایسه دو به دو میانگین گروه‌ها طراحی شده است. در این آزمون در صورت کسر مشابه آزمون t معمولی اختلاف بین دو میانگین قرار می‌گیرد ولی در مخرج کسر به جای انحراف معیار اختلاف موجود در صورت از جذر واریانس درون‌گروهی استفاده می‌شود که بدین ترتیب اثر گروه‌های دیگر مورد بررسی نیز تعدیل می‌شود. اما اشکالی که در این روش وجود دارد این است که به ازای مقایسه‌های بیش از یک مورد میزان نهایی خطا از $0/05$ بیشتر می‌شود اما توان این آزمون در آشکارسازی اختلاف‌ها بالاست. به عبارت دیگر در اشاره به دو خصوصیت مورد بررسی برای آزمون‌های تعقیبی این آزمون خاصیت اول یعنی کاهش میزان خطا در مقایسه‌های چندگانه را لحاظ نمی‌کند (عیب) ولی آشکارسازی اختلاف بین گروه‌ها را به خوبی انجام می‌دهد (مزیت).

اشکالی که در این مقایسه‌ها وجود دارد این است که به دلیل افزایش میزان خطا از $0/05$ به $0/14$ در سه مقایسه، ممکن است به دلیل افزایش در توان آزمون فرض صفرهای صحیح به اشتباه رد شوند. به همین دلیل برخی

معروف است و بسیاری از لحاظ آماری این آزمون را توصیه می‌کنند.

(۲) برای مقایسه گروه‌های مورد بررسی صرفاً با یک گروه شاهد (و نه مقایسه بین دویه‌دوی گروه‌ها مداخله با یکدیگر)، آزمون دانت پیشنهاد می‌شود که این آزمون نیز نرخ خطا را در ۰/۰۵ ثابت نگه می‌دارد و از لحاظ آشکارسازی اختلاف بین بونفرونی و LSD است.

(۳) برخی از مولفین روش کشف نرخ کاذب (False Discovery Rate :FDR) را پیشنهاد می‌کنند و آن را در مقایسه با سایر روش‌ها کارتر می‌دانند [۵].

در این روش ابتدا تمام مقایسه‌های دویه‌دوی ممکن با استفاده از آزمون t انجام شده و مقادیر P-Value محاسبه می‌شود. سپس P-Value ها به ترتیب از بزرگ به کوچک محاسبه می‌شوند. و به ترتیب به آنها شماره ۱ تا k داده می‌شود که تعداد مقایسه‌های دو به دو است. سپس برای مقایسه‌ی i -ام ($i=1, \dots, k$)، خطای تعدیل شده به صورت $(i \div k) \times 0.05 =$ خطای تعدیل شده به دست می‌آید، که در آن ۰/۰۵، مقدار خطای اولیه است.

پیش‌فرض‌های تحلیل واریانس:

در تحلیل واریانس یک سری پیش‌فرض‌ها در نظر گرفته می‌شود و بر اساس آنها شاخص آزمون محاسبه می‌گردد. بنابراین اعتبار شاخص آزمون در گروه برقراری این پیش‌فرض‌هاست و زمانی که انحراف از این پیش‌فرض‌ها ایجاد شود اعتبار شاخص آزمون F زیر سوال می‌رود (البته این شاخص به انحراف‌های جزئی از نرمال نیرومند (Robust) است ولی نابرابری واریانس گروه‌ها روی این شاخص تاثیرگذار است). به هر حال ممکن است در موقعیت‌های عملی وضعیت‌هایی وجود داشته باشد که انحراف‌های جدی از پیش‌فرض‌ها وجود داشته باشد که در این صورت نیاز است راه‌حل‌های دیگری دنبال شود. در ادامه راجع به هر یک از پیش‌فرض‌ها و راه‌حل‌های جایگزین در صورت برقرار نبودن آنها بحث خواهد شد.

(۱) پیش‌فرض نرمال بودن توزیع داده‌ها:

انحراف‌های جدی از نرمال بودن بر اعتبار شاخص F تاثیر می‌گذارد. لازم به ذکر است در این مورد کج (چوله) شدن توزیع داده‌هاست که تاثیر آن بیشتر است، در صورتی که غیرنرمال بودن به دلیل کشیدگی (نوکتیز یا پست بودن) شکل توزیع داده‌ها اتفاق افتد تاثیر زیادی بر اعتبار F ندارد. در صورتی که این پیش‌فرض برقرار نباشد، پیشنهاد اول انجام تبدیل بر روی این داده‌های کمی است. مثلاً به جای اصل داده‌ها از انجام تبدیل استفاده شود. در صورتی که استفاده از تبدیل، این پیش‌فرض‌ها را برقرار کرد، آزمون F مناسب است و در صورتی که برقرار نشد، استفاده از آزمون ناپارامتری کروسکال-والیس برای مقایسه گروه‌ها پیشنهاد می‌شود. در این آزمون به جای اصل داده‌ها از رتبه‌های متناظر با هر داده استفاده می‌شود. لازم به ذکر است در این مورد به عنوان شاخص توصیفی به جای میانگین (انحراف معیار) در هر گروه، میانه (صدک ۲۵ - صدک ۷۵) [۳] گزارش می‌شود.

(۲) پیش‌فرض همگنی واریانس‌ها:

بر اساس این پیش‌فرض، پراکندگی‌های درون‌گروهی روی هم ریخته می‌شود و واریانس درون‌گروهی را تشکیل داده و زمانی که این پیش‌فرض برقرار نباشد اعتبار مخرج کسر F زیر سوال است و معمولاً بررسی این پیش‌فرض جدی‌تر از پیش‌فرض نرمال بودن توزیع داده‌هاست. در صورتی که این پیش‌فرض برقرار نباشد مشابه حالت قبل پیشنهاد اول استفاده از تبدیلات است. در صورتی که تبدیلات بتواند این پیش‌فرض را برقرار نماید، تحلیل واریانس استفاده می‌گردد و در صورتی که استفاده از تبدیل مسئله را حل نکند در این مورد ولج و براون-فرساید آزمون‌های نیرومندی (Robust test) پیشنهاد کردند که نسبت به انحراف از این پیش‌فرض‌ها حساس نیستند (نیرومند هستند). بنابراین اگر تبدیلات کارساز نبودند برای مقایسه میانگین گروه‌ها از این آزمون‌ها و برای آزمون تعقیبی (در صورت معنی‌داری آزمون‌های فوق) از آزمون جیمز-هاول (Games-Howell) پیشنهاد می‌گردد.

اگر تبدیلات پیش‌فرض همگنی واریانس‌ها را برقرار نمایند، معمولاً برقرار شدن آن منجر به برقرار شدن نرمال

ساعت بعد و ۶ ساعت بعد از مصرف دارو در گروهی از افراد) که در این صورت دیگر پیش‌فرض فوق برقرار نیست. در این مورد باید از تحلیل واریانس دیگری (البته با برقراری پیش‌فرض نرمال بودن توزیع داده‌ها) استفاده گردد که به آن تحلیل واریانس با اندازه‌های تکراری گفته می‌شود. در این تحلیل ساختار همبستگی بین مشاهدات در محاسبه واریانس‌ها لحاظ می‌گردد. در صورت معنی‌داری این آزمون، استفاده از آزمون‌های تعقیبی زوج‌گونه بونفرونی برای مقایسه زوج میانگین‌ها پیشنهاد می‌شود. لازم به ذکر است در صورتی که پیش‌فرض نرمال بودن توزیع داده‌ها نیز برقرار نباشد، در این مورد مشابه قبل از تبدیل استفاده می‌شود و در صورتی که تبدیل کارساز نباشد، آزمون فریدمن (Friedman test) توصیه می‌شود. در این مورد نیز در صورتی که متغیر مورد بررسی رتبه‌ای ارزیابی شود، برقراری پیش‌فرض ۱ مفهومی ندارد و آزمون فریدمن انتخاب اول خواهد بود.

سایر تحلیل‌های مرتبط:

در صورتی که صرفاً مقایسه میانگین متغیر وابسته در گروه‌ها به عنوان متغیر مستقل مد نظر باشد با برقراری سایر پیش‌فرض‌ها، ANOVA توصیه می‌شود. در این صورت (از لحاظ آماری) یک متغیر مستقل و یک متغیر وابسته در تحلیل وارد می‌شوند و فرض بر آن است که مقادیر وجود دارد متغیر وابسته صرفاً تحت تاثیر متغیر مستقل است و متغیر دیگری این رابطه را تحت تاثیر قرار نمی‌دهد.

در مطالعات تجربی شاید این پیش‌فرض زیاد دور از ذهن نباشد (هر چند باز هم قطعی نمی‌توان نظر داد) ولی در مقابل در مطالعات مشاهده‌ای شرایط کنترل شده نیست و متغیرهای دیگری این رابطه را می‌توانند مخدوش کنند. بنابراین نیاز است اثر آنها تعدیل شود. برای تعدیل اثر یک متغیر کمی (با هدف مقایسه میانگین یک متغیر کمی در گروه‌ها) تحلیل کواریانس (Analysis Of Covariance) به کار می‌رود. در صورتی که متغیرهای مورد بررسی، کیفی باشند تحلیل واریانس‌های چندطرفه به کار می‌تواند اثر متغیرها را به صورت جداگانه و یا به صورت اثر متقابل با

بودن توزیع داده‌ها نیز می‌گردد (معمولاً عدم برقراری عدم همگنی واریانس با عدم نرمال بودن توزیع داده‌ها مرتبط است و شاید یک دلیل مشترک نظیر وجود داده‌های پرت سبب هر دو مسئله گردد). بنابراین در آزمون‌های نیرومند که آزمون‌های پارامتری هستند برقراری پیش‌فرض نرمال بودن توزیع داده‌ها نیاز است و زمانی که واریانس‌ها برابر نیستند و پیش‌فرض نرمالیتی هم برقرار نباشد استفاده از این آزمون‌ها نیز توصیه نمی‌شود. بنابراین تبدیلات «یک تیر و دو نشان» هستند. یکی از بهترین تبدیلات برای داده‌های کمی پیوسته تبدیل لگاریتمی است، البته در مورد سایر داده‌ها یعنی داده‌های شمارشی، نسبت‌ها و زمان لازم تا رخداد یک پیشامد به ترتیب تبدیل‌های جذر، آرکسینوس جذر و معکوس پیشنهاد می‌شود. در صورتی که به طور همزمان دو پیش‌فرض برقرار نشدند در این صورت استفاده از آزمون ناپارامتری کروسکال-والیس توصیه می‌شود.

۳) کمی بودن متغیر وابسته:

در صورتی که متغیر مورد بررسی کمی پیوسته باشد، پیش‌فرض اول و بررسی آن مفهوم پیدا می‌کند بنابراین در حالت ایده‌آل، متغیر وابسته باید کمی پیوسته باشد اما در مورد متغیرهای کمی گسسته نیز با اعمال تبدیل می‌توان به نتایج مطلوبی دست یافت.

ممکن است در عمل موقعیت‌هایی نیز وجود داشته باشد که متغیر مورد بررسی به صورت رتبه‌ای (مثلاً رده‌های BMI یا اندازه‌گیری درد با مقیاس آنالوگ بصری) ارزیابی شوند که در این صورت مقایسه گروه‌ها توسط آزمون ناپارامتری کروسکال-والیس انجام می‌شود.

۴) گروه‌های مورد بررسی از هم مستقلند:

بر اساس مستقل بودن گروه‌ها ساختار واریانس‌های درون‌گروهی و بین‌گروهی تعریف و محاسبه شدند. زمانی که این پیش‌فرض برقرار نباشد دیگر اعتبار این مجموعه از مشاهدات زیر سوال می‌رود. در موقعیت‌هایی که اندازه‌گیری‌هایی تکراری روی مجموعه یکسانی از مشاهدات صورت می‌گیرد (نظیر اندازه‌گیری پایه، سه

با یک مثال ساده منطق تشکیل جدول ANOVA و نحوه محاسبه اجزای سازنده آن روشن خواهد شد.

مثال ۱: در مطالعه‌ای با هدف بررسی رابطه بین توزیع گلبول‌های قرمز با سطح و میزان فعالیت در بیماران مبتلا به Ulcerative Colitis، میانگین پارامتر RDW (Red Blood Cell Distribution width) در گروه‌های فعال و در فاز بهبودی این بیماری با گروه کنترل مقایسه شدند (۷). برای روشن شدن محاسبات تحلیل واریانس، داده‌های سه نفر در هر گروه استفاده خواهد شد که به صورت جدول ۱۰ ارائه می‌شوند:

جدول ۱۰- مقادیر مشاهدات RDW در گروه‌های مطالعه

نام گروه‌ها	کد گروه‌ها	مشاهدات
کنترل	۱	۱۲ ۱۱ ۱۳
بهبودی	۲	۱۴ ۱۶ ۱۵
فعال	۳	۱۸ ۱۷ ۱۹

متغیر گروه در تحلیل وارد نمود. در ضمن اگر هدف بررسی اثر دو یا چند عامل به طور همزمان باشد، تحلیل واریانس دو یا چند طرفه (Two Way ANOVA) به کار می‌رود که در آنها می‌توان اثر اصلی یا اثر متقابل متغیرها را بررسی نمود. برای مطالعه بیشتر در این زمینه می‌توان [۶] را ملاحظه نمود.

ضمیمه ۱: نحوه انجام محاسبات تحلیل واریانس در قالب یک مثال ساده

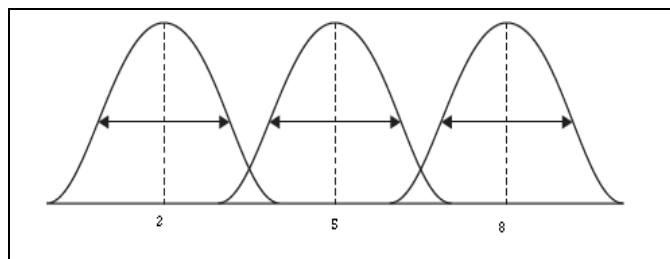
برای انجام محاسبات تحلیل واریانس توصیه می‌شود از نرم‌افزارهای آماری نظیر نرم‌افزار SPSS، STATA یا Minitab استفاده می‌شود. ولی برای آشنایی با مفاهیم کلی،

با مراجعه به جدول، داده‌ها به صورت زیر بازسازی شدند:

کد گروه‌ها	مشاهدات
۱	۲ ۱ ۳
۲	۴ ۶ ۵
۳	۸ ۷ ۹

برابر ۲، ۵ و ۸ و واریانس (انحراف معیار) آنها نیز به ترتیب برابر ۱، ۱ و ۱ است بنابراین شکل توزیعی مشاهدات را در سه گروه (با فرض نرمال بودن مشاهدات) در شکل ۶ می‌توان ملاحظه نمود:

همان‌طور که ملاحظه می‌شود اعداد جدول فوق با کم کردن عدد ۱۰ از مشاهدات اصلی بدست آمدند (کم کردن مشاهدات از یک عدد ثابت تاثیری در محاسبه تحلیل واریانس‌ها ندارد). میانگین گروه‌های ۱، ۲ و ۳ به ترتیب



شکل ۶- مقادیر میانگین در گروه‌های مطالعه

می‌شود، بنابراین نیاز است تغییرات کل، تغییرات بین گروهی و تغییرات درون گروهی محاسبه شود. برای این

همان‌طور که اشاره شد، پراکندگی کل داده‌ها به پراکندگی بین گروه‌ها و پراکندگی درون گروه‌ها تجزیه (افراز)

حال با در نظر گرفتن این پیش‌فرض که واریانس گروه‌ها یکسان است (که این پیش‌فرض باید بررسی شود)، و تغییرات درون گروه‌ها روی هم ریخته و مجموع تغییرات درون‌گروهی را به صورت زیر محاسبه می‌شود:

$$SSW = 2 + 2 + 2$$

همان‌طور که ملاحظه می‌شود جمع دو مولفه تغییرات بین‌گروهی و درون‌گروهی برابر تغییرات کل می‌شود یعنی:

$$SST = SSB + SSW$$

به عبارت دیگر، تغییرات کلی مشاهدات به دو مولفه تغییرات بین‌گروهی و تغییرات درون‌گروهی تجزیه می‌شود (تجزیه واریانس).

محاسبه واریانس‌ها

« اگر واریانس بین‌گروهی بزرگتر از واریانس

درون‌گروهی شود، میان میانگین‌ها اختلاف وجود دارد». بنابراین تنها نیاز است واریانس‌های بین‌گروهی و درون‌گروهی محاسبه گردند، به علاوه نیاز است قبل از محاسبه واریانس‌ها موضوع دیگری را نیز یادآوری شود و آن این‌که برای محاسبه واریانس ساده زمانی که تغییرات (SS) محاسبه می‌شود، این مقدار SS تقسیم بر (n-1) می‌شد که در واقع درجه آزادی واریانس است. بنابراین برای محاسبه واریانس‌های بین‌گروهی و درون‌گروهی نیاز است که تغییرات مربوط به هر یک تقسیم بر درجات آزادی متناظر شوند. از این رو نیاز است درجه آزادی هر یک از آنها به دست آید.

درجه آزادی بین‌گروهی:

در محاسبه تغییرات (SS) بین‌گروهی، میانگین هر یک از گروه‌ها منهای میانگین کل شد بنابراین برای محاسبه این درجه آزادی باید به این سوال پاسخ داد که تعداد آزادی عمل برای انتخاب میانگین‌های هر یک از گروه‌ها، با توجه به این که میانگین کل معلوم است، چقدر باید باشد [۱]؟ ۲ مورد از میانگین‌ها آزادانه انتخاب می‌شوند و مورد سوم آزادانه انتخاب نمی‌شود و بر اساس محدودیت میانگین کل محاسبه می‌شود، بنابراین از سه مورد ۲ مورد با آزادی و ۱ مورد با محدودیت مواجه است یعنی

$$df_B = 3 - 1 = 2$$

منظور، در مرحله اول همه مشاهدات روی هم ریخته می‌شوند و یک تغییرات (SS یعنی Sum of Squared) کلی برای مشاهدات محاسبه می‌شود. علت این‌که از همان ابتدا تغییرات (SS) تقسیم بر n-1 و واریانس محاسبه نمی‌شود این است که قرار است تغییرپذیری مشاهدات به دو دسته بین گروه‌ها و درون گروه‌ها تقسیم شود و برای این منظور باید از SSها استفاده کرد. در نهایت واریانس‌ها نیز محاسبه خواهد شد.

تغییرات کل (SST):

برای این منظور باید نیاز است هر یک از مشاهدات (برای هر سه گروه) منهای میانگین کل (میانگین همه مشاهدات یعنی ۵) شوند و به توان ۲ برسند و همه آنها با هم جمع شوند:

$$SST = (1-5)^2 + (2-5)^2 + \dots + (1-5)^2 = 60$$

تغییرات بین گروهی (SSB):

برای محاسبه آن نیاز است که اختلاف میانگین هر یک از گروه‌ها با میانگین کل بررسی شود. عملاً میانگین هر یک از گروه‌ها منهای میانگین کل می‌شود و به توان ۲ می‌رسد و برای این که اثر تعداد مشاهداتی که میانگین هر یک از گروه‌ها بر اساس آن ساخته می‌شوند در نظر گرفته شود، هر یک از این اختلافات در تعداد مشاهدات گروه‌های نظیر ضرب شده و در نهایت همه این کمیت‌ها با هم جمع می‌شوند:

$$SSB = 3 \times (2-5)^2 + 3 \times (5-5)^2 + 3 \times (8-5)^2 = 3 \times ((2-5)^2 + (5-5)^2 + (8-5)^2) = 54$$

تغییرات درون گروهی (SSW):

برای محاسبه آن نیاز است درون هر گروه تغییرات مشاهدات در اطراف میانگین همان گروه محاسبه شود. نتایج عبارت است از:

$$SSW = (1-2)^2 + (2-2)^2 + (3-2)^2 = 2$$

$$SSW = (4-5)^2 + (5-5)^2 + (6-5)^2 = 2$$

$$SSW = (7-8)^2 + (8-8)^2 + (9-8)^2 = 2$$

و در حالت کلی برای k گروه

$$df_B = k - 1$$

برای محاسبه درجه آزادی درون گروهی باید به این سوال پاسخ داد که در هر یک از گروه‌ها آزادی عمل در انتخاب مشاهدات چیست و مشابه با آن چه در محاسبه واریانس معمولی عمل شد، این بار به ازای هر یک از گروه‌ها داریم:

$$df_1 = 3 - 1 = 2 \quad \text{برای گروه اول}$$

$$df_2 = 3 - 1 = 2 \quad \text{برای گروه دوم}$$

$$df_3 = 3 - 1 = 2 \quad \text{برای گروه سوم}$$

بنابراین با روی هم ریختن این درجات آزادی (با توجه به روی هم ریختن تغییرات درون گروهی) درجه‌ی آزادی درون گروهی به صورت زیر به دست می‌آید:

$$df_w = 2 + 2 + 2 = 3 [2] = 6$$

$$df_w = df_1 + df_2 + df_3 = k(m-1)$$

که در آن k تعداد گروه‌ها و m تعداد مشاهدات در گروه i ام ($i=1, 2, 3$) است.

لازم به ذکر است همان‌طور که SST با جمع SSB و SSW به دست آمده بود، درجات آزادی کل (df_T) نیز با جمع درجات آزادی بین گروهی (df_B) و درون گروهی (df_w) به دست می‌آید:

$$df_T = df_B + df_w = (k-1) + k(m-1) = k + km + k-1 = n-1$$

محاسبه‌ی واریانس‌ها:

$$27 = (2 \div 54) = \text{درجه آزادی بین گروهی یا } (df_B)$$

$$\text{تغییرات بین گروهی یا } (SSB) = \text{واریانس بین گروهی}$$

$$(MSB)$$

$$6 = (6 \div 6) = \text{درجه آزادی درون گروهی یا } (df_w)$$

$$\text{تغییرات درون گروهی یا } (SSW) = \text{واریانس درون}$$

$$\text{گروهی } (MSW)$$

بر اساس ایده اساسی تحلیل واریانس می‌توان گفت که اگر

$$\text{واریانس درون گروهی} > \text{واریانس بین گروهی}$$

آنگاه میان میانگین گروه‌ها اختلاف وجود دارد. عبارت فوق را می‌توان به صورت نسبت بیان کرد یعنی اگر:

$$1 > (\text{واریانس درون گروهی} \div \text{واریانس بین گروهی})$$

در این صورت میان میانگین گروه‌ها اختلاف وجود دارد. عدد ۱ به این دلیل ملاک مقایسه است که تحت فرض صفر برابری میانگین گروه‌ها انتظار می‌رود که واریانس‌های

درون گروهی و بین گروهی برابر باشند و در نتیجه نسبت آنها برابر یک است. به عبارت دیگر اگر واریانس بین گروه‌ها با واریانس درون گروه‌ها برابر باشد آنگاه بین میانگین گروه‌ها اختلافی وجود ندارد. به این ترتیب می‌شود که شاخص سررنالد فیشر تشکیل می‌شود یعنی F که همان نسبت واریانس بین گروهی (بیان شده توسط اختلاف بین گروه‌ها) به واریانس درون گروهی (بیان نشده توسط اختلاف گروه‌ها یا خطا) است. در مثال بالا این شاخص عبارت است از:

$$1 < 27 = (27 \div 1) = \text{واریانس درون گروهی} \div$$

$$F = \text{واریانس بین گروهی}$$

با توجه به این که مقدار نسبت فوق برابر ۲۷ به دست آمد که بزرگتر از ۱ (مقدار ملاک مبتنی بر شانس) است، بنابراین به نظر می‌رسد که بین میانگین گروه‌ها اختلاف وجود داشته باشد. البته باید تصمیم‌گیری بر اساس مقایسه با توزیع نمونه‌برداری F شاخص فیشر یعنی توزیع F صورت گیرد. البته توزیع آماری شاخص فوق را سررنالد فیشر به دست آورده است که توزیع آن F است و چون دو واریانس یعنی واریانس‌های بین گروهی و درون گروهی به ترتیب در صورت و منخرج کسر F حضور دارند و هر کدام بر اساس درجه آزادی بین گروهی و درون گروهی محاسبه شده‌اند، توزیع F نیز دو درجه آزادی یک در صورت و یکی در منخرج دارد. برای این تصمیم که فرض صفر را رد یا پذیرفته شود باید مقدار محاسبه شده F با مقدار جدولی آن (جدول ۱۶) به ازای درجات آزادی صورت و منخرج به ترتیب برابر $df_B=2$ و $df_w=6$ و خطای ۰/۰۵ مقایسه گردد. مقدار این شاخص بر اساس جدول برابر ۵/۱۴ است.

با توجه به این که مقدار محاسبه شده F (یعنی ۲۷) بزرگتر از مقدار جدولی آن (یعنی ۵/۱۴) است (جدول ۱۶) بنابراین فرض صفر برابری میانگین‌ها رد می‌شود ($P < 0/05$) و این یعنی این که میانگین حداقل یکی از گروه‌ها با سایرین اختلاف دارد. همان‌طور که اشاره شد، تحلیل واریانس نمی‌تواند تعیین کند که کدام یک از گروه‌ها با هم اختلاف دارند و این موضوع باید توسط آزمون‌های تعقیبی (Post Hoc Tests) بررسی گردد.

چون $2/97 > | -3/67 |$ فرض صفر رد می‌شود و این یعنی که بین گروه‌های ۲ و ۳ اختلاف معنی‌داری وجود داشت.

نحوه محاسبه فاصله اطمینان

فاصله اطمینان مبتنی بر این آزمون را نیز می‌توان، به صورت زیر محاسبه کرد ([۱]) و بدین ترتیب اندازه اثر مطلق اختلاف و فاصله اطمینان ۹۵ درصدی آن را گزارش کرد:

$$((\text{انحراف معیار اختلاف}) \times t_{LSD} \pm \text{اختلاف در شاخص‌های نمونه‌ای}) = \text{فاصله اطمینان اختلاف‌ها}$$

که در آن مقدار شاخص t_{LSD} مقدار شاخص t مبتنی بر آزمون LSD و در واقع مقدار شاخص t جدول با درجه آزادی درون‌گروهی است و انحراف معیار اختلاف برابر جذر واریانس درون‌گروهی است که بدین ترتیب اثر گروه سوم نیز در محاسبات لحاظ می‌شود:

$$((\text{جذر واریانس درون‌گروهی}) \times t_{LSD} \pm \text{اختلاف در شاخص‌های نمونه‌ای}) = \text{فاصله اطمینان اختلاف‌ها}$$

برای $t_{LSD} = 2/97$ ، به عنوان مثال اختلاف بین گروه‌های ۱ و ۲ به صورت زیر به دست می‌آید:

$$\begin{aligned} & (2 - 5) \pm 2/97 \times 0/816 \\ & = (-3) \pm 2/42 \\ & = (-5/42, -0/58) \end{aligned}$$

نتیجه مقایسه دو به دوی گروه‌ها در جدول ۱۱ خلاصه شده‌اند:

گروه‌های مقایسه	LSD	$T_{6,0.05}$	P-Value	فاصله اطمینان ۹۵ درصدی اختلاف میانگین‌ها
۲-۱	-۳/۶۷	۲/۹۷	۰/۰۱	(-۵/۴۲, -۰/۵۸)
۳-۱	-۶/۱۲	۲/۹۷	۰/۰۰۰۹	(-۷/۴۲, -۲/۵۸)
۳-۲	-۳/۶۷	۲/۹۷	۰/۰۱	(-۵/۴۲, -۰/۵۸)

مورد). فاصله اطمینان ۹۵ درصدی اختلاف میانگین‌ها بر اساس اختلاف میانگین‌های ساده شده‌ی ۲، ۵ و ۸ به دست آمده‌اند ولی بر اساس میانگین‌های واقعی ۱۲، ۱۵ و ۱۸ نیز

آزمون‌های تعقیبی

در داده‌های مورد بررسی آزمون‌های تعقیبی برای انجام سه مقایسه به صورت زیر به کار می‌روند:

۱) آزمون کمترین تفاوت معنی‌داری

(Least Significant Difference or LSD)

الف - مقایسه گروه‌های ۱ و ۲:

$$(-3) \div \sqrt{(1(1 \div 3 + 1 \div 3))} = ((-3) \div 0/816) = -3/67$$

$$LSD_{12} = ((2$$

که شاخص فوق دارای توزیع t با ۶ درجه آزادی (df_w) است. با مراجعه به مقدار جدولی t با ۶ درجه آزادی و ۰/۰۵ خطا مقدار آن برابر ۲/۹۷ به دست می‌آید (جدول ۱۵) و در نتیجه چون $2/97 > | -3/67 |$ فرض صفر رد می‌شود و این یعنی که بین گروه‌های ۱ و ۲ اختلاف معنی‌داری وجود داشت.

ب - مقایسه‌ی گروه‌های ۱ و ۳

$$(-8) \div \sqrt{(1(1 \div 3 + 1 \div 3))} = ((-8) \div 0/816) = -6/12$$

$$LSD_{13} = ((2$$

چون $2/97 > | -6/12 |$ فرض صفر رد می‌شود و این یعنی که بین گروه‌های ۱ و ۳ اختلاف معنی‌داری وجود داشت.

ج - مقایسه‌ی گروه‌های ۲ و ۳

$$(-5) \div \sqrt{(1(1 \div 3 + 1 \div 3))} = ((-3) \div 0/816) = -3/67$$

$$LSD_{23} = ((2$$

آزمون LSD

در همه مقایسه‌ها بین همه گروه‌ها به صورت دو به دو اختلاف معنی‌داری وجود داشت ($P < 0/05$) در هر سه

مطلق اختلاف و فاصله اطمینان ۹۵ درصدی آن را گزارش کرد:

((انحراف معیار اختلاف) $\times t_{Bonferoni} \pm$ اختلاف در

شاخص‌های نمونه‌ای) = فاصله اطمینان اختلاف‌ها

که در آن $t_{Bonferoni}$ مقدار شاخص t مبتنی بر آزمون LSD و با اعمال تصحیح بونفرونی است و در واقع مقدار شاخص t جدول با درجه آزادی درون‌گروهی به ازای $0/017$ خطا است و انحراف معیار اختلاف برابر جذر واریانس درون گروهی است که بدین ترتیب اثر گروه سوم نیز در محاسبات لحاظ می‌شود:

((جذر واریانس درون‌گروهی) $\times t_{Bonferoni} \pm$ اختلاف در

شاخص‌های نمونه‌ای) = فاصله اطمینان اختلاف‌ها

برای $3/27 = t_{Bonferoni}$ ، به عنوان مثال اختلاف بین گروه‌های ۱ و ۲ به صورت زیر به دست می‌آید:

$$(2 - 5) \pm 3/27 \times 0/816$$

$$= (-3) \pm 2/67$$

$$= (-5/68, -0/32)$$

نتیجه مقایسه دو به دوی گروه‌ها در جدول ۱۲ خلاصه شده‌اند:

همین فاصله‌ها به دست می‌آیند. همان طور که ملاحظه می‌شود این فاصله‌ها، مقدار صفر یعنی ملاک عدم اختلاف تحت فرض صفر را در بر ندارند و این یعنی که بین دو به دوی گروه‌های مورد بررسی اختلاف معنی‌داری وجود داشت.

۲) تعدیل بونفرونی روی آزمون LSD

همان‌طور که اشاره شد برای تصحیح این اشکال بونفرونی روشی را پیشنهاد داده است که بر اساس آن نیاز است نرخ خطای اولیه متناسب با تعداد مقایسه‌ها کمتر از $0/05$ در نظر گرفته شود (برای سه مقایسه برابر $0/017 = 3/27$). بنابراین همه شرایط کاملاً شبیه آزمون LSD است با این تفاوت که مقدار جدولی t به جای خطای $0/05$ با خطای $0/017$ به دست می‌آید ($T_{6,0.017} = 3/27$)، و مقادیر اختلاف‌ها باید با این عدد مقایسه شوند.

نحوه محاسبه فاصله اطمینان

فاصله اطمینان مبتنی بر این آزمون را نیز می‌توان، به صورت زیر محاسبه کرد ([۱]) و بدین ترتیب اندازه اثر

جدول ۱۲- نتایج مقایسه RDW بین گروه‌های فعال ($n=3$)، فاز بهبودی ($n=3$) و کنترل ($n=3$) با استفاده از روش بونفرونی

گروه‌های مقایسه	LSD	$T_{6,0.017}$ بونفرونی	P-Value تصحیح شده بونفرونی	اختلاف میانگین‌ها (فاصله اطمینان ۹۵ درصدی)
۲-۱	- ۳/۶۷	۳/۲۷	۰/۰۳	(-۵/۶۷, -۰/۳۳)
۳-۱	- ۶/۱۲	۳/۲۷	۰/۰۰۰۲	(-۸/۶۷, -۳/۳۳)
۳-۲	- ۳/۶۷	۳/۲۷	۰/۰۳	(-۵/۶۷, -۰/۳۳)

آزمون بونفرونی

درهمه‌ی موارد فوق با توجه به این که مقدار شاخص آزمون بزرگتر از $3/27$ به دست آمده است فرض صفر رد می‌شود و این یعنی که بین دو به دوی گروه‌های مورد بررسی اختلاف معنی‌داری وجود داشت ($P < 0/017$) برای هر سه مورد). ولی همان‌طور که مشاهده می‌شود مقادیر P-Value در مقایسه با روش LSD بزرگتر هستند که توان کمتر این آزمون را نشان می‌دهد.

همان‌طور که ملاحظه می‌شود فاصله‌ها مقدار صفر یعنی ملاک عدم اختلاف تحت فرض صفر را در بر ندارند و این

یعنی که بین دو به دوی گروه‌های مورد بررسی اختلاف معنی‌داری وجود داشت. ولی این فاصله‌ها در مقایسه با آزمون LSD نرخ خطا تعدیل شده و پس از انجام سه مقایسه مورد نظر نرخ آن روی $0/05$ ثابت باقی می‌ماند (مزیت) ولی همان‌طور که ملاحظه می‌شود این فاصله‌ها بهتر از فاصله‌های LSD هستند و آزمون آن نیز محافظه‌کارانه‌تر از آزمون LSD است (عیب).

۳) روش کشف نرخ کاذب (False Discovery Rate) (FDR):

مورد انحراف معیار اختلاف باید برای مقایسه‌های دو به دو و با احتساب واریانس برای دو گروه مورد بررسی محاسبه شود ولی به دلیل اینکه واریانس درون گروهی برای هر گروه برابر ۱ به دست آمده بود، در این مورد هم انحراف معیار اختلاف به طور مشابه برابر $0/816 = (1 \div 3 + 1 \div 3)$ $\sqrt{1}$ به دست می‌آید.

به عنوان مثال، برای $t_{FDR,0.05} = 2/97$ فاصله اطمینان اختلاف بین گروه‌های ۱ و ۲ و به طور مشابه فاصله اطمینان اختلاف بین گروه‌های ۲ و ۳ به صورت زیر به دست می‌آید:

$$0/816 \times 2/97 \pm (2 - 5) \text{ یا } (5 - 8) \text{ برای مقایسه گروه}$$

های ۲ و ۳)

$$= (-3) \pm 2/42$$

$$= (-0/58, -0/42)$$

و برای $t_{FDR,0.033} = 2/75$ فاصله اطمینان اختلاف بین گروه‌های ۱ و ۳ به صورت زیر به دست می‌آید:

$$0/816 \times 2/75 \pm (2 - 8)$$

$$= (-6) \pm 2/24$$

$$= (-3/76, -8/24)$$

نتایج در جدول ۱۳ خلاصه شده است:

برای مقایسه گروه‌ها در این روش ابتدا تمام مقایسه‌های دوبه‌دوی ممکن با استفاده از آزمون t انجام شد و مقادیر P-Value محاسبه گردید. سپس P-Value ها به ترتیب از بزرگ به کوچک محاسبه شدند. سپس سطح معنی‌داری تعدیل شده به صورت

$$0/05 = 0/05 \times (3 \div 3) = \text{خطای تعدیل شده مقایسه‌های}$$

اول و دوم و مقایسه‌های دوم و سوم و

$$0/033 = 0/05 \times (2 \div 3) = \text{خطای تعدیل شده}$$

مقایسه‌های اول و سوم

به دست می‌آید.

نحوه محاسبه فاصله اطمینان

فاصله اطمینان مبتنی بر این آزمون را نیز می‌توان، به صورت زیر محاسبه کرد ([۱]) و بدین ترتیب اندازه اثر مطلق اختلاف و فاصله اطمینان ۹۵ درصدی آن را گزارش کرد:

$$(\text{انحراف معیار اختلاف}) \times t_{FDR} \pm \text{اختلاف در}$$

شاخص‌های نمونه‌ای) = فاصله اطمینان اختلاف‌ها

که در آن t_{FDR} مقدار شاخص t مبتنی بر روش FDR است و در واقع مقدار شاخص t جدول با درجه آزادی درون‌گروهی به ازای ۰/۰۵ و ۰/۰۳۳ خطا است. در این

جدول ۱۳- نتایج مقایسه RDW بین گروه‌های فعال ($n=3$)، فاز بهبودی ($n=3$) و کنترل ($n=3$) با استفاده از روش FDR

گروه‌های مقایسه	t	P-Value	سطح معنی‌داری تعدیل شده	نتیجه	اختلاف میانگین‌ها (فاصله اطمینان ۹۵ درصدی)
۲-۱	- ۵/۲۰	۰/۰۰۷	$0/05 = (3 \div 3) \times 0/05$	معنی‌دار	(-0/58, -0/42)
۳-۲	- ۵/۲۰	۰/۰۰۷	$0/05 = (3 \div 3) \times 0/05$	معنی‌دار	(-0/58, -0/42)
۳-۱	- ۸/۴۰	۰/۰۰۰۵	$0/033 = (2 \div 3) \times 0/05$	معنی‌دار	(-3/76, -8/24)

آزمون FDR

که در جدول فوق مقدار شاخص t همان مقدار شاخص آزمون t مستقل است. در هر یک از موارد مقایسه‌های فوق با توجه به این که مقدار P-Value محاسبه شده از سطح معنی‌داری تعدیل شده کوچکتر بود بنابراین نتیجه معنی‌دار حاصل شد و بین دو به دو میانگین گروه‌ها اختلاف وجود داشت. یا به طور مشابه، فاصله‌ها مقدار صفر یعنی

ملاک عدم اختلاف تحت فرض صفر را در بر ندارند و این یعنی که بین دو به دو گروه‌های مورد بررسی اختلاف معنی‌داری وجود داشت.

نحوه ارائه یافته‌ها

یافته‌های حاصل از تحلیل واریانس و آزمون تعقیبی FDR برای این مثال به صورت جدول ۱۴ ارائه می‌شود:

جدول ۱۴- نتایج مقایسه RDW بین گروه‌های فعال (n=۳)، فاز بهبودی (n=۳) و کنترل (n = ۳)

متغیر	گروه فعال	گروه فاز بهبودی	گروه کنترل	نتایج تحلیل واریانس	اختلاف میانگین‌ها (فاصله اطمینان \$ (%.۹۵)	درصد تغییرات #
RDW (%)	[۱] ۱۸	[۱] ۱۵	[۱] ۱۲	F(2,6)=27, P<0.001	(ب-ک) ۳- (-۰/۵۸ , -۵/۴۲)	(ب-ک) = ۲۵٪
					(ف-ب) ۳- (-۵/۴۲ , -۰/۵۸)	(ف-ب) = ۲۰٪
					(ف-ک) ۶- (-۳/۷۶ , -۸/۲۴)	(ف-ک) = ۵۰٪

آزمون t مستقل

*: P < ۰/۰۵

میانگین (انحراف معیار) گزارش شده است.

\$: فاصله اطمینان ۹۵٪ مبتنی بر تصحیح FDR

(ب-ک): مقایسه گروه‌های بهبودی-کنترل، (ف-ب): مقایسه گروه‌های فعال-بهبودی، (ف-ک): مقایسه گروه‌های فعال-کنترل،

درصد تغییرات نسبت به گروه کنترل با فرمول (اختلاف میانگین‌ها ÷ میانگین گروه کنترل × ۱۰۰) به دست می‌آید و تغییرات (اندازه اثر) نسبی (نسبت به گروه کنترل) نامیده می‌شود. در مقابل اختلاف میانگین‌ها، تغییرات (اندازه اثر) مطلق نامیده می‌شود.

ضمیمه ۲: جداول مقادیر شاخص‌های توزیع t و F

جدول ۱۵: مقادیر شاخص t

سطح معنی داری					درجه آزادی
۰/۰۰۰۵	۰/۰۰۵	۰/۰۱	۰/۰۲۵	۰/۰۵	
۰/۰۰۱	۰/۰۱	۰/۰۲	۰/۰۵	۰/۱۰	
۶/۸۷	۴/۰۳	۳/۳۶	۲/۵۷	۲/۰۲	۵
۴/۵۹	۳/۱۷	۲/۷۶	۲/۲۳	۱/۸۱	۱۰
۴/۰۷	۲/۹۵	۲/۶۰	۲/۱۳	۱/۷۵	۱۵
۳/۹۲	۲/۸۸	۲/۵۵	۲/۱۰	۱/۷۳	۱۸
۳/۸۵	۲/۸۵	۲/۵۳	۲/۰۹	۱/۷۲	۲۰
۳/۷۳	۲/۷۹	۲/۴۹	۲/۰۶	۱/۷۱	۲۵
۳/۶۵	۲/۷۵	۲/۴۶	۲/۰۴	۱/۷۰	۳۰
۳/۵۹	۲/۷۲	۲/۴۴	۲/۰۳	۱/۶۹	۳۵
۳/۵۵	۲/۷۰	۲/۴۲	۲/۰۲	۱/۶۸	۴۰
۳/۵۳	۲/۷۰	۲/۴۲	۲/۰۲	۱/۶۸	۴۳
۳/۵۲	۲/۶۹	۲/۴۱	۲/۰۱	۱/۶۸	۴۵
۳/۵۰	۲/۶۸	۲/۴۰	۲/۰۱	۱/۶۸	۵۰
۳/۴۶	۲/۶۶	۲/۳۹	۱/۹۹	۱/۶۷	۶۰
۳/۴۲	۲/۶۴	۲/۳۷	۱/۹۹	۱/۶۶	۸۰
۳/۳۹	۲/۶۳	۲/۳۶	۱/۹۸	۱/۶۶	۱۰۰
۳/۳۷	۲/۶۲	۲/۳۶	۱/۹۸	۱/۶۶	۱۲۰

جدول ۱۶- مقادیر شاخص F برای سطح معنی‌داری ۰/۰۵

		درجه آزادی مخرج																
		۵	۶	۱۰	۱۵	۲۰	۲۵	۳۰	۳۵	۳۷	۴۰	۴۲	۴۵	۵۰	۶۰	۸۰	۱۰۰	۱۲۰
درجه آزادی صورت	۲	۵/۷۹	۵/۱۴	۴/۱۰	۳/۶۸	۳/۴۹	۳/۳۹	۳/۳۲	۳/۲۷	۳/۲۵	۳/۲۳	۳/۲۲	۳/۲۰	۳/۱۸	۳/۱۵	۳/۱۱	۳/۰۹	۳/۰۷
	۳	۵/۴۱	۴/۷۶	۳/۷۱	۳/۲۹	۳/۱۰	۲/۹۹	۲/۹۲	۲/۸۷	۲/۸۶	۲/۸۴	۲/۸۳	۲/۸۱	۲/۷۹	۲/۷۶	۲/۷۲	۲/۷۰	۲/۶۸
	۴	۵/۱۹	۴/۵۳	۳/۴۸	۳/۰۶	۲/۸۷	۲/۷۶	۲/۶۹	۲/۶۴	۲/۶۳	۲/۶۱	۲/۵۹	۲/۵۸	۲/۵۶	۲/۵۳	۲/۴۹	۲/۴۶	۲/۴۵
	۵	۵/۰۵	۴/۳۹	۳/۳۳	۲/۹۰	۲/۷۱	۲/۶۰	۲/۵۳	۲/۴۹	۲/۴۷	۲/۴۵	۲/۴۴	۲/۴۲	۲/۴۰	۲/۳۷	۲/۳۳	۲/۳۱	۲/۲۹
	۶	۴/۹۵	۴/۲۸	۳/۲۲	۳/۷۹	۲/۶۰	۲/۴۹	۲/۴۲	۲/۳۷	۲/۳۶	۲/۳۴	۲/۳۲	۲/۳۱	۲/۲۹	۲/۲۵	۲/۲۱	۲/۱۹	۲/۱۸
	۷	۴/۸۸	۴/۲۱	۳/۱۴	۳/۷۱	۲/۵۱	۲/۴۰	۲/۳۳	۲/۲۹	۲/۲۷	۲/۲۵	۲/۲۴	۲/۲۲	۲/۲۰	۲/۱۷	۲/۱۳	۲/۱۰	۲/۰۹
	۸	۴/۸۲	۴/۱۵	۳/۰۷	۳/۶۴	۲/۴۵	۲/۳۴	۲/۲۷	۲/۲۲	۲/۲۰	۲/۱۸	۲/۱۷	۲/۱۵	۲/۱۳	۲/۱۰	۲/۰۶	۲/۰۳	۲/۰۲
	۹	۴/۷۷	۴/۱۰	۳/۰۲	۳/۵۹	۲/۳۹	۲/۲۸	۲/۲۱	۲/۱۶	۲/۱۴	۲/۱۲	۲/۱۱	۲/۱۲	۲/۱۰	۲/۰۷	۲/۰۴	۲/۰۰	۱/۹۷
	۱۰	۴/۷۴	۴/۰۶	۲/۹۸	۲/۵۴	۲/۳۵	۲/۲۴	۲/۱۶	۲/۱۱	۲/۱۰	۲/۰۸	۲/۰۶	۲/۰۵	۲/۰۳	۱/۹۹	۱/۹۵	۱/۹۳	۱/۹۱
	۱۵	۴/۶۲	۳/۹۴	۲/۸۵	۲/۴۰	۲/۲۰	۲/۰۹	۲/۰۱	۱/۹۶	۱/۹۵	۱/۹۲	۱/۹۱	۱/۸۹	۱/۸۷	۱/۸۴	۱/۷۹	۱/۷۷	۱/۷۵
	۲۰	۴/۵۶	۳/۸۷	۲/۷۷	۲/۳۳	۲/۱۲	۲/۰۱	۱/۹۳	۱/۸۸	۱/۸۶	۱/۸۴	۱/۸۳	۱/۸۱	۱/۷۸	۱/۷۵	۱/۷۰	۱/۶۸	۱/۶۶
	۲۵	۴/۵۲	۳/۸۳	۲/۷۳	۲/۲۸	۲/۰۷	۱/۹۶	۱/۸۸	۱/۸۲	۱/۸۱	۱/۷۸	۱/۷۷	۱/۷۵	۱/۷۳	۱/۶۹	۱/۶۴	۱/۶۲	۱/۶۰
	۳۰	۴/۵۰	۳/۸۱	۲/۷۰	۲/۲۵	۲/۰۴	۱/۹۲	۱/۸۴	۱/۷۹	۱/۷۷	۱/۷۴	۱/۷۳	۱/۷۱	۱/۶۹	۱/۶۵	۱/۶۰	۱/۵۷	۱/۵۵
	۳۵	۴/۴۸	۳/۷۹	۲/۶۸	۲/۲۲	۲/۰۱	۱/۸۹	۱/۸۱	۱/۷۶	۱/۷۴	۱/۷۲	۱/۷۰	۱/۶۸	۱/۶۶	۱/۶۲	۱/۵۷	۱/۵۴	۱/۵۲
	۳۷	۴/۴۷	۳/۷۸	۲/۶۷	۲/۲۱	۲/۰۱	۱/۸۸	۱/۸۰	۱/۷۵	۱/۷۳	۱/۷۱	۱/۶۹	۱/۶۷	۱/۶۵	۱/۶۱	۱/۵۶	۱/۵۳	۱/۵۱
	۴۰	۴/۴۶	۳/۷۷	۲/۶۶	۲/۲۰	۱/۹۹	۱/۸۷	۱/۷۹	۱/۷۴	۱/۷۲	۱/۶۹	۱/۶۸	۱/۶۶	۱/۶۳	۱/۵۹	۱/۵۴	۱/۵۲	۱/۵۰
	۴۲	۴/۴۶	۳/۷۷	۲/۶۶	۲/۲۰	۱/۹۹	۱/۸۶	۱/۷۸	۱/۷۳	۱/۷۱	۱/۶۹	۱/۶۷	۱/۶۵	۱/۶۳	۱/۵۹	۱/۵۴	۱/۵۱	۱/۴۹
	۴۵	۴/۴۵	۳/۷۶	۲/۶۵	۲/۱۹	۱/۹۸	۱/۸۶	۱/۷۷	۱/۷۲	۱/۷۰	۱/۶۷	۱/۶۶	۱/۶۴	۱/۶۱	۱/۵۷	۱/۵۲	۱/۴۹	۱/۴۷
	۵۰	۴/۴۴	۳/۷۵	۲/۶۴	۲/۱۸	۱/۹۷	۱/۸۴	۱/۷۶	۱/۷۰	۱/۶۸	۱/۶۶	۱/۶۵	۱/۶۳	۱/۶۰	۱/۵۶	۱/۵۱	۱/۴۸	۱/۴۶
	۶۰	۴/۴۳	۳/۷۴	۲/۶۲	۲/۱۶	۱/۹۵	۱/۸۲	۱/۷۴	۱/۶۸	۱/۶۶	۱/۶۴	۱/۶۲	۱/۶۰	۱/۵۸	۱/۵۳	۱/۴۸	۱/۴۵	۱/۴۳
۸۰	۴/۴۱	۳/۷۲	۲/۶۰	۲/۱۴	۱/۹۲	۱/۸۰	۱/۷۱	۱/۶۵	۱/۶۳	۱/۶۱	۱/۵۹	۱/۵۷	۱/۵۴	۱/۵۰	۱/۴۵	۱/۴۱	۱/۳۹	
۱۰۰	۴/۴۱	۳/۷۱	۲/۵۹	۲/۱۲	۱/۹۱	۱/۷۸	۱/۷۰	۱/۶۳	۱/۶۲	۱/۵۹	۱/۵۷	۱/۵۵	۱/۵۲	۱/۴۸	۱/۴۳	۱/۳۹	۱/۳۷	
۱۲۰	۴/۴۰	۳/۷۰	۲/۵۸	۲/۱۱	۱/۹۰	۱/۷۷	۱/۶۸	۱/۶۲	۱/۶۰	۱/۵۸	۱/۵۶	۱/۵۴	۱/۵۱	۱/۴۷	۱/۴۱	۱/۳۸	۱/۳۵	

سپاسگزاری

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

مأخذ

1. 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 Under Press [In Persian].
2. Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Probability and Distributions. *Journal of Diabetes and Lipid Disorders* 2013 12(2): 83-100 Under Review [in Persian].
3. Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Summarizing and Displaying Data. *Journal of Diabetes and Lipid Disorders* 2013 12(2): 83-100 Under Press [in Persian].
4. Godarzi M, Ebrahimzadeh E, Rabi A, SaediPour B, Asghari Jafarabadi M. Evaluation the relationship Between Knowledge, Attitude and Participation with Self Efficacy in Type 2 Diabetes Patients in Karaj. *Journal of Diabetes and Lipid Disorders* 2012 11(3):269-81[In Persian].
5. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B* 1995 57(1):289-300.
6. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York Wiley and Sons 1986.
7. Abedimanesh N, Alipour B, Soumi MH, Ostadrahimi A, Abedimanesh S, Asghari Jafarabadi M, et al. The relationship between Red Cell Distribution with Levels and the Disease Activity in Ulcerative Colitis. *Govaresh* 2013 17(3):183-8.