

سری آمار: مقدمه‌ای بر آمار استنباطی (بر آورد عددی، فاصله اطمینان و آزمون فرض)

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

چکیده

مقدمه: ارائه برآوردی از پارامترهای جامعه و آزمون‌های آماری را، تقریباً در همه مطالعات علوم پزشکی می‌توان یافت. هدف از این مقاله، معرفی مفاهیم و روش‌های ساده و کاربردی برآورد عددی، فاصله اطمینان و آزمون فرض در این مطالعات بود. روش‌ها: مبانی و نحوه محاسبه برآورد عددی و فاصله اطمینان، مفاهیم خطای معیار برای داده‌های کمی و کیفی ارائه گردید. مفاهیم آزمون فرض، شامل نحوه تبدیل فرضیات پژوهشی به فرضیات آماری، انواع خطاها در آزمون فرض، P-value، تصمیم‌گیری بر اساس فاصله اطمینان، معنی‌داری آماری در مقابل اهمیت بالینی و مسئله چند آزمونی بررسی شد. یافته‌ها و نتیجه‌گیری: برای متغیرهای کیفی و کمی، برآورد عددی و فاصله اطمینان پارامترهای جامعه توسط نسبت نمونه‌ای و میانگین به عنوان برآوردهای نا اریب، محاسبه و ویژگی‌های آنها ارزیابی شد. براساس مثال‌های مطالعاتی، فرضیات پژوهشی به فرضیات آماری تبدیل شدند. برای تصمیم‌گیری برای فرضیات، شاخص آزمون و در نتیجه، P-value محاسبه و برای مقادیر کمتر از ۰/۰۵ این شاخص، یا به طور معادل اگر فاصله اطمینان مقدار تحت فرض صفر را در بر نداشته باشد، اختلاف یا رابطه معنی‌دار حاصل می‌شود. علاوه بر معنی‌داری آماری باید به اهمیت بالینی آن نیز توجه کرد. استفاده بیش از یک آزمون برای یک فرضیه، چند آزمونی نامیده می‌شود و در چنین وضعیتی نرخ خطا افزایش می‌یابد که باید تصحیح گردد.

واژگان کلیدی: آمار استنباطی، برآورد، فاصله اطمینان، آزمون فرض، P-Value، چندآزمونی

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

* نشانی: تبریز، خیابان گلگشت، خیابان عطار نیشابوری، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، گروه آمار و اپیدمیولوژی، کدپستی: ۵۱۶۶۱۴۷۱۱، تلفن: ۰۲-۴۱۱۳۳۵۷۵۸۰، نامبر: ۰۴۱۱۳۳۴۰۶۳۴، پست الکترونیک: asgharimo@tbzmed.ac.ir

آمار استنباطی

در تعریف علم آمار بیان می‌شود آمار علم جمع‌آوری، طبقه‌بندی و تلخیص داده‌ها، انتخاب بخشی از جامعه به عنوان نمونه، تجزیه و تحلیل داده‌های نمونه‌ای، تعمیم نتایج به جامعه و محاسبه خطای تعمیم‌پذیری است. بخش دوم از تعریف، آمار استنباطی نام دارد. در بین منابع اطلاعاتی شامل: سرشماری، آمارهای ثبتی و نمونه‌گیری، بهترین منبع برای انجام آمار استنباطی، نمونه‌گیری است که با تعریف آن هم همخوانی دارد ولی برای آمارهای ثبتی (نظیر داده‌های مراکز بهداشت یا ثبت سرطان [۱]) هم با اغماض آمار استنباطی انجام می‌شود (در مورد سرشماری‌ها معمولاً گزارش یافته‌ها با استفاده از روش‌های آمار توصیفی انجام می‌شود).

تعاریف و مفاهیم اولیه

جامعه آماری^۱ شامل همه افرادی است که هدف، اندازه‌گیری ویژگی‌های مطالعه در آنهاست که معمولاً منظور از آن، جامعه هدف^۲، یعنی جامعه‌ای است که هدف، انجام توصیف یا استنباط آماری درباره آن است، مثلاً در مطالعه گودرزی و همکاران [۲]، همه افراد دیابتی نوع ۲ مراجعه کننده به انجمن دیابت شهر کرج جامعه هدف مطالعه را تشکیل دادند. به صفت مورد ارزیابی در جامعه اصطلاحاً پارامتر^۳ گفته می‌شود که توصیفی از ویژگی مورد نظر در جامعه فراهم می‌کنند؛ مثلاً پارامتر میانگین و انحراف معیار (به ترتیب μ و σ) متغیر سن یا پارامتر نسبت مردان (P) در جامعه فوق برآورد گردید.

در مطالعات، جز در مواردی که جامعه آماری کوچک باشد، معمولاً انتخاب همه افراد موجود در جامعه آماری امکان‌پذیر نیست (به دلیل محدودیت در هزینه، زمان و نیروی انسانی و همچنین دستیابی به دقت بیشتر). بنابراین بخشی از جامعه مورد مطالعه^۴ (یعنی جامعه‌ای که نمونه‌گیری از آن انجام می‌شود)، براساس لیست

نمونه‌گیری^۵ انتخاب شده و بررسی‌های لازم در این قسمت کوچک‌تر انجام می‌شود. به این بخش کوچک نمونه^۶ گفته می‌شود و انتظار می‌رود که این نمونه نماینده‌ای از افراد موجود در جامعه باشد، یعنی خصوصیات مشابه با جامعه مورد مطالعه داشته باشد و هر نتیجه‌ای که از این نمونه حاصل شود به کل جامعه تعمیم داده می‌شود. برای دستیابی به نماینده بودن، نمونه نیاز است که از روش‌های نمونه‌گیری تصادفی (با شانس برابر افراد جامعه برای شرکت در نمونه) استفاده شود، ولی در برخی از موقعیت‌ها ممکن است امکان انتخاب نمونه تصادفی وجود نداشته باشد و یک نمونه آسان یا در دسترس انتخاب گردد. به هر حال، روش‌های مختلفی برای انتخاب نمونه تصادفی (شامل نمونه‌گیری تصادفی ساده، سیستماتیک، طبقه‌ای و خوشه‌ای) ارائه شده است که هدف از به کارگیری این روش‌ها، کمتر نمودن سوگرایی در انتخاب افراد و دستیابی به نمونه‌ای نماینده برای جامعه است. روش‌های مختلف نمونه‌گیری و تعیین تعداد مناسب نمونه در سری‌های آتی ارائه خواهد شد.

به هر حال، پارامتر جامعه توسط یک کمیت نمونه‌ای متناظر با آن، که اصطلاحاً شاخص یا آماره^۷ نامیده می‌شود، برآورد می‌گردد؛ مثلاً برای تخمین پارامترهای جامعه فوق، میانگین نمونه‌ای (\bar{x}) و انحراف معیار نمونه‌ای (SD یا S) متغیر سن، به ترتیب برابر ۵۰/۰۶ و ۱۰/۲۳، یا پارامتر نسبت مردان (P) بر اساس نمونه برابر ۲۱ درصد به دست آمد [۲].

حوزه‌های استنباط آماری شامل برآورد و آزمون فرض می‌باشد. در حوزه برآورد، هدف یافتن یک عدد یا یک فاصله برای صفت مورد نظر در جامعه است که به ترتیب برآورد عددی و فاصله اطمینان را فراهم می‌کنند. در بحث آزمون فرض، هدف ارزیابی یک ادعا درباره ویژگی مورد بررسی (یا ارزیابی فرضیه پژوهشی) است.

5. Sampling Frame
6. Sample
7. Statistics

1. Statistical Population
2. Target Population
3. Parameter
4. Study Population

برآورد عددی و ویژگی های آن

همانطور که ملاحظه شد، برای تخمین ویژگی مورد نظر در جامعه براساس نمونه یک عدد گزارش می شود که به آن برآورد عددی (نقطه ای) گفته می شود. برآورد نقطه ای بهترین تخمین برای برآورد پارامتر مورد نظر در جامعه است. برای محاسبه برآورد عددی از تکنیک های آمار توصیفی برای محاسبه انواع شاخص ها استفاده می شود [۳]. انتظار می رود، عددی که براساس نمونه برآورد می شود تا حد ممکن به پارامتر جامعه نزدیک باشد. برای دستیابی به این هدف، نیاز است تا چند ویژگی در برآوردگر مورد نظر برقرار باشد. برای بررسی این ویژگی ها و روشن شدن موضوع، مثال زیر را ببینید.

مثال ۱: به منظور دستیابی به اهداف مورد نظر در قالب محاسبات ساده، برای این مثال یک جامعه آماری کوچک با اعداد ۱، ۲، ۳ و ۴ در نظر بگیرید. در صورت انجام سرشماری، میانگین و واریانس این جامعه به ترتیب برابر $2/5$ و $1/25$ به دست می آید (برای محاسبه میانگین و

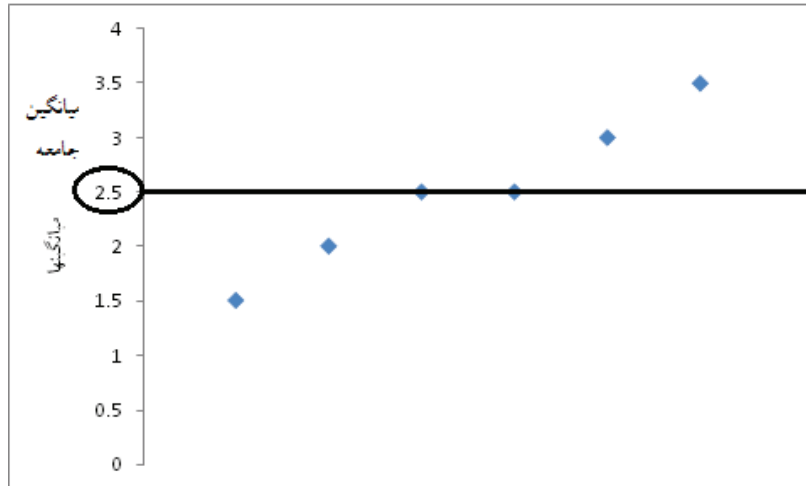
واریانس، از فرمول های موجود در [۳] استفاده شد). حال فرض کنید محققین مختلف، نمونه هایی به تصادف و با تعداد ۲ از این جامعه انتخاب نمایند. این امکان وجود دارد که برای محقق اول، آزمودنی اول و دوم در این نمونه قرار گیرند که شامل اعداد ۱ و ۲ هستند که برای این دو عدد میانگین و واریانس نمونه ای (به عبارت دیگر، برآورد میانگین و واریانس در این جامعه آماری بر اساس نمونه انتخابی) به ترتیب برابر $1/5$ و $0/5$ به دست می آید. محقق دوم از این جامعه آماری نمونه گیری می کند و مثلاً نمونه های ۱ و ۳ در نمونه او قرار می گیرند که برای این دو عدد میانگین و واریانس نمونه ای (به عبارت دیگر، برآورد میانگین و واریانس در این جامعه آماری بر اساس نمونه انتخابی) به ترتیب برابر 2 و 2 به دست می آید. ممکن است برای محققین مختلف نمونه های مختلفی انتخاب شود. همه نمونه های ممکن دوتایی از این جامعه در جدول ۱ ارائه شده اند:

جدول ۱- نمونه های دوتایی ممکن در جامعه آماری و برآورد میانگین و واریانس ها در هر نمونه

شماره نمونه ها	۱	۲	۳	۴	۵	۶
اعداد موجود در نمونه	۱	۱	۱	۲	۲	۳
میانگین	$1/5$	$2/5$	$2/5$	$3/5$	$3/5$	$3/5$
واریانس	$0/5$	$2/5$	$4/5$	$0/5$	$2/5$	$0/5$
انحراف معیار	$0/71$	$1/41$	$2/12$	$0/71$	$1/41$	$0/71$

دارد. برای هر یک از نمونه ها، میانگین، واریانس و انحراف معیار نیز ارائه شده است. میانگین های نمونه ای روی نمودار به صورت زیر ارائه می شود:

همانطور که ملاحظه می شود، در صورتی که نمونه های ممکن (غیر تکراری) مد نظر باشد در انتخاب نمونه های ۲ تایی از جامعه ۴ تایی، ۶ نمونه ممکن غیر تکراری وجود



نمودار ۱- مقادیر میانگین‌های نمونه‌ای در ۶ نمونه ممکن دوتایی از یک جامعه با ۴ عضو

معیار) بر اساس تنها یک نمونه به دست می‌آید و با توجه به این ویژگی، برآوردهای حاصل، بهترین تخمین برای پارامترهای جامعه هستند.

۳) میانگین نمونه به عنوان یک متغیر دارای توزیع نمونه‌گیری است. توزیع نمونه‌گیری تجربی آن را می‌توان به صورت زیر مشاهده کرد:

۳/۵	۳	۲/۵	۲	۱/۵	مقادیر میانگین:
۱/۶	۱/۶	۲/۶	۱/۶	۱/۶	احتمال رخداد:

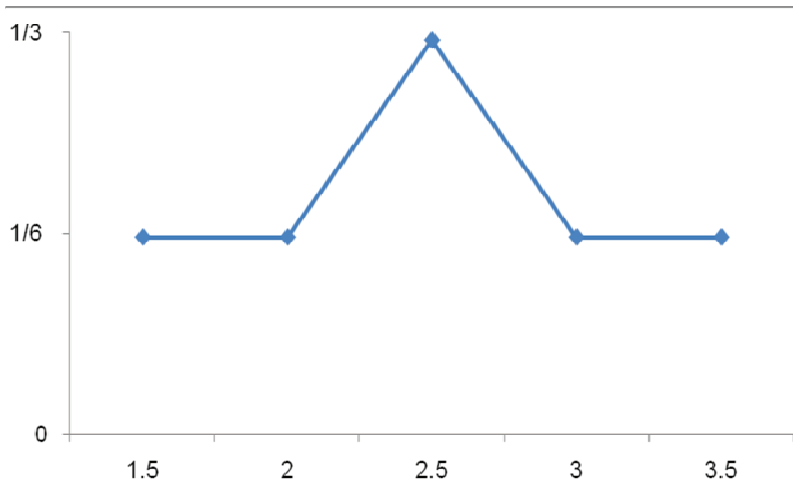
یعنی میانگین ۱/۵، ۲، ۳ و ۳/۵ هر کدام یک بار از ۶ بار و میانگین ۲/۵ دو بار از ۶ بار رخ داده است. برای توزیع فوق، نمودار زیر را می‌توان رسم نمود (نمودار ۲):

خط وسط نمودار ۱، روی نقطه ۲/۵ یعنی نقطه میانگین جامعه قرار گرفته است. همان‌طور که ملاحظه می‌شود میانگین نمونه‌های مختلف در اطراف میانگین جامعه متغیر هستند. اما همان‌طور که اشاره شد، در یک نمونه خاص، برآورد نقطه‌ای بهترین تخمین برای برآورد پارامتر مورد نظر در جامعه است. بنابراین:

۱) میانگین نمونه‌ای یک متغیر است یعنی از نمونه‌ای به نمونه دیگر تغییر می‌کند (مطالعات مختلف انجام شده در یک جامعه آماری برآوردهای مختلفی از یک ویژگی ارائه می‌کنند).

۲) متوسط میانگین‌های نمونه‌ای (ردیف میانگین در جدول ۲) برابر ۲/۵ می‌شود (۱۵ یعنی مجموع تقسیم بر ۶) که دقیقاً برابر میانگین جامعه است. به عبارت دیگر انتظار می‌رود (به طور متوسط)، شاخص محاسبه شده در نمونه روی پارامتر جامعه قرار گیرد، یعنی شاخص نمونه، پارامتر جامعه را بیش و کم برآورد نمی‌کند. به این خصوصیت اصطلاحاً ناریبی^۱ گفته می‌شود. میانگین و واریانس نمونه‌ای، برآوردهای ناریب میانگین و واریانس جامعه هستند. در عمل فقط یک نمونه انتخاب می‌شود و همه نمونه‌های ممکن یا سری نمونه‌ها در یک مطالعه وجود ندارد، بنابراین میانگین و واریانس (انحراف

1. Unbiaseness



نمودار ۲- توزیع تجربی میانگین‌های نمونه‌ای در ۶ نمونه ممکن دوتایی از یک جامعه با ۴ عضو

کرد که برای نمونه‌های بزرگ توزیع نمونه‌گیری میانگین نمونه، تقریباً نرمال با میانگین و انحراف معیار برابر میانگین جامعه و خطای معیار می‌باشد.

لازم به ذکر است که در عمل فقط یک نمونه انتخاب می‌شود (همه نمونه‌های ممکن یا سری نمونه‌ها در یک مطالعه وجود ندارد)، بنابراین واریانس (انحراف معیار) میانگین بر اساس تنها یک نمونه به دست می‌آید. بر اساس نمونه انتخابی، برآورد خطای معیار برابر با انحراف معیار نمونه تقسیم بر جذر تعداد نمونه است (انحراف معیار جذر n). با داشتن رادیکال n در مخرج کسر هر چه نمونه بزرگتر انتخاب شود، مخرج کسر بزرگتر و در نتیجه کل کسر یعنی خطای معیار کوچکتر و دقت آن بیشتر می‌شود (یعنی به ازای هزینه بیشتر، دقت بیشتر کسب می‌گردد). برای روشن‌تر شدن موضوع، در مثال فوق نمونه‌هایی به تعداد ۳ انتخاب شد و نتایج برای همه نمونه‌های ممکن، در جدول ۲ ارائه شده است:

جدول ۲- نمونه‌های سه‌تایی ممکن در جامعه آماری و برآورد

میانگین و واریانس‌ها در هر نمونه			
شماره نمونه‌ها	۱	۱	۲
اعداد موجود	۲	۲	۳
در نمونه	۳	۴	۴
میانگین	۲	۲/۳۳	۳
واریانس	۱	۲/۳۳	۱
انحراف معیار	۱	۱/۵۳	۱

همان‌طور که ملاحظه می‌شود، در این نمودار، منحنی شکسته توزیع نمونه‌گیری میانگین بسیار شبیه توزیع نرمال است که با قضیه حد مرکزی مطابقت دارد. قضیه حد مرکزی بیان می‌کند که توزیع اولیه داده‌ها هر چه که باشد، برای حجم نمونه‌های بزرگ ($n \geq 30$)، توزیع نمونه‌گیری میانگین نرمال می‌شود [۴]. حجم نمونه‌ی این مثال، برابر ۲ بود ولی در حجم نمونه‌های بالاتر منحنی شکسته نخواهد بود و منحنی هموارتری حاصل می‌شود. بنابراین، می‌توان این گونه نتیجه گرفت که در عمل، برای تنها یک نمونه و به ازای حجم نمونه‌های بزرگ ($n \geq 30$) توزیع نمونه‌گیری میانگین نرمال می‌شود. این موضوع برای استنباط درباره میانگین (و همچنین نسبت‌ها به عنوان میانگین متغیرهای کیفی) کاربرد زیادی دارد.

۴) در عمل میزان دقت میانگین نمونه‌ای و میزان تغییر پذیری آن حین تصمیم‌گیری درباره آن بسیار مهم است. بنابراین، نیاز است از تغییر پذیری این شاخص اطلاعاتی در دست باشد. در مثال فوق، ۶ مقدار برای میانگین به دست آمد؛ واریانس (انحراف معیار) توزیع نمونه‌گیری میانگین‌های فوق برابر $0/5$ ($0/71$) به دست می‌آید که این انحراف معیار، خطای معیار^۱ (SE یا SEM) نامیده می‌شود. در نتیجه، خطای معیار، انحراف معیار میانگین و به عبارت دقیق‌تر، انحراف معیار توزیع نمونه‌گیری میانگین است. بنابراین براساس این نکته و نکات ۲ و ۳، می‌توان استنباط

1 Standard Error

استنباطی است و یک معیار توصیفی نیست، بنابراین در ارائه کردن داده‌ها از انحراف معیار استفاده می‌شود [۳] ولی زمانی که هدف انجام استنباط درباره میانگین باشد (آزمون فرض و فاصله اطمینان)، در این صورت خطای معیار استفاده می‌شود. در عمل به دلیل این که خطای معیار کوچکتر از انحراف معیار است، و با تصور این که پراکندگی را کمتر نشان می‌دهد، به اشتباه، به جای انحراف معیار استفاده می‌شود. بنابراین توصیه می‌شود صرفاً در مسایل استنباطی و آن هم در قالب فاصله اطمینان (بخش بعدی) از خطای معیار استفاده شود [۳، ۵]. خطای معیار شاخص‌های نمونه‌ای مختلف فرمول‌های خاص خود را دارند (جدول ۳).

واریانس (انحراف معیار) توزیع نمونه‌گیری میانگین‌های فوق برابر $0.26 / 0.51$ به دست می‌آید که تغییرپذیری کمتر (دقت بیشتر) میانگین را به ازای افزایش (حتی یک) نمونه نشان می‌دهد. دلیل آن، این است که با نمونه‌های بزرگتر اثر داده‌های پرت با حضور تعداد بیشتر داده‌ها کم‌رنگ‌تر می‌شود.

انحراف معیار یا خطای معیار

انحراف معیار تغییر پذیری داده‌ها در جامعه را نشان می‌دهد و خطای معیار، تغییرپذیری میانگین را در اطراف میانگین جامعه، و به عبارت دقیق‌تر، انحراف معیار توزیع نمونه‌گیری میانگین را نشان می‌دهد. خطای معیار یک

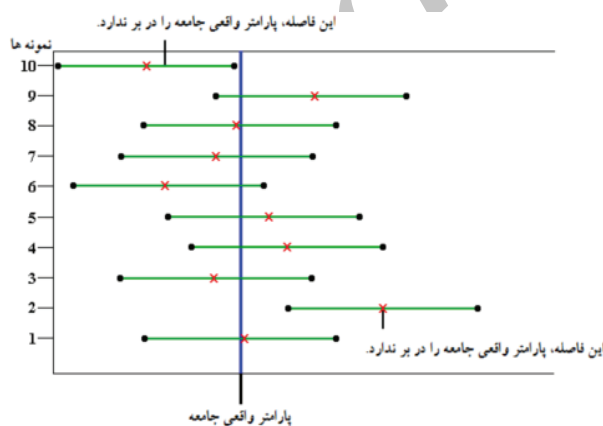
جدول ۳- فرمول خطای معیار برای شاخص‌های مختلف

نوع متغیر	توزیع احتمالی	پارامتر جامعه	نماد	شاخص نمونه	فرمول خطای معیار	فرمول ریاضی
کمی پیوسته	نرمال	میانگین	μ	\bar{x}	انحراف معیار \div جذر n	$s \div \sqrt{n}$
کیفی دوحالتی	دوجمله‌ای	نسبت	π	p	جذر((نسبت \times یک منهای نسبت) \div n)	$\sqrt{p \times (1-p) \div n}$
کمی گسسته (شمارشی)	پواسن	میزان	λ	r	جذر (میزان \div n)	$\sqrt{r \div n}$

اطمینان مبتنی بر ۱۰ سری نمونه و مقدار واقعی پارامتر جامعه نشان داده شده است.

فاصله اطمینان

زمانی که یک عدد به عنوان برآوردی از پارامتر جامعه معرفی می‌شود، خود به تنهایی برای معرفی صفت مورد بررسی کافی نیست و همراه با آن باید شاخص پراکندگی آن هم گزارش شود تا میزان دقت آن هم به صورت هم‌زمان بررسی شود. بیان یک برآورد به همراه شاخص پراکندگی در کنار آن که معمولاً به صورت واحدی از شاخص پراکندگی در سمت چپ و راست برآوردگر بیان می‌شود، ایده‌ای برای تعریف یک فاصله برای پارامتر جامعه را فراهم می‌کند. به عبارت دیگر، بر اساس این فاصله می‌توان دامنه‌ای از مقادیر را می‌توان تعریف نمود که با یک سطح معلومی از اطمینان، پارامتر جامعه می‌تواند آن مقادیر را بپذیرد. بنابراین فاصله اطمینان برآوردی از پارامتر جامعه و میزانی از دقت آن را به طور هم‌زمان ارائه می‌کند و هرچه این فاصله کوتاهتر باشد دقت بیشتری را در برآورد پارامتر جامعه فراهم می‌کند. در شکل ۱، فواصل



شکل ۱- فواصل اطمینان مبتنی بر ۱۰ سری نمونه و مقدار واقعی پارامتر جامعه

در این شکل، از ۱۰ فاصله، ۸ فاصله، پارامتر واقعی جامعه را در بر ندارند (یعنی ۸۰ درصد اطمینان). بنابراین، مفهوم

((انحراف معیار میانگین) $\times 2 \pm$ میانگین) = فاصله اطمینان میانگین یا

((خطای معیار) $\times 2 \pm$ میانگین) = فاصله اطمینان میانگین یا

((\sqrt{n} \pm انحراف معیار داده‌ها) $\times 2 \pm$ میانگین) = فاصله اطمینان میانگین

برای مثال، در داده‌های جدول ۳ و بر اساس نمونه سری اول، که در آن میانگین برابر ۲ به دست آمد و انحراف معیار جامعه که برابر $\sqrt{1/25}$ یعنی $1/5$ بود، فاصله اطمینان به صورت زیر محاسبه می‌شود:

$$\begin{aligned} & ((\sqrt{3}) \pm 2 \times (1/5)) = \text{حد پایین} \\ & = (0.71 \text{ تا } 3.29) \end{aligned}$$

بنابراین با اطمینان ۹۵ درصد می‌توان گفت که میانگین جامعه آماری مورد بررسی در فاصله ۰/۷۱ و ۳/۲۹ قرار می‌گیرد (تفسیر مبتنی بر نمونه انفرادی).

طول فاصله اطمینان، برابر با حد بالای فاصله منهای حد پایین فاصله اطمینان است. در مثال فوق طول فاصله اطمینان برابر $2/58 (= 3/15 - 0/85)$ به دست می‌آید. هر چه طول فاصل اطمینان کوتاه‌تر باشد، یعنی محدوده کوچک‌تری از مقادیر برای پارامتر جامعه در نظر گرفته شده است که تغییرپذیری کمتر میانگین نمونه‌ای در برآورد میانگین جامعه و دقت بیشتر آن را نشان می‌دهد. در این مثال، به دلیل حجم نمونه کوچک، طول فاصله اطمینان پهن است.

فرمول کلی فاصله اطمینان

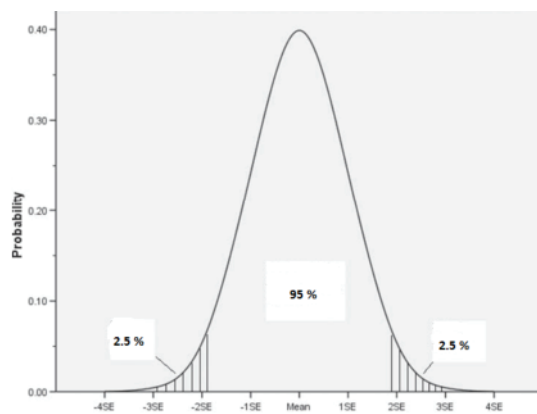
برای محاسبه فاصله‌هایی با اطمینان بیشتر یا کمتر نیاز است از عددهای Z متناظر با سطوح دیگر اطمینان در توزیع نرمال استفاده شود. برای اطمینان‌های ۹۹ و ۹۰ درصد (به جای ۹۵ درصد) به ترتیب اعداد $2/58 \pm$ و $1/65 \pm$ شاخص Z نرمال (به جای $2 \pm$)، به کار می‌روند. فرمول کلی فاصله اطمینان به صورت زیر است:

((خطای معیار) $\times Z \pm$ میانگین) = فاصله اطمینان میانگین که در آن Z نمادی برای مقادیر (چندک) توزیع نرمال می‌باشد. مثلاً برای مثال فوق، فاصله اطمینان‌های ۹۹ و ۹۰ درصد به ترتیب برابر $(3/49)$ و $(0/51)$ و $(2/95)$ و $(1/91)$ به

واقعی فاصله اطمینان را در قالب نمونه‌های تکراری می‌توان تفسیر نمود؛ یعنی برای تفسیر یک فاصله اطمینان ۹۵ درصد می‌توان گفت اگر ۱۰۰ بار نمونه انتخاب شود و بر اساس فاصله اطمینان ۹۵ درصدی ساخته شود، ۹۵ تا از این فاصله‌ها پارامتر واقعی جامعه را در بر می‌گیرند و ۵ مورد در بر نمی‌گیرند و در واقع این استنباط آماری با ۵ درصد خطا انجام می‌شود. در نتیجه، فاصله اطمینان انفرادی ساخته شده بر اساس یک نمونه ممکن است پارامتر واقعی جامعه را در بر داشته باشد یا نداشته باشد.

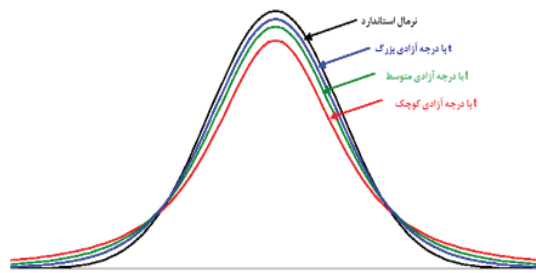
مولفه‌هایی که در محاسبه فاصل اطمینان دخیل هستند، پراکندگی موجود در داده‌ها یعنی انحراف معیار، حجم نمونه، و میزان اطمینان که در واقع همان شانس یا تصادف موجود در داده‌ها است که به ترتیب با طول فاصله اطمینان روابط مستقیم، معکوس و مستقیم دارند.

برای محاسبه ساده فاصله اطمینان ۹۵٪، همان طور که در [۴] اشاره شد، در توزیع نرمال ۹۵ درصد مشاهدات بین $2 \pm$ انحراف معیار از مقدار میانگین قرار می‌گیرند. بر اساس نکات ۲ و ۴ ذکر شده فوق، در توزیع نرمال، ۹۵ درصد میانگین‌ها بین $2 \pm$ انحراف معیار از میانگین جامعه قرار می‌گیرند (شکل ۲) (البته مقدار دقیق آن $1/96 \pm$ است ولی برای سادگی محاسبات، مقدار تقریبی $2 \pm$ در محاسبات استفاده خواهد شد).



شکل ۲- نحوه توزیع احتمال در دامنه‌های ۲ انحراف معیار میانگین (خطای معیار) در اطراف آن

بنابراین یک فاصله اطمینان ۹۵ درصدی برای میانگین جامعه با فرمول زیر به دست می‌آید:



شکل ۳- توزیع t با درجات آزادی مختلف و توزیع نرمال استاندارد

درجه آزادی، آزادی در انتخاب مقادیر نمونه است. به عنوان مثال در صورتی که هدف یافتن ۵ عدد باشد که میانگین آنها برابر ۱۰ شود و یا به طور معادل مجموع آنها برابر ۵۰ شود، ۴ عدد اول را آزادانه می‌توان انتخاب کرد ولی عدد پنجم باید به گونه‌ای انتخاب شود که حاصل مجموع ۴ عدد قبلی و این عدد برابر ۵۰ شود. بنابراین در انتخاب ۵ نمونه، ۴ مقدار آزادانه انتخاب شد ولی عدد پنجم با اعمال محدودیت محاسبه شد. در نتیجه درجه آزادی در ۵ نمونه برابر ۴ ($= 5 - 1$) شد. در مسئله برآورد میانگین در یک جامعه، درجه آزادی برابر $n - 1$ و در حالت کلی درجه آزادی برابر (تعداد محدودیت‌ها - تعداد نمونه) می‌شود. درجات آزادی چون متعادل حجم نمونه هستند، درجات آزادی بالاتر دقت بیشتر در استنباط را فراهم می‌کند.

در مورد تاریخچه t باید گفت که این توزیع توسط ویلیام سیلی گوست^۱ (۱۹۰۸)، ارائه شد و به دلیل این که شرکت ایرلندی گاینس بروری^۲ یعنی شرکت کارفرمای او، اجازه چاپ مقاله با نام اصلی خود را نمی‌داد، مقاله خود را در این مورد این توزیع در مجله بیومتریکا^۳ با نام مستعار استودنت چاپ کرد و به همین دلیل این توزیع، t استودنت نام گرفت [۶].

برآورد نقطه‌ای، فاصله‌ای و خطای معیار برای نسبت‌ها و

میزان‌ها

در صورتی که هدف محاسبه فاصله اطمینان برای شاخص‌های دیگری نظیر نسبت‌ها یا میزان رخدادها

دست می‌آید. همان‌طور که ملاحظه می‌شود، طول فاصله اطمینان ۹۰ درصدی کوتاه‌تر است چون برای اطمینان ۹۰ درصد یعنی اطمینان کمتر، به فاصله کوتاه‌تری برای قرار مقادیر پیشنهادی میانگین جامعه نیاز است (برای گرفتن ماهی کمتر، تور کوچک‌تری نیاز است).

توزیع t

در محاسبه فاصله اطمینان برای میانگین با توجه به معلوم بودن مقدار انحراف معیار در جامعه، در محاسبات فاصله اطمینان از این انحراف معیار معلوم استفاده شد و بر اساس توزیع نرمال برای اطمینان ۹۵ درصد، از عدد ۲ برای Z نرمال استفاده شد. ممکن است این سوال پیش آید که اگر انحراف معیار جامعه معلوم نباشد، که در اغلب مطالعات چنین است، چه باید کرد؟ در چنین مواردی به جای انحراف معیار جامعه از انحراف معیار مبتنی بر نمونه استخراجی استفاده می‌شود ولی در این صورت باید به جای Z نرمال در فرمول زیر:

((خطای معیار جامعه) \pm میانگین) = فاصله اطمینان میانگین از t استفاده می‌شود یعنی:

((خطای معیار نمونه‌ای) $\pm t \times$ میانگین) = فاصله اطمینان میانگین که در آن t نمادی برای مقادیر (چندک) توزیع t می‌باشد.

این توزیع شبیه توزیع نرمال متقارن است ولی نسبت به توزیع نرمال ارتفاع پایین‌تری دارد. هر چه حجم نمونه کوچک‌تر باشد ارتفاع قله این توزیع، نسبت به توزیع نرمال پایین‌تر است ولی با افزایش حجم نمونه ارتفاع قله به نرمال نزدیک‌تر می‌شود به طوری که برای حجم نمونه‌های بزرگ ($n \geq 50$) این فاصله خیلی کم می‌شود. توزیع t با مشخصه‌ای به نام درجه آزادی شناخته می‌شود که متناسب با حجم نمونه است و در ساده‌ترین حالت برابر تعداد نمونه منهای یک می‌شود (شکل ۳).

1. William Sealy Gosset
2. Guinness Berwery
3. Biometrika

مثال ۳: در مطالعه‌ای با هدف تعیین نرخ مرگ و میر ناشی از سرطان کولون [۱] در یک مدت زمان پیگیری معین با میانه ۲۱/۱، تعداد ۲۲۳ مورد مرگ مشاهده شد (از تعداد کل ۸۱۷ مورد) مشاهده شد که به طور متوسط مرگ و میر ماهانه را برابر تقریباً ۱۰/۵ مورد با انحراف معیار ۳/۲، نتیجه می‌دهد. بنابراین نحوه محاسبه فاصله اطمینان برای میزان فوق به صورت زیر است:

$$(n \div \text{تعداد رخداد پیشامد (ماهانه)}) = \sqrt{\text{انحراف معیار میزان}}$$

(ماهانه)

$$= \sqrt{(10.5 \div 817)} = 0.11$$

((انحراف معیار میزان) $\times 2 \pm$ میزان نمونه‌ای) = فاصله

اطمینان پارامتر میزان در جامعه

$$= (10.5 \pm 2 \times (0.11)) = (9.78 \text{ تا } 10.22)$$

بنابراین، با اطمینان ۹۵ درصد پارامتر میزان مرگ و میر ماهانه در این جامعه در فاصله فوق قرار گرفت.

خطای معیار اختلاف‌ها (تفاضل‌ها)

در برخی از مسایل هدف مقایسه دو گروه یا دو جامعه در پارامتر مشخصی نظیر میانگین، نسبت و یا میزان است که مسئله مورد نظر به اختلاف یا تفاضل بین این شاخص‌ها در دو جامعه می‌باشد. بنابراین، هدف محاسبه فاصله اطمینان برای این کمیت‌ها خواهد بود که در این صورت می‌توان فرمول بیان شده برای محاسبه فاصله اطمینان میانگین را به صورت زیر تعمیم داد:

((انحراف معیار اختلاف) $\times 2 \pm$ اختلاف در شاخص‌های

نمونه‌ای) = فاصله اطمینان اختلاف‌ها

برای محاسبه، کفایت اختلاف در شاخص‌های نمونه‌ای و انحراف معیار آن مشخص باشد (جدول ۴).

(فرآیندهای شمارشی) باشد، در این صورت می‌توان فرمول بیان شده برای محاسبه فاصله اطمینان میانگین را به صورت زیر تعمیم داد:

((انحراف معیار شاخص) $\times 2 \pm$ شاخص نمونه‌ای) =

فاصله اطمینان پارامتر

برای محاسبه کفایت شاخص نمونه‌ای ویژگی مورد نظر و انحراف معیار آن مشخص باشد. انحراف معیار نسبت و میزان به صورت زیر می‌باشد:

(نسبت) $\div (1 - \text{نسبت}) \times \sqrt{\text{انحراف معیار نسبت}}$

(نسبت) \div تعداد رخداد پیشامد) = انحراف معیار میزان

مثال ۲: در مطالعه‌ای با هدف تعیین شیوع مصرف سیگار در جوانان دانش‌آموز، شاغل و بیکار [۷]، از تعداد کل ۷۳۸ نفر، تعداد ۱۸۹ نفر سیگاری بودند که در این صورت برآورد عددی نسبت برابر ۰/۲۵۶ ($189 \div 738$) به دست می‌آید که با ضرب در عدد صد برابر ۲۵/۶ به دست می‌آید. نحوه محاسبه فاصله اطمینان برای نسبت فوق به صورت زیر است:

$$= \sqrt{((0.256) \times (1 - 0.256) \div 738)} = 0.003$$

((انحراف معیار نسبت) $\times 2 \pm$ نسبت نمونه‌ای) = فاصله

اطمینان پارامتر نسبت در جامعه

$$= (0.256 \pm 2 \times (0.003)) = (0.224 \text{ تا } 0.288)$$

و برای بیان این فاصله بر حسب درصد، کفایت حدود در عدد ۱۰۰ ضرب شوند که به صورت (۲۸/۸ تا ۲۲/۴) درصد به دست می‌آید. لازم به ذکر است که برای محاسبه فاصله اطمینان نیاز است از نسبت استفاده شود و پس از محاسبه، حدود در عدد ۱۰۰ ضرب می‌شوند که فاصله مبتنی بر درصد را نتیجه می‌دهند.

جدول ۴- فرمول خطای معیار برای شاخص‌های مختلف

نوع متغیر	توزیع احتمالی	پارامتر (جامعه)	نماد	شاخص نمونه‌ای	فرمول خطای معیار	فرمول ریاضی
کمی پیوسته	نرمال	اختلاف میانگین‌ها	$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$	جذر ((واریانس اول $\div n_1$) + (واریانس دوم $\div n_2$))	$\sqrt{[(S_1^2 \div n_1) + (S_2^2 \div n_2)]}$
کیفی دوحالتی	دوجمله‌ای	اختلاف نسبت‌ها	$\pi_1 - \pi_2$	$p_1 - p_2$	جذر ((نسبت اول \times یک منهای نسبت اول $\div n_1$) + (نسبت دوم \times یک منهای نسبت دوم $\div n_2$))	$\sqrt{\{[p_1 \times (1-p_1) \div n_1] + [p_2 \times (1-p_2) \div n_2]\}}$
کمی گسسته (شمارشی)	پواسن	اختلاف میزان‌ها	$\lambda_1 - \lambda_2$	$r_1 - r_2$	جذر ((میزان اول $\div n_1$) + (میزان دوم $\div n_2$))	$\sqrt{[(r_1 \div n_1) + (r_2 \div n_2)]}$

مثال ۴: در مطالعه‌ی بررسی کیفیت زندگی بیماران دیابتی، میانگین (انحراف معیار) امتیاز بعد روانی-اجتماعی در دو گروه افراد با سابقه کمتر مساوی ۶ سال و ۷ سال و بیشتر، به ترتیب برابر ۱۰/۹۶ (۴/۰) و ۱۲/۲۲ (۴/۲۵) و تعداد افراد در این دو گروه به ترتیب برابر ۹۳ و ۸۷ بود [۸]، بنابراین فاصله اطمینان اختلاف این دو میانگین (با هدف بررسی اختلاف امتیاز این بعد در دو گروه)، به صورت زیر محاسبه می‌شود:

$$\begin{aligned} & \text{((جذر ((n1 \div \text{واریانس اول}) + (n2 \div \text{میزان دوم}))) \times 2 \pm \\ & \text{اختلاف در میزان‌های نمونه‌ای) =} \\ & \text{((جذر ((\text{واریانس اول} \div 6) + (\text{واریانس دوم} \div 10/5))) \times 2 \pm 6/0) -} \\ & \text{10/5} \\ & \text{= (4/84 تا 4/16) = 0/17} \end{aligned}$$

بنابراین، با اطمینان ۹۵ درصد پارامتر اختلاف میزان‌ها در دو سرطان و در این جامعه در فاصله فوق قرار گرفت.

آزمون فرضیه

آزمون‌های معنی‌داری بخشی از آمار استنباطی هستند که در آن بر اساس نتایج حاصل نمونه در مورد جامعه‌ای که نمونه از آن استخراج شده است، تعمیمی صورت می‌گیرد. آزمون فرض به مطالعات سررنالد فیشر (قرن نوزدهم و بیستم)، جرزی نیمن (قرن نوزدهم) و کارل پیرسون (قرن نوزدهم و بیستم) برمی‌گردد. آزمون فرض مدرن ترکیبی از کارهای این افراد است که در قرن بیست عنوان آزمون فرض به آنها اختصاص داده شد [۹].

در بحث استنباط آماری، یعنی تجزیه و تحلیل داده‌ها و تعمیم نتایج آن به جامعه آماری مورد بررسی، تصمیم‌گیری در مورد ادعاهای مطرح شده نیز وجود دارد. به عبارت دیگر فرضیه‌های پژوهشی که در مطالعات مطرح می‌شوند باید توسط روش‌های مناسب (آزمون‌های فرض آماری) بررسی شوند و تأیید و یا رد گردند. در این مورد نیز نقش شانس در داده‌ها با استفاده از الگوهای احتمالی نمونه‌گیری مدل می‌شوند. در این بحث هر فرضیه و پژوهش (یا سوال پژوهشی مبتنی بر اختلاف یا رابطه) در مطالعات تحلیلی باید به زبان آماری (یعنی فرضیه آماری) تبدیل شود تا از طریق روش‌های آماری قابل بررسی باشد. آزمون فرض شامل دو بخش فرض یعنی همان ادعای مطرح شده و

بنابراین، با اطمینان ۹۵ درصد پارامتر اختلاف میانگین‌ها در دو گروه و در این جامعه در فاصله فوق قرار گرفت.

$$\begin{aligned} & \text{((جذر ((n1 \div \text{واریانس اول}) + (n2 \div \text{واریانس دوم}))) \times 2 \pm \\ & \text{اختلاف در میانگین‌های نمونه‌ای) =} \\ & \text{((جذر ((\text{واریانس اول} \div 10/96) + (\text{واریانس دوم} \div 12/22))) \times 2 \pm 4/0) -} \\ & \text{10/96} \\ & \text{= (12/22} \\ & \text{= 1/26 \pm 2 \times (0/62)} \\ & \text{= (0/05 تا 2/47)} \end{aligned}$$

بنابراین، با اطمینان ۹۵ درصد پارامتر اختلاف میانگین‌ها در دو گروه و در این جامعه در فاصله فوق قرار گرفت.

مثال ۵: در مطالعه بررسی شیوع کشیدن سیگار، تعداد دانش‌آموزان و جوانان شاغل سیگاری به ترتیب برابر ۱۱۶ و ۴۸ بودند و تعداد کل افراد در این دو گروه به ترتیب برابر ۵۲۱ و ۱۵۶ نفر بود که به ترتیب برابر ۲۲/۳ و ۳۰/۸ درصد را نتیجه داد [۷]. بنابراین فاصله اطمینان اختلاف این دو نسبت (با هدف بررسی اختلاف نسبت شیوع در دو گروه)، به صورت زیر محاسبه می‌شود:

$$\begin{aligned} & \text{((جذر ((n1 \div \text{یک منهای نسبت اول} \times \text{نسبت اول}) + (n2 \div \text{یک منهای نسبت دوم} \times \text{نسبت دوم}))) \times 2 \pm \\ & \text{اختلاف در نسبت‌های نمونه‌ای) =} \\ & \text{((جذر ((\text{نسبت اول} \div 10/96) + (\text{نسبت دوم} \div 12/22))) \times 2 \pm 4/0) -} \\ & \text{10/96} \\ & \text{= (12/22} \\ & \text{= 1/26 \pm 2 \times (0/62)} \\ & \text{= (0/05 تا 2/47)} \end{aligned}$$

که بر حسب درصد ۱ درصد تا ۱۷ درصد می‌شود. بنابراین، با اطمینان ۹۵ درصد پارامتر اختلاف نسبت‌ها در دو گروه و در این جامعه در فاصله فوق قرار گرفت.

مثال ۶: در مطالعه بررسی خطر مرگ ناشی از سرطان کولورکتال، تعداد مرگ ناشی از سرطان‌های کولون و رکتال به ترتیب برابر ۲۲۳ و ۱۲۱ مورد به ترتیب در مدت زمان

۳- میانگین Cholesterol برای داروی جدید در مقایسه با گروه شاهد اختلاف دارد.

همان‌طور که ملاحظه می‌شود فرضیه‌های پژوهشی حدس هوشمندانه‌ای است که در مورد نتایج احتمالی مطالعه در راستای اهداف جزئی آن و در نهایت هدف کلی آن ارائه می‌شود و نتایج حاصل از تحلیل این حدس را تأیید و یا آن را رد می‌کند. به عبارت دیگر، اگر شواهد مبنی بر تأیید حدس باشد، در این صورت فرضیه پژوهشی تأیید می‌گردد و در غیر این صورت رد می‌شود.

نکته قابل توجه در بیان فرضیه‌های پژوهشی آن است که فرضیه‌ها با زبان تخصصی زمینه‌ای نوشته می‌شوند، که مطالعه در آن زمینه انجام می‌شود. مثلاً فرضیه‌های پژوهشی فوق به زبان علوم پزشکی نوشته شده‌اند، اما برای تأیید و یا رد آنها نیاز است از روش‌های آماری استفاده شود. برای اینکه فرضیه‌های پژوهشی به یک زبان قابل فهم توسط روش‌های آماری تبدیل شوند، نیاز است به زبان آماری ترجمه گردند. به عبارت دیگر، فرضیه‌های پژوهشی باید به فرضیه‌های آماری تبدیل شوند. همان‌طور که اشاره شد، تصمیم‌گیری در مورد یک فرضیه و پژوهش به صورت دو حالتی یعنی تأیید و یا رد فرضیه می‌باشد، بنابراین فرضیه‌های آماری نیز دو حالتی بیان می‌شوند که تأیید یکی رد دیگری را در بر دارد. به علاوه در نگارش در زمینه‌های آماری، یکی از آنها حالت جاری و متداول و وضعیت موجود را نشان می‌دهد و فرضیه دیگر وضعیت تغییر در شرایط را نشان خواهد داد. فرضیه‌های آماری به شکل H_0 و H_1 (به ترتیب فرضیه صفر^۱ و فرضیه مقابل^۲ نوشته می‌شوند که در آن فرض صفر (پوچ) وضعیت جاری یا موجود (معمولاً عدم اختلاف یا رابطه) و فرض مقابل (فرض یک) وضعیت تغییر یافته (معمولاً وجود اختلاف یا رابطه) را نشان می‌دهد. برای فرمول کردن فرضیه پژوهشی به صورت فرض‌های آماری، همیشه این موضوع را باید رعایت کرد که مساوی در فرض صفر قرار می‌گیرد. به عنوان مثال، برای فرضیه پژوهش اول مطرح شده در مثال فوق فرضیه‌های آماری را می‌توان به صورت زیر نوشت:

آزمون یعنی ابزار بررسی این ادعاست که براساس نتیجه آزمون فرضیه پژوهشی تأیید یا رد می‌شود.

فرضیه پژوهشی و فرضیه‌های آماری

برای انجام مطالعات متداول آن است که یک طرح از پیش نوشته می‌شود و یا حتی زمانی که طرحی از پیش تعیین شده وجود نداشته باشد، یک هدف کلی در نظر گرفته می‌شود که پس از پایان مطالعه باید به آن هدف برسد. اما ممکن است هدف کلی عملیاتی نباشد و برای آن که بتوان در عمل و با انجام مطالعه بدان دست یافت، نیاز است هدف کلی به اجزای کوچک‌تری تقسیم شود که به آنها اهداف جزئی گفته می‌شود. اهداف جزئی در زمینه تخصصی مورد نظر و با آن زبان نوشته می‌شوند. منظور اینکه به عنوان مثال، اگر هدف بررسی تأثیر یک داروی جدید بر میزان چربی خون در مقایسه با داروهای روتین یا با دارونما باشد [۱۰]، هدف کلی را می‌توان به صورت «تأثیر داروی جدید بر پروفایل‌های لیپیدی» تعریف کرد و اهداف جزئی را به صورت زیر تنظیم نمود:

۱- تعیین میزان تأثیر داروی جدید روی HDL و مقایسه آن با گروه شاهد

۲- تعیین میزان تأثیر داروی جدید روی LDL و مقایسه آن با گروه شاهد

۳- تعیین میزان تأثیر داروی جدید روی Cholesterol و مقایسه آن با گروه شاهد

همان‌طور که ملاحظه می‌شود برای دستیابی به هدف کلی نیاز است که ابتدا به اهداف جزئی (عملیاتی شده) دست یافت. به علاوه اهداف در قالب ادبیات پزشکی نوشته شده است. حال برای دستیابی به اهداف جزئی معمولاً فرضیه‌های پژوهشی مبتنی بر آنها ساخته می‌شود. به عنوان مثال، برای اهداف جزئی فوق، فرضیه‌های پژوهشی زیر را می‌توان به صورت زیر تنظیم نمود:

۱- میانگین HDL برای داروی جدید در مقایسه با گروه شاهد اختلاف دارد.

۲- میانگین LDL برای داروی جدید در مقایسه با گروه شاهد اختلاف دارد.

1. Null Hypothesis
2. Alternative Hypothesis

موضوع باشد که داروی جدید (با عوارض جانبی کمتر) از لحاظ تاثیر بهبودی حداقل به اندازه داروی روتین (که عوارض جانبی بیشتری دارد) عمل می‌کند یا خیر، در این صورت در این صورت فرضیه پژوهشی برای مثال فوق به صورت: داروی جدید در مقایسه با داروی در اثرگذاری روی میانگین HDL حداقل بهتر است و این یعنی این که میانگین HDL در داروی جدید بزرگتر یا مساوی میانگین HDL در داروی روتین است. بنابراین فرضیه پژوهشی فوق، به صورت زیر به فرضیه‌های آماری تبدیل می‌شوند، که در آن فرضیه پژوهشی در فرضیه صفر آماری نوشته شده است:

$$H_0: \mu_1 \geq \mu_2$$

$$H_1: \mu_1 < \mu_2$$

برای تصمیم‌گیری در مورد فرضیه‌های آماری، همان‌طور که در بحث احتمال راجع آن بحث شد، ابتدا باید یک مدل یا الگوی احتمالی را برای اختلاف بین دو گروه باید در نظر گرفته شود، که در این مورد با فرض نرمال بودن توزیع داده‌ها، توزیع اختلاف بین دو میانگین، نیز نرمال خواهد بود. بنابراین با در نظر گرفتن توزیع نرمال برای اختلاف بین دو میانگین مورد انتظار تحت شانس برای این اختلاف برابر صفر است. به عبارت دیگر، انتظار می‌رود تحت H_0 میانگین اختلاف‌ها برابر صفر باشد، یعنی:

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{یا} \quad H_0: \text{Dif} = 0$$

زمانی از H_0 انحراف ایجاد شده است که میانگین مشاهده شده برای اختلاف بین میانگین‌ها از صفر فاصله بگیرد. شایان ذکر است از نظر ریاضی ممکن است مقادیر بسیار کوچک (مثلاً ۰/۰۰۱) هم به عنوان اختلاف در نظر گرفته شود ولی در مسایل استنباط آماری، اختلاف زمانی معنی‌دار^۱، در نظر گرفته می‌شود که از حد معینی بگذرد. برای حالتی که فرضیه بدون جهت بیان می‌شود، این انحراف می‌تواند در سمت چپ و یا در سمت راست عدد صفر در توزیع احتمال نرمال رخ دهد (شکل ۴):

میانگین HDL در گروه شاهد=میانگین HDL در گروه

داروی جدید H_0 :

میانگین HDL در گروه شاهد \neq میانگین HDL در گروه

داروی جدید H_1 :

و یا اگر میانگین HDL در گروه داروی جدید با μ_1 و میانگین H_1 در گروه شاهد با μ_2 نشان داده شود (که در آن μ_1 ، μ_2 یونانی است)، در این صورت فرضیه‌ها به صورت زیر خواهند بود:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

همان‌طور که ملاحظه می‌شود، مساوی در فرض صفر قرار گرفته است و هر چه که در H_0 قرار گرفته، خلاف آن در H_1 قرار می‌گیرد.

همان‌طور که در فرضیه‌های آماری و یا در فرضیه‌ی پژوهشی اول مشاهده می‌شود، مشخص نشده است که داروی جدید باعث افزایش HDL و یا کاهش آن در مقایسه با گروه شاهد می‌گردد. به عبارت دیگر جهت تاثیر گذاری آن مشخص نشده است. ممکن است این نحوه بیان فرضیه پژوهشی به این دلیل بوده است که پژوهشگر نسبت به جهت تاثیر گذاری داروی جدید، نظری نداشته است. حال اگر پژوهشگر می‌دانست که داروی جدید باعث افزایش HDL در مقایسه با شاهد می‌شود در این صورت فرضیه پژوهشی به صورت: میانگین HDL در داروی جدید در مقایسه با گروه شاهد افزایش می‌یابد (بزرگتر است)، نوشته می‌شود. به فرضیه فوق که در آن جهت تاثیر گذاری مداخله مشخص است، فرضیه پژوهشی جهت‌دار گفته می‌شود. در مقابل به فرضیه‌هایی که قبلاً مطرح شده بودند و جهت تاثیر گذاری مداخله در آن مشخص نبود، فرضیه پژوهشی بدون جهت گفته می‌شود. فرضیه پژوهشی فوق، به صورت زیر به فرضیه‌های آماری تبدیل می‌شوند، که در آن فرضیه پژوهشی در فرضیه مقابل آماری نوشته شده است:

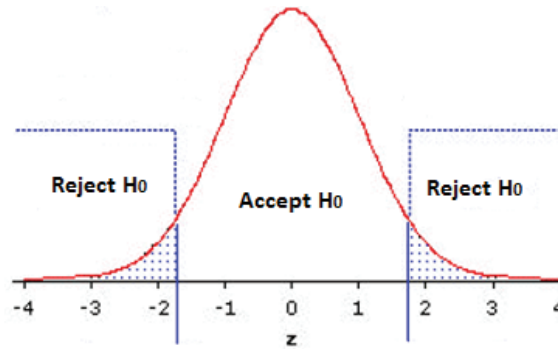
$$H_0: \mu_1 \leq \mu_2$$

$$H_1: \mu_1 > \mu_2$$

در اغلب مطالعات، فرضیه‌ی پژوهشی در فرض مقابل تعریف می‌شود ولی لزوماً این‌گونه نیست؛ مثلاً در یک مطالعه کارآزمایی بالینی ممکن است هدف بررسی این

1. Significant

تصمیم‌گیری درباره‌ی فرضیه‌ها مسئله دوباره از زمینه آماری به زمینه‌ی تخصصی مورد نظر بر می‌گردد. همان‌طور که قبلاً اشاره شد، زمانی که انحراف معیار جامعه معلوم نباشد (که در عمل تقریباً چنین است)، از انحراف معیار نمونه‌ای به جای آن استفاده می‌شود، که در این صورت توزیع احتمال اختلاف (به جای نرمال)، توزیع t خواهد بود.

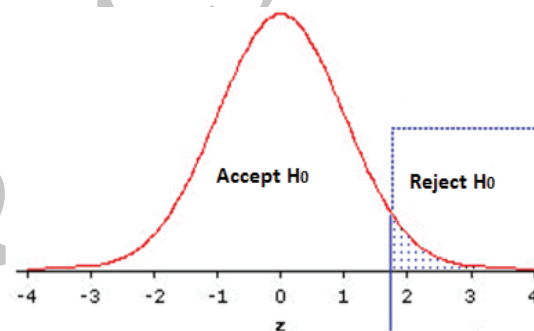


شکل ۴- ناحیه‌های رد فرض صفر در آزمون‌های دو دنباله‌ای

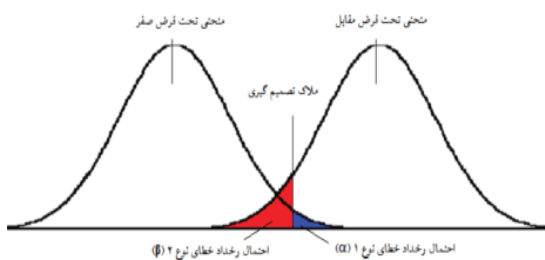
انواع خطاها در آزمون فرض

چون آزمون فرض یک ابزار تصمیم‌گیری براساس نمونه مشاهده شده در مورد جامعه است، یعنی معیاری است که از روی نمونه راجع به جامعه تصمیم می‌گیرد، بنابراین بدون خطا نمی‌باشد. به علاوه، با توجه به این که نتایج تصمیم‌گیری مبتنی بر آزمون فرض دو حالتی است یعنی تأیید یا رد یک فرضیه، بنابراین در رد یا تأیید H_0 (و یا H_1) ممکن است خطا رخ دهد به عبارت دیگر ممکن است H_0 به اشتباه رد شود و یا به اشتباه پذیرفته شود که به ترتیب خطای نوع اول و دوم نامیده می‌شوند و احتمال رخداد این دو نوع خطا به ترتیب با α (آلفا) و β (بتا) نشان داده می‌شوند (شکل ۶).

و یا زمانی که فرضیه به صورت جهت‌دار بیان می‌شود (فرضیه آماری دوم) در این صورت زمانی فرض صفر رد می‌شود که: $H_0: \mu_1 - \mu_2 > 0$ یا $H_0: Dif > 0$ یا به عبارت دیگر انحراف در سمت راست عدد صفر در توزیع احتمال نرمال رخ دهد (شکل ۵):



شکل ۵- ناحیه‌های رد فرض صفر در آزمون‌های یک دنباله‌ای



شکل ۶- احتمال رخداد خطای نوع ۱ (α) و خطای نوع ۲ (β) در آزمون فرض

به عبارت دقیق‌تر، ممکن است فرض صفر به اشتباه رد شود، به عبارت دیگر فرض صفر در جامعه‌ی آماری واقعاً صحیح باشد ولی بر اساس نمونه به اشتباه این فرضیه رد شود و یا ممکن است فرض صفر به اشتباه پذیرفته شود. به عبارت دیگر، فرض صفر در جامعه آماری واقعاً صحیح نباشد ولی براساس نمونه به اشتباه این فرضیه پذیرفته شود.

به آزمون‌های آماری معادل با فرضیه بدون جهت آزمون دو دنباله‌ای (دو دامنه) و به آزمون‌های معادل با فرضیه جهت‌دار آزمون یک دنباله‌ای (یک دامنه) گفته می‌شود. نکته:

- فرض صفر، در یک تصمیم دو حالتی، رد و یا پذیرفته می‌شود، در صورت رد فرض صفر، با شرایط نمونه فرض مقابل پذیرفته می‌شود و تأیید فرضیه‌ی صفر و یا فرضیه‌ی مقابل به معنی اثبات آنها نیست و تنها در شرایط نمونه حاضر این تأیید صورت گرفته است و این امکان وجود دارد که در نمونه دیگر نتیجه دیگری حاصل گردد.
- پس از رد یا تأیید فرض صفر، برای تفسیر آن باید به بیان فرضیه پژوهشی ملاحظه نمود. در واقع پس از

➤ توان آزمون با خطای نوع دوم رابطه‌ی معکوس و با خطای نوع اول رابطه‌ی مستقیم دارد.

➤ معمولاً آزمون‌های آماری با در نظر گرفتن مقدار ثابتی برای خطای نوع اول (در اغلب مطالعات ۰/۰۵) و با مینیمم خطای نوع دوم و یا ماکسیمم کردن توان آزمون ساخته می‌شوند.

➤ زمانی که در یک مطالعه اختلاف معنی‌دار حاصل نمی‌شود ممکن است به دلیل کم بودن حجم نمونه و در نتیجه توان پایین آزمون، اختلاف قابل آشکارسازی نباشد و یا این که واقعاً اختلاف قابل ملاحظه‌ای وجود نداشته است. در مورد اول مطالعه جای نقد دارد چون برای انجام آن حجم نمونه کافی در نظر گرفته نشده است و این نقد هم از دیدگاه اخلاقی و هم از دیدگاه داوران ارزیابی کننده طرح مطالعه کاملاً به جا و وارد است. چون در مرحله طراحی و شروع مطالعه و قبل از جمع‌آوری داده‌ها نیاز بود توان کافی برای تعیین حجم نمونه در نظر گرفته می‌شد و این موضوع همراه با جزئیات کامل باید در قسمت روش‌های گزارش قید شود. بنابراین برای بررسی معنی‌داری یک اختلاف یا رابطه، نیاز است توان آزمون هم در نظر گرفته شود، چون ممکن است عدم معنی‌داری به دلیل توان پایین آزمون باشد و توصیه می‌شود توان آماری آزمون هم به عنوان بخش از یافته‌ها در مطالعات گزارش شود [۲].

مقدار احتمال^۵ (P-value)

لازم به ذکر است که برای تصمیم‌گیری در مورد رد یا پذیرش فرض صفر، باید ملاک آماری تعیین شود که این ملاک همان P-value یا P است. در نرم‌افزارهای آماری (مثل SPSS)، برای تصمیم‌گیری در مورد رد یا پذیرش فرض صفر، این ملاک با نام احتمال معنی‌داری^۶ یا مقدار احتمال (P-value)، بر اساس داده‌ها محاسبه می‌شود. ملاک P-value، شاخصی است که از آزمون معنی‌داری نتیجه می‌شود و شدت سازگاری فرضیه مورد بررسی درباره جامعه را با شواهد (داده‌ها) ارزیابی می‌کند. فرضیه صفر

از طرف دیگر، تصمیم‌گیری‌های صحیح در مورد پذیرش یا رد H. نیز به ترتیب با پذیرش صحیح H. و رد H. بیان می‌شود که به ترتیب متمم احتمال رخداد خطای نوع یک و خطای نوع دوم هستند که به ترتیب برابر $1 - \alpha$ و $1 - \beta$ می‌شوند. به احتمال اول ضریب اطمینان (فاصله اطمینان) و به احتمال دوم توان آزمون (توان آشکارسازی اختلاف یا رابطه) گویند. آزمون مناسب آزمونی است که فرض صفر را به درستی رد کند. اصطلاحاً آزمونی که توان بیشتری برای رد فرض صفر داشته باشد، آزمون بهتری است. خلاصه‌ای از خطاها و تصمیم‌گیری‌های صحیح آزمون در جدول زیر آمده است (جدول ۵):

جدول ۵: خطاها و تصمیم‌گیری‌های صحیح در آزمون فرض

وضعیت H_0 در جامعه		درست	
		غلط	درست
α	$1 - \beta$	رد	تصمیم‌گیری براساس آزمون فرض
$1 - \alpha$	β	قبول	

همان‌طور که ملاحظه می‌شود، خطاها در آزمون‌های معادل تصمیم‌گیری‌های غلط آزمون تشخیصی است یعنی احتمال خطای نوع اول و دوم به ترتیب همان احتمال رخداد مثبت کاذب^۱ و احتمال رخداد منفی کاذب^۲ هستند. به علاوه توان آزمون و فاصله اطمینان به ترتیب معادل تصمیم‌گیری‌های صحیح در آزمون تشخیصی یعنی به ترتیب احتمال مثبت واقعی^۳ یا همان حساسیت آزمون و احتمال منفی واقعی^۴ یا همان ویژگی آزمون می‌باشند.

نکته:

- خطای نوع اول و دوم زمانی مطرح می‌شوند که به ترتیب فرض صفر و پذیرفته شود.
- در یک نمونه ثابت، خطای نوع اول و نوع دوم با یکدیگر رابطه معکوس دارند. به عبارت دیگر با افزایش یکی دیگری کاهش می‌یابد و بالعکس و با افزایش نمونه هر دو خطا کاهش می‌یابند و بالعکس.

1. False positive Rate
2. False Negative Rate
3. True Positive Rate
4. True Negative Rate

5. Probability Value

6. Significance Probability

مثال ۷: در حالت عادی برای یک سکه سالم زمانی که ۱۰ پرتاب انجام شود انتظار می‌رود که ۵ شیر مشاهده شود. منصفانه‌تر این است که برای شانس هم حدی قابل شد. فرض کنید در عمل در ۱۰ پرتاب سکه، ۸ شیر مشاهده شد. بنابراین در این مسئله، فرض صفر، سالم بودن سکه و یا وجود احتمال ۰/۵ برای رخداد شیر است و اگر احتمال مشاهده این تعداد شیر بر اساس فرض صفر سالم بودن سکه عددی کوچک (کمتر از ۰/۰۵) باشد در این صورت، مشاهدات با فرض صفر انطباق (سازگاری) ندارند و در نتیجه فرض صفر رد خواهد شد. بنابراین طبق تعریف، P-Value برابر با احتمال مشاهده ۸ شیر و بیشتر است ولی با توجه به این که ممکن بود با مشاهده تعداد کم شیر هم این نتیجه حاصل می‌شد بنابراین تعداد ۲ شیر و کمتر نیز برای این احتمال در نظر گرفته شده و در نتیجه، P-Value برابر با احتمال مشاهده ۸ شیر و بیشتر یا تعداد ۲ شیر و کمتر خواهد بود. زمانی که این احتمال از مقدار تعیین شده ۰/۰۵ کمتر شد، فرض صفر سالم بودن سکه رد می‌شود. به عبارت دیگر، در این صورت احتمال سالم بودن سکه کم است و می‌توان آن را به حساب شانس گذاشت و اگر این احتمال از ۰/۰۵ بزرگتر باشد در این صورت دیگر نمی‌توان آن را به حساب شانس گذاشت و در نتیجه فرض صفر سالم بودن سکه پذیرفته می‌شود. با توجه به این که توزیع مشاهده شیر در پرتاب سکه، یعنی توزیع دو جمله‌ای، متقارن است، بنابراین کافی است احتمال مشاهده ۸ شیر و بیشتر محاسبه و در عدد ۲ ضرب شود. احتمال مشاهده تعداد مختلف برای شیر به صورت زیر آورده شده است:

احتمال	تعداد شیرها
۰/۰۰۱	۰
۰/۰۱۰	۱
۰/۰۴۴	۲
۰/۱۱۷	۳
۰/۲۰۵	۴
۰/۲۴۶	۵
۰/۲۰۵	۶
۰/۱۱۷	۷
۰/۰۴۴	۸
۰/۰۱۰	۹
۰/۰۰۱	۱۰

آماري معمولاً مبتنی بر عدم وجود رابطه یا اختلاف در جامعه‌ای است که نمونه از آن استخراج شده است. آماره آزمون شاخصی است که تحت فرضیه صفر، از یک توزیع آماری معلوم تبعیت می‌کند. احتمال معنی‌داری تحت فرضیه صفر، احتمال این است که مقدار آماره آزمون تا چه حد از آنچه که بر اساس داده‌ها مورد انتظار است، انحراف دارد. اگر P-value کوچک باشد (معمولاً کمتر از ۰/۰۵)، شواهد برخلاف فرضیه صفر است و این فرضیه رد می‌شود. چون مشاهده شدن شانس‌ی یک نتیجه برابر ۰/۵ در نظر گرفته می‌شود، بنابراین برحسب مقایسه P با ۰/۵ تصمیم‌نهایی در مورد رد یا پذیرش فرضیه گرفته می‌شود؛ زمانی که P کمتر از ۰/۵ باشد در این صورت تعداد نتایج مشاهده شده از ۰/۵ شانس‌ی کمتر بوده و فرض مقابل شانس‌ی نیست و در نتیجه فرض صفر را باید رد کرد و برعکس. هر چه P کوچک‌تر باشد، شواهد با توان بیشتری فرض صفر را رد می‌کنند. اگر این احتمال بزرگ باشد، شواهدی ضعیف علیه فرضیه صفر وجود دارد و این فرضیه رد نمی‌شود. به عبارت دیگر، داده‌ها با فرضیه صفر سازگاری دارند و فرضیه وجود رابطه یا اختلاف معنی‌دار نیست. اگر اختلاف، معنی‌دار نباشد، فرضیه وجود اختلاف در جامعه رد می‌شود ولی ممکن است در واقع این اختلاف وجود داشته باشد و ممکن است نمونه آن قدر بزرگ نباشد که این اختلاف را آشکار نماید. به هر حال، این که اختلافی وجود ندارد، اثبات نمی‌گردد.

P-Value احتمال مشاهده یک اختلاف حداقل به اندازه آن چیزی است که در داده‌ها تحت صحیح بودن فرض صفر مشاهده می‌شود و در نتیجه میزان انطباق داده‌ها با این فرض را نشان می‌دهد و همان‌طور که گفته شد این اختلاف در ۵ (α) درصد مواقع قابل اغماض است یعنی ممکن است بر اساس شانس باشد بنابراین زمانی که P کوچکتر از ۵ درصد باشد میزان انطباق داده‌ها با فرض صفر وجود ندارد و در نتیجه، اختلاف یا رابطه واقعا وجود دارد ولی اگر تعداد موارد از ۵ درصد بیشتر شود در این صورت دیگر نمی‌توان آن را به حساب شانس گذاشت و بین داده‌ها با فرض صفر انطباق وجود دارد و در نتیجه، اختلاف یا رابطه وجود ندارد. برای روشن‌تر شدن موضوع، مثال زیر را ببینید.

شاخص آزمون، تعداد نمونه و توزیع احتمالی داده‌ها محاسبه می‌شود. در واقع α یک نقطه برش برای P-Value تعیین می‌کند به طوری که مقادیر کمتر از α برای P-Value، معنی‌دار بودن رابطه یا اختلاف را نشان می‌دهد.

۲. P-Value احتمال صحیح بودن فرض صفر یا غلط بودن فرض مقابل نیست. به عبارت دیگر از دیدگاه مکتب فراوانی نمی‌توان برای صحیح یا غلط بودن یک فرضیه احتمالی قائل شد و این موضوع را در مکتب بی‌زی می‌توان دنبال کرد.

۳. P-Value این احتمال نیست که تا چه مقدار یافته‌ها حاصل شانس هستند، و این موضوع متفاوت از آن است که P-Value شانس به دست آوردن چنین نتیجه‌ای در داده‌ها است اگر فرض صفر صحیح باشد.

۴. P-Value این احتمال را نتیجه نمی‌دهد که با تکرار مطالعه همان یافته‌ها حاصل خواهد شد به عبارت دیگر، P-Value تنها بر مبنای نتایج یک مطالعه محاسبه می‌شود و این احتمال تکرار پذیر نیست و تکرار پذیری نتایج را تضمین نمی‌کند.

۵. P-Value اندازه اثر (اندازه اختلاف یا رابطه) را نشان نمی‌دهد. اگر چه هر دو معیار با هم تغییر می‌کنند ولی P-Value علاوه بر اندازه اختلاف یا رابطه به حجم نمونه هم بستگی دارد، به طوری که برای موارد زیر تناقض وجود دارد: زمانی که اندازه اثر بزرگ باشد ولی حجم نمونه کوچک باشد ممکن است اختلاف یا رابطه مورد مطالعه معنی‌دار نشود و برعکس.

شایان ذکر است، برآورد نهایی یک آزمون معنی‌داری P-Value است و نه رد یا پذیرش فرض صفر، و در نتیجه ممکن است برای داده‌های حاصل از یک مطالعه واحد با تغییر مورد ادعا شده در فرضیه، مقادیر مختلفی برای P-Value به دست آید که البته این موضوع منطقی است چون فرضیات مختلفی بررسی می‌شوند و برای آنها P-Value های مختلفی به دست می‌آید. بنابراین نیاز است به عنوان احتمالی از سازگاری داده‌ها با فرض صفر و بر اساس مبنای نظری و میزان اعتقاد افراد خبره در زمینه و همچنین با دخالت دادن اندازه اثر بالینی تعبیر شود. همچنین برخی

همان‌طور که ملاحظه می‌شود، احتمال رخداد ۵ شیر یعنی همان مقدار مورد انتظار، دارای بیشترین احتمال است. احتمال مشاهده ۸ شیر و بیشتر یعنی مشاهده ۸، ۹ یا ۱۰ شیر برابر $0/055 (= 0/001 + 0/010 + 0/044)$ و در نتیجه احتمال مشاهده ۸ شیر و بیشتر یا ۲ شیر و کمتر برابر $0/110 (= 2 \times 0/055)$ می‌شود. چون P-Value بزرگتر از $0/05$ شده است در نتیجه فرض صفر رد نمی‌شود یعنی با داشتن سکه سالم حتی امکان مشاهده ۸ شیر و بیشتر یا ۲ شیر و کمتر (تا ۱۱ درصد موارد) وجود دارد.

اگر در عمل ۹ شیر مشاهده شود با روند فوق برای محاسبات، احتمال مشاهده ۹ شیر و بیشتر یا ۱ شیر و کمتر برابر $0/022 (= 2 \times 0/011)$ می‌شود. در این مورد چون P-Value کوچکتر از $0/05$ شده است در نتیجه فرض صفر رد می‌شود یعنی با داشتن سکه سالم حتی امکان مشاهده ۹ شیر و بیشتر یا ۱ شیر و کمتر تنها در ۲ درصد موارد وجود دارد که آن را می‌توان به حساب شانس گذاشت.

نکته:

برای گزارش معنی‌داری یا عدم معنی‌داری یک اختلاف یا رابطه در یافته‌ها، توصیه اکید می‌شود که مقدار P-Value تا سه رقم اعشار گزارش شود و از گزارش صرف معنی‌دار (Sig) یا غیر معنی‌دار^۱ (NS) و یا حتی $P < 0.05$ یا $P < 0.01$ خودداری شود. برای مواردی که خروجی نرم‌افزار مقادیر $0/000$ را ارائه می‌کند، شیوه ارائه $0/001 < P$ برای مناسب است.

اشتباهات متداولی که درباره P-Value وجود دارد عبارتند از:

۱. همواره بین α و P-Value اشتباه گرفته می‌شود، P-Value همان α نیست؛ یا سطح معنی‌داری^۲ مقدار ثابت از پیش تعیین شده‌ای است که توسط محقق، قبل از جمع‌آوری داده‌ها و بر اساس حساسیت موضوع، منابع و شرایط انتخاب می‌شود ولی P یا احتمال معنی‌داری متغیری است که پس از جمع‌آوری بر اساس

1. Non-Significant
2. Significance Level

که برای ارائه موثر یافته‌ها بسیار مفید است و به همین دلیل، در اغلب مجلات، ارائه مبتنی بر فاصله اطمینان توصیه می‌شود. به عبارت دیگر، چون نیاز است ابتدا پارامتر مورد بررسی در جامعه برآورد شود، اگرچه شاخص نمونه‌ای متناظر بهترین برآورد برای پارامتر مورد نظر است ولی لزوماً دقیقاً همان مقدار نخواهد بود. فاصله اطمینان دامنه‌ای از مقادیر را برای این منظور ارائه می‌کند و با احتمال مشخصی بیان می‌دارد که پارامتر جامعه می‌تواند هر یک از این مقادیر باشند.

۲. تصمیم‌گیری دوحالتی بر اساس P-Value موجب کاهش اطلاعات حاصل از نمونه به وجود یا عدم اختلاف یا رابطه می‌گردد و این که ممکن است در صورت وجود آنها، شدت آنها چقدر است اطلاعی فراهم نمی‌شود و یا اینکه مثلاً با این که بین دو عدد ولی منجر به دو تصمیم متفاوت می‌گردند. در مقایسه فاصله اطمینان میزانی از شدت اختلاف یا رابطه را و همچنین جهت آنها را هم نشان می‌دهد، به عبارت دیگر هر چه حدود فاصله اطمینان از عدد ادعا شده در فرض صفر فاصله بیشتری داشته باشند، شدت اختلاف یا رابطه بیشتر است و این که حدود اطمینان در سمت راست یا چپ عدد ادعا شده در فرض صفر باشند، جهت رابطه را نیز نشان می‌دهد.

۳. فاصله اطمینان دامنه‌ای از مقادیر احتمالی برای اندازه اثر مطالعه را فراهم می‌کند و از این جهت، منطبق بر داده‌هاست و حتی حدود فاصله اطمینان، بعد اندازه‌گیری معادل با داده‌ها را دارند.

۴. با استفاده از فاصله اطمینان می‌توان نه تنها اختلاف یا رابطه صفر را بررسی نمود، سایر مقادیر از پیش تعریف شده برای اختلاف یا رابطه را نیز بررسی کرد؛ مثلاً اگر محقق از پیش تعیین کرده باشد که داروی جدید اگر در مقایسه با داروی روتین بیش از ۲ واحد اختلاف در متغیر اصلی ایجاد کند در این صورت در فرض صفر مقدار اختلاف دیگر صفر نخواهد بود و مقدار اختلاف برابر ۲ خواهد بود که در این صورت اگر فاصله این مقدار را در بر داشته باشد، آنگاه فرض صفر پذیرفته

معتقدند که در قالب روش‌های بیزی، شواهد موجود تا به حال با نتایج حاصل از مطالعه حاضر ترکیب شود و نتیجه‌گیری نهایی مبتنی بر آن باشد.

فاصله اطمینان در مقایسه با P-Value و مزایای فاصله اطمینان

فاصله اطمینان که همان تصمیم‌گیری صحیح در مورد پذیرش فرض صفر است، برای تصمیم‌گیری در مورد پذیرش یا رد فرض صفر به کار می‌رود. مطابق تعریف فوق، اگر مقداری که در فرض صفر ادعا می‌شود در فاصله اطمینان پارامتر مورد نظر قرار گیرد در این صورت فرض صفر تایید و در غیر این صورت رد می‌شود. مقدار تحت H₀ وابسته به مسئله ممکن است متفاوت باشد؛ در اغلب آزمون‌هایی که اختلاف یا رابطه را بررسی می‌کنند مقدار تحت فرض صفر، عدد صفر است و در آزمون‌های که نسبت‌هایی نظیر نسبت شانس‌ها^۱ (OR) یا نسبت خطر^۲ (RR) را آزمون می‌کنند مقدار تحت فرض صفر برابر ۱ می‌باشد. بنابراین اگر مقدار تحت فرض صفر در فاصله قرار نداشته باشد اختلاف یا رابطه معنی‌دار خواهد بود.

مثلاً برای مثال‌های ۴ تا ۶، که اختلاف میانگین‌ها، نسبت‌ها و میزان‌ها را بررسی کردند، در هیچ یک از فواصل اطمینان عدد صفر (عدد تحت فرض صفر برای اختلاف) وجود نداشت (به ترتیب فواصل اطمینان شامل: (۲/۴۷ تا ۰/۰۵)، (۰/۱۷ تا ۰/۰۱) و (۴/۸۴ تا ۴/۱۶))، بنابراین فرض صفر برابری پارامترهای مورد بررسی در مثال‌های مزبور رد می‌شود و برای پارامترهای مورد بررسی در گروه‌های تعریف شده اختلاف معنی‌داری وجود داشت.

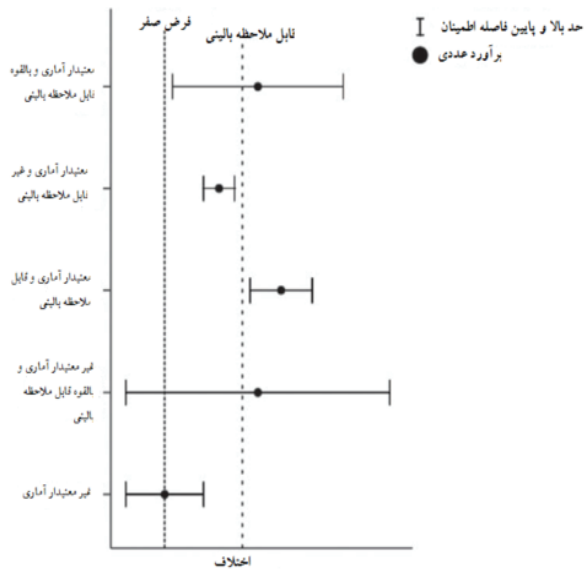
فاصله اطمینان علاوه بر تصمیم‌گیری در مورد رد یا پذیرش فرض صفر، نسبت به P-Value اطلاع بیشتری درباره رابطه یا اختلاف مورد نظر فراهم می‌کند و از این نظر نسبت به آزمون‌های معنی‌داری مزایایی دارد، از جمله:

۱. فاصله اطمینان، علاوه بر تصمیم‌گیری در مورد رد یا پذیرش فرض صفر، برآوردی از اثر (اندازه اثر مطلق) در مطالعه و میزانی از تغییرپذیری آن را فراهم می‌کند

¹ Odds Ratio

² Relative Risk (Risk ratio)

مطالعات با تعداد نمونه کوچک یک اثر بزرگ ممکن است به دلیل حساسیت پایین آزمون غیر معنی‌دار گزارش شود. در مقابل، فاصله اطمینان با ارائه محدوده احتمالی برای اثر مطالعه، اندازه‌ای از اهمیت بالینی (عملی) اثر مطالعه را فراهم می‌کند (شکل ۷).



شکل ۷- معنی‌داری آماری در مقابل قابل ملاحظه بودن عملی (کلینیکی)

معنی‌داری آماری باید در کنار اهمیت بالینی یک اختلاف یا رابطه بررسی شود و تصمیم‌گیری براساس وضعیت این دو کمیت صورت پذیرد (جدول ۶).

جدول ۶- تصمیم‌گیری در مورد وضعیت‌های مختلف معنی‌داری آماری و اهمیت بالینی

معنی‌داری آماری	اهمیت بالینی	توضیح	تصمیم‌گیری
وجود دارد	وجود دارد	--	اختلاف یا رابطه قابل توصیه است
وجود دارد	وجود ندارد	معمولاً برای حجم نمونه‌های بزرگ	اختلاف یا رابطه قابل توصیه نیست
وجود ندارد	وجود دارد	معمولاً برای حجم نمونه‌های کوچک	اختلاف یا رابطه قابل بحث است
وجود دارد	وجود دارد	--	اختلاف یا رابطه غیر قابل بحث و توصیه است

اختلاف یا رابطه، به آن عنوان نتیجه منفی^۱ داده می‌شود و منظور این است که اختلاف یا رابطه وجود نداشته است

شده و در غیر این صورت رد می‌شود. در نتیجه برای مقادیر مختلفی از اختلاف یا رابطه، فاصله اطمینان می‌توان ملاک تصمیم‌گیری و دامنه مقادیر قابل قبول را فراهم کند. در نتیجه فاصله اطمینان ناحیه پذیرش فرض صفر را برای بی‌نهایت آزمون معنی‌داری فراهم می‌کند.

۵. فاصله اطمینان با تعریف محدوده‌ای از مقادیر قابل قبول برای اندازه اثر مطالعه، تفسیر نتایج مطالعه را ساده‌تر و تصمیم‌گیری در مورد میزان تاثیر مورد مطالعه را به قضاوت مخاطبین قرار می‌دهد و در نتیجه انعطاف‌پذیری بیشتری فراهم می‌کند.

۶. طول فاصله اطمینان اندازه‌ای از توان آزمون را فراهم می‌کند؛ هر چه طول فاصله اطمینان کوتاه‌تر باشد، توان آزمون بالاتر است، ولی P-Value صرفاً معنی‌داری را بدون توجه به توان آزمون نشان می‌دهد.

۷. معنی‌دار نبودن اختلاف یا رابطه به معنای عدم اختلاف یا رابطه نیست و همچنین معنی‌دار بودن اختلاف یا رابطه به معنای وجود اختلاف یا رابطه نیست، اما با ارائه فاصله اطمینان از اندازه اثر مورد مطالعه، این تفسیر اشتباه رخ نخواهد داد.

۸. معنی‌داری آماری لزوماً به معنای اهمیت کلینیکی اختلاف یا رابطه نیست و این موضوع توسط P-Value قابل ارائه نیست، چون در مطالعات با تعداد نمونه زیاد یک اثر کوچک ممکن است به دلیل حساسیت زیاد آزمون به شدت معنی‌دار گزارش شود و بر عکس برای

تصمیم‌گیری صرف بر اساس معنی‌داری آماری این اشکال را به وجود می‌آورد که در صورت معنی‌دار نشدن یک

نرخ خانواده خطا را تعدیل (تصحیح) نمود تا نرخ خطای نهایی یا خانواده آزمون‌ها از ۰/۵٪ بیشتر نشود. برای این منظور، روش‌های مختلفی وجود دارد که از متداول‌ترین روش‌ها برای تعدیل نرخ خطا، روش بونفرونی است. در این روش نرخ خطای اولیه آنقدر از ۰/۵٪ کمتر در نظر گرفته می‌شود که در مجموع آزمون‌ها، نرخ خطا از ۰/۵٪ بیشتر نشود. براین منظور خطای اولیه برابر ۰/۵٪ تقسیم بر تعداد آزمون‌ها در نظر گرفته می‌شود؛ مثلاً برای انجام سه آزمون مقایسه به صورت دو به دو، با توجه به این که برای هر آزمون ۰/۵٪ برای خطا در نظر گرفته می‌شود بنابراین در هر آزمون ۹۵٪ اطمینان وجود دارد. با در مستقل در نظر گرفتن مقایسه‌ها از یکدیگر، اطمینان کلی بر اساس حاصل ضرب ۳ آزمون به دست می‌آید (بر اساس اصل استقلال [۴]) که برابر است با $0/86 (= 0/95 \times 0/95 \times 0/95)$ و در نتیجه α کلی حاصل تفاضل این عدد از ۱ برابر ۰/۱۴ به دست می‌آید که از ۰/۰۵ اولیه بزرگتر است و اصطلاحاً مثبت کاذب اضافی ایجاد شده است و در چنین موقعیتی ممکن معنی‌دار شدن نتایج واقعی نخواهد بود و نیاز به تصحیح نرخ خطا است. در روش بونفرونی، خطای اولیه یعنی ۰/۰۵ به تعداد مقایسه‌ها تقسیم می‌شود. در این مثال، ۰/۰۵ به ۳ تقسیم می‌شود (تقریباً ۰/۰۱۷)، و در نتیجه با در نظر گرفتن ۰/۰۱۷ برای خطای اولیه، با انجام سه مقایسه، نرخ خطای نهایی از ۰/۰۵ بیشتر نخواهد شد و به این ترتیب تصحیح نرخ خطا صورت گرفته است. در عمل برای آزمون‌های معنی‌داری، به جای تقسیم α کلی به تعداد مقایسه‌ها، برای راحتی کار مقدار P-Value در تعداد مقایسه‌ها ضرب و تصحیح نرخ خطا انجام می‌شود.

این موضوع از این جهت اهمیت دارد که در مقالات ممکن است تعداد زیادی از روابط جستجو شود (مثلاً انجام تحلیل‌های زیرگروهی و تحلیل‌های فرعی) که بالاخره از بین آنها نتایج معنی‌دار حاصل شود. همان‌طور که اشاره شد، معنی‌دار شدن نتایج واقعی نخواهد بود و نیاز به تصحیح نرخ خطا است و ممکن است با تصحیح نرخ خطا نتایج معنی‌دار نخواهند شد. چون مثلاً در یک تحلیل زیر گروهی با ۵ گروه اگر بر اساس نرخ تصحیح نشده خطا، یک نتیجه معنی‌دار حاصل شود (فرض کنید $P=0/020$)

(استفاده از این واژه توصیه نمی‌شود). در حالی که، نتیجه آزمون معنی‌داری آماری در چنین موردی بیان می‌دارد که شواهدی مبنی بر وجود اختلاف یا رابطه وجود نداشت. به عبارت دیگر، نبودن شواهد مبنی بر وجود اختلاف یا رابطه به معنای شواهد مبنی بر نبود آن نیست [۱۱]. همان‌طور که قبلاً اشاره شد ممکن است برخی از موارد غیر معنی‌دار به دلیل حجم نمونه کوچک و توان پایین آزمون باشد، و اگر ارائه یک خدمت بهداشتی یا ارائه یک درمان، با توجه به اهمیت آن از دیدگاه بهداشت عمومی وابسته به وجود شواهد باشد، اتکاء صرف معنی‌داری آماری ممکن است باعث تاخیر در ارائه آن خدمت گردد، بنابراین به استفاده از فاصله اطمینان و توجه به اهمیت بالینی در چنین مواردی توصیه اکید می‌شود. در پایان نیاز است این نکته مد نظر قرار گیرد که دو معیار P-Value و فاصله اطمینان مکمل هم هستند و با در دست داشتن یکی و اندکی اطلاعات دیگر، کمیت دیگر قابل محاسبه است. براساس P-Value یک تصمیم قطعی و سریع در مورد پذیرش یا رد فرض صفر فراهم می‌کند ولی در مقایسه فاصله اطمینان دامنه‌ای از مقادیر مورد پذیرش تحت فرض صفر را فراهم می‌کند. بنابراین P-Value و فاصله اطمینان نتایج متناقض ارائه نمی‌کنند و در نتیجه توصیه می‌گردد هر دو شاخص برای توصیف اختلاف یا رابطه در یک مطالعه ارائه شوند و در صورتی که انتخاب بین یکی از این دو کمیت است، با توجه به مزایای آن، استفاده از فاصله اطمینان پیشنهاد می‌شود.

چند آزمونی^۱

نکته دیگری که در مورد آزمون‌ها باید دقت نمود این است که اگر برای بررسی یک فرضیه پژوهشی نیاز به انجام چند آزمون باشند، در این صورت چون در هر آزمون ۰/۵٪ خطا در نظر گرفته می‌شود، مقدار خطا در چند آزمون از ۰/۵٪ بیشتر می‌شود که با قرارداد اولیه مبنی برداشتن ۰/۵٪ خطا هم‌خوانی ندارد. بنابراین نیاز است در آزمون‌های چندگانه برای یک فرضیه نرخ خانواده خطا را تصحیح نمود. به این پدیده چند آزمونی گفته می‌شود. در چنین مواردی باید

1. Multiple Testing

مشاهده‌ای) نیز رخ می‌دهد. شایان ذکر است روش بونفرونی به دلیل استفاده از قاعده ضرب تعداد مقایسه‌ها یا تعداد روابط در P-Value روشی بسیار محافظه‌کار است (حساسیت آن پایین است) و تمایل به غیر معنی‌دار شدن نتایج دارد. در این مورد توصیه می‌شود روش‌های دیگری نظیر میزان کشف اشتباه False Discovery Rate (FDR) [۱۲] استفاده شود.

ولی ممکن است با تصحیح نرخ خطا این نتیجه دیگر معنی‌دار نباشد ($P = 0.020 \times 5 = 0.100$) تصحیح شده). مسئله چند آزمونی در مسایل دیگری نظیر: بررسی اختلاف یا رابطه در چند پیامد با اولویت یکسان، مقایسه دو به دو گروه‌ها در مطالعات چندگروهی، مقایسه دو به دو زمان‌ها در اندازه‌گیری‌های تکراری، و همچنین بررسی رابطه میان تعداد زیادی متغیر (غالباً در مطالعات

مأخذ

1. Asghari-Jafarabadi M, Hajizadeh E, Kazemnejad A, Fatemi SR. Site-Specific Evaluation of Prognostic Factors in Iranian Colorectal Cancer Patients: A Competing Risks Survival Analysis. *Asian Pacific Journal of Cancer Prevention* 2009 10:815-22.
2. Godarzi M, Ebrahimzadeh E, Rabi A, SaeediPour 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.
3. Asghari Jafarabadi M, Soltani A, Mohammadi SM. Statistical Series: Summarizing and Displaying Data. *Journal of Diabetes and Lipid Disorders* 2013; Under Press [in Persian].
4. Asghari Jafarabadi M, Soltani A, Mohammadi SM. Statistical Series: Probability and Distributions. *Journal of Diabetes and Lipid Disorders* 2013; Under Review [in Persian].
5. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340:1-28.
6. Mankiewicz R. The Story of Mathematics. New Jersey: Princeton University Press 2004.
7. Asghari Jafarabadi M, Allahverdi-pour H, Bashirian S, Jannati A. Modeling the Underlying Predicting Factors of Tobacco Smoking among Adolescents. *Iranian Journal of Public Health* 2012; 41(5):46-57.
8. Mirfeizi M, Asghari Jafarabadi M, Mehdizadeh Toorzani, Mohammadi SM, Dehghan Azad M, Vizheh Mohammadi A, et al. Feasibility, Reliability and Validity of the Iranian Version of the Diabetes Quality of Life Brief Clinical Inventory (IDQOL-BCI). *Diabetes Research and Clinical Practice* 2012; 96(2):237-47.
9. Anders H. A History of Mathematical Statistics. New York: John Wiley and Sons 1998.
10. Ejtahed HS, Mohtadi-Nia J, Homayouni-Rad A, Niafar M, Asghari Jafarabadi M, Mofid V, et al. Effect of Probiotic Yoghurt Containing Lactobacillus acidophilus and Bifidobacterium lactis on Lipid Profile in Individuals with Type 2 Diabetes Mellitus. *Journal of Dairy Science* 2011; 94(7): 3288-94.
11. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311:485.
12. 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.