سری آمار: احتمال و توزیع‌های احتمالی

محمد اصغری جعفرآبادی[۱*]، سیده مومنه محمدی [۳]

چکیده

برای انجام استنباط آماری روی داده‌هایی که در قالب نمونه‌گیری‌های تصادفی انتخاب شده‌اند، احتمالات با ایفای نقش پل ارتباطی بین داده‌ها و آمار استنباطی، نقش تصادف یا شانس در داده‌ها را مدل می‌کند. هدف از این مقاله، معرفی احتمال و قواعد آن در قالب مثال‌های مطالعات علوم پزشکی و معرفی توزیع‌های احتمالی پرکاربرد در این مطالعات می‌باشد. تعاریف و مفاهیم اولیه احتمال و روش‌های محاسبه آن در مثال‌های ساده و کاربردی مشخص شد، و در ادامه توزیع‌های متداول احتمالی و ویژگی‌های هر یک از آنها معرفی گردید. با استفاده از قواعد خیلی ساده و در قالب مثال‌های ساده و کاربردی، احتمال در مطالعات پزشکی، محاسبه و براساس آن شاخص‌های بررسی رابطه در این مطالعات معرفی شد. همچنین برای مدل احتمالی متغیرهای کیفی دو حالتی، توزیع دوجمله‌ای و برای متغیرهای کمی گسسته و پیوسته به ترتیب توزیع پواسن و نرمال پیشنهاد گردید.

واژگان کلیدی: احتمال، تصادف، توزیع، نرمال، دوجمله‌ای، پواسن

[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: ۵۱۶۶۶۱۴۷۱۱، تلفن: ۲- ۰۴۱۱۳۳۵۷۵۸۰، نمابر: ۰۴۱۱۳۳۴۰۶۳۴، پست الکترونیک: asgharimo@tbzmed.ac.ir

تاریخ دریافت: ۱۳۹۱/۰۷/۱۵ **تاریخ درخواست اصلاح:** ۱۳۹۱/۰۷/۲۵ **تاریخ پذیرش:** ۱۳۹۱/۰۸/۱۸

مقدمه

در ابتدا ممکن است این سوال پیش آید که چرا باید احتمال بخوانیم؟ برای داشتن روایی بیرونی در مطالعات، یکی از شرط‌های اساسی، انتخاب تصادفی افراد حاضر در مطالعه یا در برخی دیگر از انواع مطالعات، تصادفی‌سازی افراد است [۱]. بنابراین در انتخاب نمونه‌ها و اندازه‌گیری‌های حاصل، شانس یا تصادف دخالت دارد. بنابراین نیاز است نقش شانس یا تصادف موجود در داده‌ها به نحوی در محاسبات لحاظ شود، که توزیع‌های احتمالی به عنوان یک پل ارتباطی، رابطه بین داده‌ها و استنباط آماری را برقرار می‌کنند. به عبارت دیگر به این دلیل که در مبحث استنباط آماری، داده‌ها بر اساس یک نمونه تصادفی به دست می‌آیند، بنابراین نقش شانس در انجام استنباط توسط توزیع‌های نمونه‌گیری فوق اصطلاحاً مدل می‌شوند. در نتیجه نقش الگوهای احتمال، مدل نمودن نقش شانس در استنباط آماری است.

در کتاب "حکایت دولت و فرزانگی" (نوشته مارک فیشر) [۲]، حکایت جوانی نقل می‌شود که به علت مشکلات مالی به عموی ثروتمند خویش مراجعه می‌کند. و عموی متمول به جای دادن ماهی راه ماهی‌گیری را به او یاد می‌دهد و برای این منظور او را به‌یک دولتمند معرفی می‌کند که بتواند از او " راز دولت و فرزانگی" را یاد بگیرد. ابتدای کار برای این که برای جوان انگیزه لازم ایجاد کند، چند ترفند را به کار می‌برد. در یکی از این ترفندها دولتمند به جوان می‌گوید که می‌تواند یک سکه را که به هوا پرت می‌کند و وقتی به زمین برمی‌گردد و شیر یا خط بودن آن تا ۱۰ بار دقیقاً مطابق با آنچه باشد که او می‌خواهد و ادعا می‌کرد که در این کار استعداد قوی ندارد، چون پدر او این کار را تا ۱۵ بار انجام می‌داد. حال چنین فضایی را تصور کنید و بگویید آیا ادعای دولتمند را به راحتی قبول می‌کنید؟ فرد خردمند حداقل برای تایید و یا رد ادعا، آزمایشی را انجام می‌دهد. به عبارت دیگر به دولتمند می‌گوید که سکه را برای ۱۰ بار پرتاب کند و نتیجه را بررسی نماید. در عمل چند بار نتیجه درست باشد، یعنی همان باشد که دولتمند می‌گوید، در این صورت ادعای دولتمند را می‌پذیرید یا حداقل اذعان می‌کنید که این کار او شگفت انگیز بوده است؟ ۵ بار یا ۶ بار؟ خیر زیرا ۵ یا ۶ بار ممکن است برحسب شانس نیز رخ دهد و در این صورت کار بزرگی انجام نداده است. ممکن است بگویید ۱۰ بار؟ خوب بله اگر این گونه شود که به توانایی بدون شک اذعان خواهید نمود ولی این رخداد خیلی ایده آل است. حال بیایید یک راه حل بینابینی در نظر بگیریم و این گونه در نظر بگیریم که اگر بیش از ۸ بار نتیجه را درست بگوید ادعای او را بپذیریم. در این صورت یک قضاوت منصفانه خواهیم نمود. به عبارت دیگر، اگر از ۱۰ بار و یا بیش از ۸ بار یعنی ۸٪ و بیش از ۸۰٪ موارد صحیح بگوید در این صورت ادعای او را قبول می‌کنیم و در غیر این صورت ادعای او رد می‌شود. حال بیاییم کل فرآیند را به صورت مفصل و جزئی بررسی کنیم:

یک سکه سالم دو وجه دارد، شیر و خط که اصطلاحاً به آمدن هر وجه سکه یک "برآمد" یا "پیامد" (outcome) گفته می‌شود. هر دو حالت شیر و خط که در یک سکه امکان رخداد آن وجود دارد، فضای نمونه گفته می‌شود. به عبارت دیگر تمام حالات ممکن رخداد پیشامدها را فضای نمونه نامند. حال اگر قبل از پرتاب سکه از شما پرسیده شود که نتیجه پرتاب چه خواهد بود، شما صرفاً حدس خود را بیان می‌کنید مثلاً می‌گویید حدس می‌زنم نتیجه شیر باشد ولی دقیقاً نمی‌توانید بگویید که نتیجه چه خواهد بود. بنابراین در هر بار پرتاب سکه، شما صرفاً حدسی از نتیجه را خواهید گفت. به عبارت دیگر، در هر بار اجرای این کار نتیجه برحسب شانس رخ می‌دهد، که این موضوع در بحث احتمال بسیار حائز اهمیت است. به عبارت دیگر در این بحث، زمانی می‌توان احتمال رخداد یک پیشامد را حساب کرد که آن پیشامد در هر بار برحسب شانس یا تصادف رخ دهد و محاسبه احتمال برای پیشامدهایی که قطعی هستند معنی ندارد. برای روشنتر شدن موضوع، اگر شما شیمی ندانید، برای اولین بار در آزمایشگاه، و فقط برای اولین بار، نمی‌توانید بگویید که واکنش O با H2 (یعنی واکنش اکسیژن با دو واحد هیدروژن) آب را به وجود می آورد ولی فقط در اولین بار نتیجه را دقیقاً نمی‌دانید ولی در بارهای بعدی نتیجه دقیقاً تکرار خواهد شد. بنابراین یک نتیجه دقیق و قطعی برای این آزمایش وجود دارد که

محاسبه احتمال در این مورد بی مفهوم خواهد بود. اصطلاحاً به این پدیده‌ها دقیق، قطعی یا ریاضی (Deterministic) گفته می‌شود. ولی در مورد سکه، در هر بار تکرار نتیجه شانسی است و به این پدیده‌ها اصطلاحاً تصادفی (Stochastic) گفته می‌شود. احتمال صرفاً در مورد پدیده‌های تصادفی بحث می‌کند و زمانی می‌توان احتمال را محاسبه کرد که‌یک پیشامد شانسی باشد یعنی نتوان از قبل نتیجه آن را دقیقاً گفت.

حال فرض کنید که سکه سالم باشد و شانس رخداد شیر یا خط بیشتر از دیگری نباشد. در این صورت شیر و خط شانس یکسانی برای رخ دادن دارند. بنابراین برای تمام حالت‌های ممکن که امکان رخداد آن ۱۰۰٪ است، نصف آن را به شیر و نصف آن را به خط اختصاص می‌دهیم (به دلیل سالم بودن سکه)، یعنی آمدن شیر و آمدن خط هرکدام با شانس ۵۰٪ رخ می‌دهد. به این نحوه محاسبه احتمال، محاسبه احتمال مبتنی بر **تعبیر هم شانسی** پیشامدها گفته می‌شود. به عبارت دیگر شیر و خط هم شانس بودند و در نتیجه امکان ۱۰۰٪ به طور مساوی بین آن‌ها (هرکدام ۵٪) تقسیم شد. در نمونه‌گیری‌هایی که با عنوان تصادفی ساده انجام می‌شود نیز این موضوع مشاهده می‌گردد. مثلاً اگر بخواهیم از بین ۱۰ بیمار یک نفر را به تصادف انتخاب کنیم آن گاه اگر هیچ گونه سوگرایی در انتخاب افراد وجود نداشته باشد، در این صورت هر کدام از ۱۰ بیمار هم شانس بوده و احتمال انتخاب هر کدام از آن‌ها می‌شود ۱ به ۱۰ یعنی ۰/۱ یا ۱۰ درصد.

نگاه دومی که در محاسبه احتمال وجود دارد، این است که فراوانی نسبی رخداد پیشامد مورد نظر محاسبه گردد. مثلاً اگر در ۱۰ پرتاب یک سکه ۸ بار شیر آمده باشد در این صورت شانس ۸ به ۱۰ یا فراوانی نسبی ۸ از ۱۰ بار یعنی ۸۰٪ شانس رخداد شیر را نتیجه می‌دهد به عبارت دیگر تعداد رخداد مورد نظر (شیر) به تعداد کل پرتاب‌های انجام شده احتمال رخداد شیر را در ۱۰ پرتاب نتیجه داده است. به این نحوه محاسبه احتمال، محاسبه مبتنی بر **تعبیر فراوانی نسبی** از احتمال گفته می‌شود.

در مقایسه دو تعبیر فوق، مورد دوم کاربردی‌تر است و در عمل بیشتر رخ می‌دهد. مثلاً شیوع یک بیماری (به عنوان احتمال بیمار بودن فردی از جامعه)، بر اساس تعداد موارد مشاهده شده از یک بیماری به تعداد کل افراد موجود در جمعیت به دست می‌آید که در واقع فراوانی نسبی محاسبه می‌شود. یا به طور مشابه در محاسبه میزان بروز، تعداد موارد جدید یک بیماری نسبت به تعداد کل افراد در معرض خطر در یک دوره زمانی معین در نظر گرفته می‌شود که در واقع فراوانی نسبی موارد جدید یک بیماری در افراد در معرض خطر محاسبه می‌شود و هر یک از موارد فوق احتمال مبتنی بر فراوانی نسبی هستند.

حال اگر از پزشکی جراح پرسیده شود که احتمال زنده ماندن فردی که قرار است عمل جراحی شود چقدر است (که معمولاً آن را تحت عنوان Risk و در قالب درصد بیان می‌کنند) فکر می‌کنید از کدام یک از تعبیرهای فوق استفاده می‌شود؟ بگذارید بیشتر توضیح دهیم آیا پزشک جراح با خود این فکر را می‌کند که:

۱) این عمل دو پیامد دارد مرگ و زندگی و بر اساس تعبیر هم شانسی هر کدام با شانس۵۰٪ رخ خواهند داد.

۲) مثلاً در ۱۰۰ مورد عملی که انجام داده است تعداد ۷۰ مورد زنده مانده اند و براساس تعبیر فراوانی نسبی شانس فرد تحت عمل ۷۰٪ است.

تعبیر اول به این دلیل اشکال دارد که نمی‌توان شانس مرگ و زندگی را در عمل جراحی یکسان گرفت و بنابراین براساس تعریف هم شانسی نمی‌توان برآورد ارائه شده پزشک را برای احتمال زنده ماندن توجیه نمود.

تعبیر دوم به این دلیل اشکال دارد که نیاز است فرض شود همه ۱۰۰ عمل شرایط کاملاً یکسانی داشتند تا بتوان آن‌ها را روی هم گذاشت و به یک نتیجه کلی رسید. اما می‌دانیم که ممکن نیست حتی دو فرد کاملاً یکسان از لحاظ ویژگی‌های زمینه‌ای یکسان باشند، باز هم ممکن است متغیرهای پنهان دیگری نظیر عوامل ژنتیکی وجود داشته باشد که باز هم شرایط آن‌ها را در عمل جراحی و نتایج پس از آن، متفاوت نماید و این تازه زمانی است که فرض کنیم مهارت و شرایط پزشک و تیم عمل جراحی، ابزار و ... تغییر نکنند. بنابراین در این مورد پزشک یک قضاوت شخصی از احتمال زنده ماندن را ارائه می‌کند که برحسب

تجارب شخصی پزشک و البته مشاهدات گذشته نتیجه شده است. به این تعبیر از احتمال، اصطلاحاً تعبیر **قضاوت شخصی** گفته می‌شود. این امکان وجود دارد که نظر پزشکان مختلف در مورد احتمال زنده ماندن فرد یکسان، کاملاً یکی نباشد و این احتمال از پزشکی به پزشک دیگر تغییر می‌کند. به عبارت دیگر کاملاً ذهنی (subjective) است. دلیل آن است که محاسبه نمی‌شود تا شواهد یکسانی در دست همه پزشکان باشد و نتایج یکسانی حاصل گردد.
در حالت کلی چهار نوع مختلف از تعبیر و ارائه برای احتمال وجود دارد که سه مورد تعابیر فوق هستند و مورد چهارم احتمالی است که بر اساس مدل‌های آماری به دست می‌آیند و به آنها **احتمال‌های مبتنی بر مدل** گفته می‌شود.
همان طور قبلاً اشاره شد، شکل دوم، شکل کاربردی آن است و بر اساس آن قاعده تعداد موارد مورد نظر به تعداد کل (فراوانی نسبی پیشامد مورد نظر) استخراج می‌شود. که یک قاعده بسیار کاربردی در محاسبات احتمال است. تعبیر اول بیشتر جنبه نظری دارد و برای درک حالات خاصی از احتمال مفید است. تعبیر سوم منطق و رویکرد جدیدی در علم آمار را به وجود آورده است که پایه‌گذار آن توماس بیز، کشیش انگلیسی (قرن هجدهم)، بوده است و به همین دلیل نام این رویکرد، بیزی می‌باشد. در این رویکرد ترکیبی از شواهد به دست آمده و قضاوت و تجربه شخصی متخصصین فن، در محاسبه احتمال دخالت می‌کنند که این موضوع طرفداران خاص خود را در مقابل رویکرد صرفاً مبتنی بر داده‌ها (یعنی رویکرد فراوانی) دارند. در ادامه، در قالب مثال‌هایی، نحوه محاسبه احتمال در حالت‌هایی بسیار ساده و کاربردی مشخص خواهد شد.

مثال ۱: در یک مطالعه طولی، برای بررسی رابطه بین مرحله پاتولوژیک تومور و خطر مرگ ناشی از سرطان کولون، دو گروه از افراد، شامل افراد با مرحله پیشرفته و اولیه مرحله تومور حضور داشتند (که در این مثال به ترتیب گروه مواجهه و گروه عدم مواجهه نام‌گذاری می‌شوند). افراد در یک مدت زمان پیگیری معین با میانه ۲۱/۱ ماهه، بررسی شده بودند و پس از اتمام دوره مطالعه خلاصه نتایج در جدول زیر ارائه شده است:

جدول ۱- خطر مرگ افراد مبتلا به سرطان کولون برحسب مرحله پاتولوژیک تومور

مرحله پاتولوژیک تومور	مرگ و میر ناشی از سرطان کولون	
	تعداد مرگ‌ومیر	تعداد کل
اولیه	۸۹	۴۳۱
پیشرفته	۱۴۹	۴۳۱
تعداد کل	۲۳۸	۸۶۲

منبع: [۴،۳]
نوع مطالعه: طولی، ارائه بر حسب تعداد، تعداد نمونه = ۱۲۱۹

در این مثال چند مورد احتمال محاسبه و در هر مورد توضیح لازم ارائه خواهد شد.
۱- احتمال مرگ ناشی از سرطان کولون در حالت کلی (در مجموع دو گروه). لازم به توضیح است که معمولاً محاسبه احتمال در حالت کلی مورد نظر نیست، زیرا در این صورت اثر مواجهه نادیده گرفته می‌شود و نمی‌توان رابطه مواجهه (مرحله پاتولوژیک تومور) و پیامد (مرگ) را به دست آورد. اما برای یادگیری محاسبه احتمال، در یک حالت ساده این کار را انجام می‌شود. همان‌طور که اشاره شد، در محاسبه احتمال از قاعده فراوانی نسبی یعنی "تعداد مورد نظر به تعداد کل" استفاده خواهد شد. بنابراین:

۲۳۸=۱۴۹+۸۹= تعداد کل افراد مبتلا به سرطان کولون که مرگ را تجربه کردند

۸۶۲=۴۳۱+۴۳۱= تعداد کل افراد تحت مطالعه (در معرض)

بنابراین احتمال بیماری برابر است با ۲۳۸ تقسیم بر ۸۶۲، که می‌شود ۰/۲۷۶ یا برحسب درصد می‌شود ۲۷/۶ درصد. به عبارت دیگر، احتمال مرگ ناشی از سرطان کولون در افراد، ۲۷/۶ درصد بوده است. این نحوه محاسبه احتمال دقیقاً مشابه محاسبه فراوانی نسبی برای خلاصه نمودن یک متغیر کیفی است که در [۳] ارائه شد.
۲- احتمال زنده ماندن (بقا) افراد مبتلا به سرطان کولورکتال در حالت کلی:
این احتمال را می‌توان به دو صورت محاسبه نمود؛ در روش اول از همان روش قبلی استفاده می‌شود یعنی:

۶۴۲=۲۸۴+۳۴۲= تعداد افراد مبتلا به سرطان کولون که مرگ را تجربه نکردند

۸۶۲=۴۳۱+۴۳۱= تعداد کل افراد تحت مطالعه (در معرض)
بنابراین احتمال سالم ماندن برابر است بـا ۶۴۲ تقـسیم بـر ۸۶۲ که می‌شود ۰/۷۲۴ یا برحـسب درصـد، ۷۲/۴ درصـد می‌شود. به عبارت دیگر احتمال زنده ماندن (بقـای) افـراد مبتلا به سرطان کولورکتـال، ۷۲/۴ درصـد بـوده اسـت. در روش دیگر می‌توان این گونه استنباط کرد که فضای نمونه، یعنی تمام حالات پیامد در ایـن مطالعـه، دو حالـت اسـت مرگ یا بقا. بنابراین کافیست از ۱۰۰٪ احتمـال کـل، سـهم مرگ ناشی از سرطان کولون یعنی ۲۷/۶ درصد را کم کنیم که می‌شود همان ۷۲/۴ درصد. این استنباط قاعـده دیگـری را در محاسبه احتمال با عنوان **قاعده متمم** احتمال را نتیجه می‌دهد که بیان می‌کند:

احتمال حالت دیگر –۱۰۰٪= احتمال حالت متمم

برای بررسی رابطـه مرحلـه پاتولوژیـک و مـرگ ناشـی از سـرطان کولـون، احتمـالات زیـر محاسـبه مـی‌شـوند کـه احتمالات شرطی نامیده می‌شوند:

۳- احتمال مرگ ناشی از سرطان کولـون در گـروه مرحلـه پیشرفته تومور (مواجهه):

در این حالت تفاوتی که به وجود آمده اسـت شـرطی روی محاسبه احتمال آورده شده است و آن شرط این اسـت کـه محدود به گروه مواجهه شده است و در نتیجه کافیست کـه محاسبات احتمال با همان قواعـد قبلـی محـدود بـه گـروه مواجهه شود یعنی صرفاً سطر اول جدول در محاسـبات در نظر گرفته شود:

۱۴۹= تعداد افراد مرده در گروه مرحله پیشرفته تومور

۴۳۱= تعداد کل افراد در مرحله پیشرفته تومور

که در این صورت احتمال مرگ ناشی از سرطان کولـون در گروه مواجهه می‌شـود ۱۴۹ تقـسیم بـر ۴۳۱ کـه مـی‌شـود ۰/۳۴۶ و برحسب درصد، ۳۴/۶ درصد می‌شود. به عبـارت دیگـر، احتمـال مـرگ ناشـی از سـرطان کولـون در گـروه مواجهه برابر ۳۴/۶ درصد می‌شـود. در ایـن وضـعیت تنهـا کاری که انجام شد این بود که فضای شرط شناسایی شد و محاسبات محدود به فضای شـرط گردیـد. بنـابراین بـرای محاسبه احتمـال شـرطی کافیـست محاسـبات محـدود بـه فضای شرط گردد.

۴- احتمال مرگ ناشی از سرطان کولـون در گـروه مرحلـه اولیه تومور (عدم مواجهه):

در این حالت نیز محاسبات به فضای شـرط عـدم مواجهـه یعنی سطر دوم جدول محدود می‌گردد:

۸۹=تعداد مرگ ناشی از سـرطان کولـون در گـروه مرحلـه اولیه تومور

۴۳۱= تعداد کل افراد در گروه مرحله اولیه تومور

بنابراین احتمال مرگ ناشی از سرطان کولون در گروه عـدم مواجهه برابر است با ۸۹ تقسیم بر ۴۳۱ که می‌شـود ۰/۲۰۷ و برحسب درصد، ۲۰/۷ درصد می‌شـود. بنـابراین احتمـال مرگ ناشی از سرطان کولون در گروه عـدم مواجهـه، برابـر ۲۰/۷ درصد می‌شود که باز هم یک احتمال شرطی است.

برای نتیجه‌گیری از دو احتمال اخیر، لازم به توضیح اسـت که،

۱- اولاً: هر یـک از احتمـال‌هـای فـوق یـک میـزان بـروز هستند، چون تعداد موارد جدید مرگ ناشی از سـرطان کولون در افراد در معرض خطر را نشان می‌دهند که در یک دوره زمانی معین بررسی شده‌اند.

۲- دوماً: زمانی کـه بتـوان گفـت احتمـال مـرگ ناشـی از سرطان کولون در گروه مواجهه بیشتر از این احتمال در گروه عدم مواجهه باشـد، در ایـن صـورت بیمـاری بـا مواجهه (مرحله پاتولوژیک تومور) رابطه داشـته اسـت (و چون مطالعه طولی است) می‌توان نتیجه گرفـت کـه مرحله پاتولوژیک تومور یکی از علت‌های مرگ ناشـی از سرطان کولون بوده است.

حال برای این که این بیان شکل کمی به خود گیرد می‌توان این گونه بیان کرد که:

احتمـال بیمـار شـدن در گـروه مرحلـه پیـشرفته تومـور (مواجهه) نسبت به احتمال بیمار شدن در گروه مرحله اولیه تومور (عدم مواجهه)

بزرگتر از یک باشد، آنگاه می‌توان گفت که مواجهه و مرگ ناشی از سرطان کولون رابطه دارند. در مثـال حاضـر، ایـن کمیت برابر است با ۱/۶۷ (۳۴/۶ تقسیم بـر ۲۰/۷ درصـد). بنابراین احتمال مـرگ ناشـی از سـرطان کولـون در گـروه مرحلـه پیـشرفته تومـور (مواجهـه)، ۶۷ درصـد بیـشتر از احتمال مرگ ناشی از سرطان کولون در گروه مرحله اولیـه

تومور (عدم مواجهه) است و در نتیجه، احتمالاً بین مرحله پاتولوژیک تومور و مرگ ناشی از سرطان کولون رابطه وجود داشته است. این شاخص همان خطر نسبی (Relative Risk (RR)) است.

لازم به ذکر است، در حالت کلی اگر RR برابر ۱ شود بیماری با مواجهه رابطه نخواهد داشت و اگر RR کوچکتر از یک شود مواجهه از بیماری جلوگیری می‌کند یعنی اثر محافظتی (Protective) دارد و زمانی که RR بزرگتر از یک شود مواجهه سبب بیماری شده است. در عمل مقدار RR به تنهایی برای تصمیم گیری کافی نیست و فاصله اطمینان ۹۵ درصدی آن محاسبه می‌شود و زمانی که عدد یک یعنی عدد متناظر با عدم رابطه در این فاصله قرار گیرد، نتیجه این می‌شود که بیماری و مواجهه رابطه ندارند.

یکی از نکات قابل توجه در محاسبه احتمال بیماری در گروه‌های مواجهه و عدم مواجهه این است که مرگ ناشی از سرطان کولون یعنی پیامد مورد نظر در افراد به تصادف رخ می‌دهد، یعنی قبل از شروع مطالعه نمی‌توان برای هر یک از افراد مبتلا به سرطان کولون حدس زد که مرگ برای آن‌ها رخ می‌دهد یا خیر، بنابراین با پیامدهای تصادفی سر وکار داریم و همان‌طور که در ابتدای این مقاله آورده شد، برای چنین پیامدهایی امکان محاسبه احتمال وجود دارد و در نهایت براساس احتمال‌های محاسبه شده شاخص RR محاسبه گردید. نقطه مقابل این بحث در مطالعات مورد-شاهدی است که پیامد در آن به تصادف رخ نمی‌دهد (و محقق پیامد را انتخاب می‌نماید و مطالعه از نقطه پیامد شروع می‌شود)، و به همین دلیل در آنها، امکان محاسبه احتمال و در نتیجه RR وجود ندارد. برای روشن شدن بیشتر موضوع آن را در قالب مثال زیر بیان می‌شود.

مثال ۲: در یک مطالعه مورد شاهدی برای بررسی رابطه رفتار مصرف سیگار با برخی از فاکتورهای محیطی در نوجوانان دانش‌آموز و غیر دانش‌آموز، تعداد ۱۸۰ مورد (سیگاری) و ۴۷۹ شاهد (غیرسیگاری) انتخاب و سابقه اصرار همسالان در آن‌ها پرسیده شد. نتایج این بررسی در جدول زیر خلاصه شده است:

جدول ۲-رابطه رفتار مصرف سیگار و سابقه اصرار همسالان

سابقه اصرار همسالان	مصرف سیگار		
	تعداد موارد سیگاری	تعداد موارد غیر سیگاری	تعداد کل
دارد	۱۱۳	۳۳	۵۵۹
ندارد	۶۷	۴۴۶	۱۰۰
تعداد کل	۱۸۰	۴۷۹	۶۵۹

منبع: [۵]

نوع مطالعه: مقطعی، ارائه بر حسب تعداد، تعداد نمونه = ۸۵۰

همان‌طور که در جدول فوق ملاحظه می‌شود از ۱۸۰ نفر افراد سیگاری ۱۱۳ نفر سابقه اصرار همسالان داشتند و از ۴۷۹ نفر افراد غیر سیگاری ۳۳ نفر آن‌ها سابقه اصرار همسالان نداشتند. هدف از این ارزیابی، بررسی رابطه بین مصرف سیگار و سابقه اصرار همسالان است، بنابراین در این مورد نیز اگر بتوان گفت که احتمال سیگاری بودن در افراد با سابقه اصرار همسالان بیشتر از احتمال سیگاری بودن در افراد بدون سابقه اصرار همسالان است آن گاه می‌توان رابطه بین مصرف سیگار و سابقه اصرار همسالان را نتیجه گرفت. اما این کار مستلزم محاسبه احتمال سیگاری بودن است که در این مطالعه به این دلیل که افراد سیگاری و غیر سیگاری توسط محقق انتخاب می‌شوند و از همان ابتدا دقیقاً پیامد مشخص است و در نتیجه بنا به توضیحات قبلی، محاسبه احتمال در مورد آن‌ها معنی ندارد. حال این سئوال پیش می‌آید که چگونه می‌توان رابطه بین بیماری و مواجهه را در این گونه مطالعات بررسی کرد. با توجه به جهت این مطالعات که گذشته‌نگر هستند و در این مطالعه، به سابقه افراد سیگاری و غیر سیگاری می‌نگرند که چقدر افراد در مواجهه قرار داشتند، برای حل مسئله می‌توان شدت مواجهه را بررسی نمود، یعنی به عبارت دقیق‌تر این موضوع را بررسی کرد افراد مورد (سیگاری) شدت مواجهه بیشتری داشتند یا افراد شاهد (غیر سیگاری) و در نتیجه اگر شدت مواجهه افراد سیگاری بیشتر باشد می‌توان گفت که مواجهه باعث مصرف سیگار شده است.

۱- احتمال سابقه اصرار همسالان (مواجهه) در گروه افراد سیگاری:

در این مورد هم با اعمال شرط گروه افراد سیگاری، با استفاده از مفهوم فراوانی نسبی، در این گروه احتمال به صورت زیر محاسبه می‌شود:

۱۱۳=تعداد افراد با سابقه اصرار همسالان در گروه سیگاری

۱۸۰=تعداد کل افراد در گروه سیگاری

که در این صورت احتمال سابقه اصرار همسالان در گروه سیگاری می‌شود ۱۱۳ تقسیم بر ۱۸۰ که می‌شود ۰/۶۲۸ و برحسب درصد، ۶۲/۸ درصد می‌شود. به عبارت دیگر، احتمال سابقه اصرار همسالان در گروه سیگاری برابر ۶۲/۸ درصد می‌شود. در این مورد هم احتمال محاسبه شده، یک احتمال شرطی است. بر اساس قاعده متمم، احتمال عدم سابقه اصرار همسالان در گروه سیگاری برابر ۳۲/۷ (=۶۲/۸-۱۰۰) درصد می‌شود.

در این مطالعات از مفهومی به نام شانس استفاده می‌شود که حاصل تقسیم دو احتمال محاسبه شده فوق است. به عبارت دیگر، شانس داشتن سابقه اصرار همسالان در گروه سیگاری به صورت زیر تعریف می شود:

احتمال سابقه اصرار همسالان در گروه سیگاری/احتمال عدم سابقه اصرار همسالان در گروه سیگاری= شانس داشتن سابقه اصرار همسالان در گروه سیگاری که مقدار عددی آن برابر ۱/۶۹ می‌شود.

۲- احتمال سابقه اصرار همسالان (مواجهه) در گروه افراد غیر سیگاری:

در این مورد هم با اعمال شرط گروه افراد غیر سیگاری، با استفاده از مفهوم فراوانی نسبی، در این گروه احتمال به صورت زیر محاسبه می‌شود:

۳۳=تعداد افراد با سابقه اصرار همسالان در گروه غیر سیگاری

۴۷۹=تعداد کل افراد در گروه غیر سیگاری

که در این صورت احتمال سابقه اصرار همسالان در گروه غیر سیگاری می‌شود ۳۳ تقسیم بر ۴۷۹ که می‌شود ۰/۰۷ و برحسب درصد، ۷/۰ درصد می‌شود. به عبارت دیگر، احتمال سابقه اصرار همسالان در گروه غیر سیگاری برابر ۷/۰ درصد می‌شود. در این مورد هم احتمال محاسبه شده یک احتمال شرطی است. بر اساس قاعده متمم، احتمال عدم سابقه اصرار همسالان در گروه غیر سیگاری برابر ۹۳/۰ (=۷/۰-۱۰۰) درصد می‌شود. شانس داشتن سابقه اصرار همسالان در گروه غیر سیگاری به صورت زیر تعریف می‌شود:

احتمال سابقه اصرار همسالان در گروه غیرسیگاری/احتمال عدم سابقه اصرار همسالان در گروه غیرسیگاری= شانس داشتن سابقه اصرار همسالان در گروه غیرسیگاری که مقدار عددی آن برابر ۱۳/۵۲ می‌شود.

حال اگر بتوان گفت که شانس داشتن سابقه اصرار همسالان در گروه سیگاری نسبت به شانس داشتن سابقه اصرار همسالان در گروه غیر سیگاری بیشتر است، در این صورت وجود رابطه بین مصرف سیگار و اصرار همسالان تایید می گردد. حال برای این که این بیان شکل کمی به خود گیرد می‌توان این گونه بیان کرد که اگر

شانس داشتن سابقه اصرار همسالان در گروه سیگاری نسبت به شانس داشتن سابقه اصرار همسالان در گروه غیر سیگاری

بزرگتر از عدد یک باشد آنگاه می‌توان گفت که مواجهه و سیگاری بودن رابطه دارند. در مثال حاضر، این کمیت برابر است ۸/۰۱ (۱۳/۵۲ تقسیم بر ۱/۶۹ درصد). بنابراین شانس داشتن سابقه اصرار همسالان در گروه سیگاری نسبت به شانس داشتن سابقه اصرار همسالان در گروه غیر سیگاری تقریباً ۸ برابر است و در نتیجه، احتمالاً بین مرحله سیگاری بودن و سابقه اصرار همسالان رابطه وجود داشته است. این شاخص همان (Odds Ratio (OR)) است.

لازم به ذکر است، در حالت کلی اگر OR برابر ۱ شود پیامد با مواجهه رابطه نخواهد داشت، اگر RR کوچکتر از یک شود مواجهه از پیامد جلوگیری می‌کند یعنی اثر محافظتی (Protective) دارد و اگر RR بزرگتر از یک شود مواجهه سبب ایجاد پیامد شده است (با آن رابطه دارد). در عمل، مقدار OR به تنهایی برای تصمیم‌گیری کافی نیست و فاصله اطمینان ۹۵ درصدی آن محاسبه می‌شود و زمانی که عدد یک یعنی عدد متناظر با عدم رابطه در این فاصله قرار گیرد نتیجه این می‌شود که بیماری و مواجهه رابطه ندارند.

مثال ۳: در مطالعه‌ای برای بررسی ارزش تشخیصی گلوکز پلاسمای ناشتا (Fasting Plasma Glucose(FPG)) در غربالگری دیابت بارداری (Gestational Diabetes Mellitus

(GDM))، ۱۰۰ زن در دوران بارداری برای این مطالعه انتخاب شدند که از میان آنها بر اساس معیار استاندارد ۳۶ نفر دارای GDM و ۶۴ نفر بقیه سالم بودند. همچنین بر اساس FPG، که در این مطالعه به عنوان تست تشخیصی در نظر گرفته شد، و با احتساب نقطه برش ۹۱ (mg/dl)، نتایج جدول زیر حاصل شد:

جدول ۳- ارزش تشخیصی FPG برای GDM

نتیجه FPG با نقطه برش ۹۱	GDM		
	دارند (بیمار)	ندارند (سالم)	تعداد کل
+	۲۳	۱۳	۳۶
–	۱۳	۵۱	۶۴
تعداد کل	۳۶	۶۴	۱۰۰

منبع: [۶]

نوع مطالعه: مقطعی، ارائه بر حسب تعداد، تعداد نمونه = ۱۰۰

در جدول فوق، به افرادی بیماری که به درستی براساس تست مثبت تشخیص داده شدند، مثبت واقعی (True Positive) (۲۳ نفر در این مثال)، به افرادی بیماری که به درستی بر اساس تست منفی تشخیص داده شدند، منفی واقعی (True Negative) (۵۱ نفر در این مثال)، به افرادی بیماری که به اشتباه براساس تست مثبت تشخیص داده شدند، مثبت کاذب (False Positive) (۱۳ نفر در این مثال) و به افرادی بیماری که به اشتباه براساس تست منفی تشخیص داده شدند، منفی کاذب (False Negative) (۱۳ نفر در این مثال)، گفته می‌شود.

برای ارزیابی یک تست تشخیصی چند سری شاخص محاسبه می‌شود که در این مثال، دو سری از این شاخص‌ها یعنی شاخص‌های مربوط به توان جداسازی افراد بیمار و سالم (به ترتیب شامل حساسیت و ویژگی) و شاخص‌های مربوط به میزان اطمینان به نتایج مثبت و منفی تست (به ترتیب شامل ارزش پیشگویی مثبت (Positive Predictive Value (PPV)) و ارزش پیشگویی منفی (Negative Predictive Value (NPV)) با دیدگاه احتمالی محاسبه می‌شوند.

الف- حساسیت:

حساسیت، توان جداسازی افراد بیمار برای تست تشخیصی است و تعریف احتمالی آن: احتمال مثبت بودن نتیجه تست در بین افراد بیمار است که یک احتمال شرطی می‌باشد. به عبارت دیگر، برای محاسبه حساسیت کافی است فراوانی نسبی نتایج مثبت تست در افراد بیمار (True Positive Rate) محاسبه شود. در مثال فوق:

۲۳=تعداد افراد با نتایج مثبت تست در افراد بیمار

۳۶=تعداد کل افراد بیمار

بنابراین حساسیت یعنی احتمال مثبت بودن تست در افراد بیمار برابر است با ۲۳ تقسیم بر ۳۶، که ۰/۶۴ می‌شود یا برحسب درصد برابر ۶۴/۰ درصد می‌شود.

ب- ویژگی:

ویژگی، توان جداسازی افراد سالم برای تست تشخیصی است و تعریف احتمالی آن: احتمال منفی بودن نتیجه تست در بین افراد سالم است که یک احتمال شرطی می‌باشد. به عبارت دیگر، برای محاسبه ویژگی کافی است فراوانی نسبی نتایج منفی تست در افراد سالم (True Negative Rate) محاسبه شود. در مثال فوق:

۵۱=تعداد افراد با نتایج منفی تست در افراد سالم

۶۴=تعداد کل افراد سالم

بنابراین ویژگی یعنی احتمال منفی بودن تست در افراد سالم برابر است با ۵۱ تقسیم بر ۶۴، که ۰/۸۰ می‌شود یا برحسب درصد برابر ۸۰/۰ درصد می‌شود.

ج- PPV:

شاخصPPV، میزان اطمینان به نتایج مثبت تست تشخیصی است و تعریف احتمالی آن: احتمال بیمار بودن فرد به شرط داشتن نتیجه مثبت است که یک احتمال شرطی می‌باشد، به عبارت دیگر برای محاسبه PPV کافی است فراوانی نسبی افراد بیمار در نتایج مثبت تست محاسبه شود. در مثال فوق:

۲۳=تعداد افراد با نتایج مثبت تست در افراد بیمار

۳۶=تعداد کل افراد بیمار

بنابراین PPV یعنی احتمال بیمار بودن فرد به شرط داشتن نتیجه مثبت برابر است با ۲۳ تقسیم بر ۳۶، که ۰/۶۴ می‌شود یا برحسب درصد برابر ۶۴/۰ درصد می‌شود.

شایان ذکر است در محاسبه PPV، از قاعده‌ای استفاده شد که اصطلاحاً قاعده بیز در احتمال نامیده می‌شود (این قاعده توسط توماس بیز کشیش انگلیسی در قرن هجدهم معرفی شد). برای روشن‌تر شدن این موضوع به مثال مراجعه می‌کنیم:

در این مثال، قبل از اعمال تست تشخیصی، فراوانی نسبی افراد GDM برابر ۳۶ به ۱۰۰ یعنی ۳۶ درصد بود که به شرط معرف بودن نمونه این عدد معادل میزان شیوع عارضه مورد بررسی در جامعه آماری است. بنابراین احتمال این که فردی در جامعه آماری مورد بررسی دارای GDM باشد برابر ۳۶ درصد است که در قاعده بیز به آن احتمال اولیه (Prior) برای بیماری مورد بررسی گفته می‌شود. بر اساس نتایج تست تشخیصی، ۲۳ نفر از ۳۶ نفر افراد بیمار (۶۴ درصد) و ۱۳ نفر از ۶۴ نفر افراد سالم (۲۰ درصد) بودند. حال بر اساس ترکیب این نتایج، شاخص PPV محاسبه می‌شود که فراوانی نسبی اولیه این بیماری (یعنی داده‌ها) را با یک میانگین وزنی از توان احتمالی موجود برای تشخیص مثبت در این تست (یعنی حساسیت و عکس ویژگی) ترکیب می‌کند واحتمال جدیدی را حاصل می نماید که در قاعده بیز به آن احتمال ثانویه (Posterior) برای بیماری مورد بررسی گفته می‌شود که در این مثال برابر ۶۴ درصد به دست آمد. بنابراین اگر این تست تشخیصی ملاک تشخیص افراد برای GDM استفاده شود، احتمال این که بر اساس نتیجه مثبت تست تشخیصی، فردی در جامعه آماری مورد بررسی دارای GDM باشد برابر ۶۴ درصد است.

د- NPV:

شاخص NPV، میزان اطمینان به نتایج منفی تست تشخیصی است و تعریف احتمالی آن: احتمال سالم بودن فرد به شرط داشتن نتیجه منفی است که یک احتمال شرطی می‌باشد، به عبارت دیگر برای محاسبه NPV کافی است فراوانی نسبی افراد سالم در نتایج منفی تست محاسبه شود. در مثال فوق:

۵۱=تعداد افراد با نتایج مثبت تست در افراد بیمار

۶۴=تعداد کل افراد بیمار

بنابراین NPV یعنی احتمال سالم بودن فرد به شرط داشتن نتیجه منفی برابر است با ۵۱ تقسیم بر ۶۴ که ۰/۸۰ می‌شود یا برحسب درصد برابر ۸۰/۰ درصد می‌شود.

مثال ۴: در مطالعه‌ای با هدف بررسی رابطه میان مصرف سیگار، مصرف الکل با مرگ ناشی از سرطان‌های کولورکتال، نتایج در جداول ۴ تا ۶ نشان داده شده است:

جدول ۴- فراوانی مرگ و میر افراد مبتلا به سرطان کولورکتال برحسب رده‌های متغیرهای وضعیت کشیدن سیگار

وضعیت کشیدن سیگار	تعداد مرگ‌و میر	تعداد کل افراد
هرگز مصرف نکرده	۲۲۵	۸۳۲
در حال حاضر یا قبلاً مصرف کرده	۹۷	۲۸۴

منبع: [۷]

نوع مطالعه: طولی، ارائه بر حسب تعداد، تعداد نمونه = ۱۲۱۹

جدول ۵- فراوانی مرگ و میر افراد مبتلا به سرطان کولورکتال برحسب رده‌های متغیرهای وضعیت مصرف الکل

وضعیت مصرف الکل	تعداد مرگ‌ومیر	تعداد کل افراد
هرگز مصرف نکرده	۲۹۵	۱۰۳۶
در حال حاضر یا قبلاً مصرف کرده	۳۳	۹۹

منبع: [۷]

نوع مطالعه: طولی، ارائه بر حسب تعداد، تعداد نمونه = ۱۲۱۹

جدول ۶- فراوانی مرگ ومیر افراد مبتلا به سرطان کولورکتال بر حسب رده‌های متغیرهای وضعیت مصرف الکل ومصرف سیگار

وضعیت مصرف الکل	وضعیت کشیدن سیگار	تعداد مرگ‌ومیر	تعداد کل افراد
هرگز مصرف نکرده	هرگز مصرف نکرده	۲۱۲	۷۹۶
	در حال حاضر یا قبلا مصرف کرده	۷۱	۲۱۰
در حال حاضر یا قبلا مصرف کرده	هرگز مصرف نکرده	۸	۲۹
	در حال حاضر یا قبلا مصرف کرده	۲۴	۶۹

منبع: [۷]

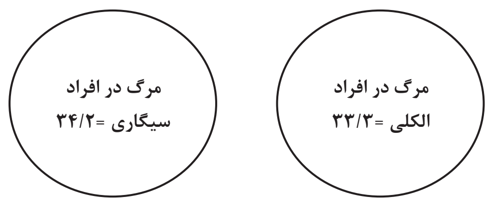
نوع مطالعه: طولی، ارائه بر حسب تعداد، تعداد نمونه = ۱۲۱۹

هدف از ارائه این مثال، محاسبه احتمالات و ارائه برخی قواعد مرتبط با احتمال می‌باشد:

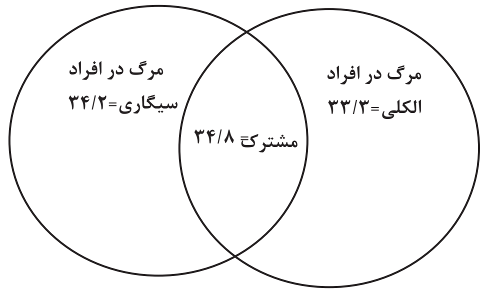
الف- احتمال (خطر) مرگ ناشی از سرطان کولورکتال در افراد سیگاری، برابر است با ۹۷ تقسیم بر ۲۸۴ که می‌شود ۰/۳۴۲، یا برحسب درصد برابر ۳۴/۲ درصد می‌شود. همان طور که قبلاً اشاره شد این احتمال، شرطی است (جدول ۴).

ب- احتمال (خطر) مرگ ناشی از سرطان کولورکتال در افراد الکلی، برابر است با ۳۳ تقسیم بر ۹۹ که می‌شود ۰/۳۳۳، یا برحسب درصد برابر ۳۳/۳ درصد می‌شود (احتمال شرطی) (جدول ۵).

ج - احتمال (خطر) مرگ ناشی از سرطان کولورکتال در افراد سیگاری یا الکلی (این احتمال در یافتن بار بیماری در بین افراد دارای این دو ویژگی کمک می‌کند)، بنابراین ممکن است با خود بگوییم این احتمال را می توان با جمع بستن دو احتمال برابر ۶۷/۵ (= ۳۳/۳ + ۳۴/۲) به دست آورد یعنی بر اساس مجموع دو دایره موجود در دیاگرام زیر:



ولی این احتمال صحیح نیست، چون تعداد ۲۴ نفر از مرگ‌های ناشی از سرطان کولورکتال، (۳۴/۸درصد) به طور مشترک سیگاری و الکلی بودند (جدول ۶). به عبارت دیگر:



براساس نمودار فوق با جمع بستن دو احتمال یعنی دو دایره، قسمت مشترک دو بار در محاسبات حساب می‌شود که نیاز است یک بار آن را کم نمود، بنابراین زمانی که هدف محاسبه احتمال برای رخداد پیامد اول **یا** دوم باشد، در صورتی که این دو پیامد وجه مشترک داشته باشند نیاز است احتمال وجه مشترک از احتمال مجموع رخداد دو پیامد کم شود؛ بنابراین احتمال مرگ ناشی از سرطان کولورکتال در افراد سیگاری **یا** الکلی باشد برابر است با:

۳۴/۲+۳۳/۳-۳۴/۸=۳۲/۷

در حالت کلی: زمانی که هدف محاسبه احتمال برای رخداد پیامد اول یا دوم باشد، در صورتی که این دو پیامد وجه مشترک داشته باشند نیاز است احتمال وجه مشترک از احتمال مجموع رخداد دو پیامد کم شود. به این احتمال، احتمال اجتماع دو پیامد گفته می‌شود.

مثال ۵: با مراجعه به مثال ۴، هدف بررسی رابطه تعاملی بین الکل و سیگار در مرگ ناشی از سرطان کولورکتال از دیدگاه احتمال است به عبارت دیگر، آیا احتمال مرگ ناشی از سرطان کولورکتال در افراد سیگاری به الکلی بودن افراد یا احتمال مرگ ناشی از سرطان کولورکتال در افراد الکلی به سیگاری بودن افراد بستگی دارد یا خیر. برای این منظور احتمالات زیر محاسبه می شود:

الف- مرگ ناشی از سرطان کولورکتال در افراد سیگاری: که در مثال ۴ برابر ۳۴/۲ درصد محاسبه شد.

ب- مرگ ناشی از سرطان کولورکتال در افراد الکلی: که در مثال ۴ برابر ۳۳/۳ درصد محاسبه شد.

ج- مرگ ناشی از سرطان کولورکتال در افراد سیگاری و الکلی (احتمال اشتراک): که در مثال ۴ برابر ۳۴/۸ درصد محاسبه شد.

اگر احتمال مرگ ناشی از سرطان کولورکتال در افراد سیگاری به الکلی بودن افراد بستگی نداشته باشد (یا احتمال مرگ ناشی از سرطان کولورکتال در افراد الکلی به سیگاری بودن افراد بستگی نداشته باشد)، در این صورت انتظار می‌رود که دو پیامد فوق به طور مستقل از هم عمل کنند، به طوری که بتوان احتمال سرطان کولورکتال در افراد سیگاری **و** الکلی یعنی اشتراک دو پیامد را بر اساس احتمالات انفرادی و با حاصل ضرب آنها به دست آورد. حاصل ضرب انفرادی دو احتمال مرگ ناشی از سرطان کولورکتال در افراد سیگاری و مرگ ناشی از سرطان

کولورکتال در افراد الکلی برابر ۱۱/۳۹=(۳۳/۳×۳۴/۲) درصد می‌شود که با احتمال مرگ ناشی از سرطان کولورکتال در افراد سیگاری **و** الکلی (احتمال اشتراک) برابر نمی‌شود ودر نتیجه دو احتمال انفرادی فوق به طور مستقل نتوانستند احتمال اشتراک را ایجاد نمایند و در نتیجه دو پیامد فوق از هم مستقل نیستند و با هم رابطه دارند. به عبارت دیگر، احتمال مرگ ناشی از سرطان کولورکتال در افراد سیگاری به الکلی بودن افراد یا احتمال مرگ ناشی از سرطان کولورکتال در افراد الکلی به سیگاری بودن افراد بستگی دارد.

در حالت کلی: اگر احتمال اشتراک دو پیامد با حاصل ضرب احتمال انفرادی آنها برابر باشد، این یعنی اینکه دو پیامد به طور مستقل از هم عمل می‌کنند و احتمال اشتراک دو پیامد را می‌توان بر اساس احتمالات انفرادی هر یک از پیامدها به دست آورد. در این صورت دو پیامد با هم رابطه ندارند.

توزیع‌های احتمال

همان‌طور که اشاره شد، برای انجام استنباط آماری روی داده‌هایی که به تصادف انتخاب شده‌اند، نیاز است که از پل ارتباطی احتمال و مدل‌های احتمالی استفاده شود که این موضوع توسط توزیع‌های احتمال فراهم می‌شود. در ادامه سه مدل احتمالی پر استفاده در علوم پزشکی معرفی خواهند شد.

۱- توزیع احتمال دو جمله‌ای

این توزیع احتمالی برای متغیرهای کیفی دو حالتی (و تعمیم آن برای متغیرهای کیفی چند حالتی) به کار می‌رود و در انواع مطالعات علوم پزشکی، مبنای محاسبه شاخص‌های بررسی رابطه پیامدهای با خصوصیت فوق و ریسک فاکتورها قرار می‌گیرد.

با مراجعه به موضوع جوان و دولتمند، هدف یافتن نحوه محاسبه آن احتمال است. چون تعداد پرتاب‌ها ده تا در نظر گرفته شده بود و محاسبه آن کمی وقت‌گیر است، مسئله در حالت سه پرتاب بررسی می‌گردد ولی نتایج، به تعداد بیشتر پرتاب قابل تعمیم است.

مثال ۶: اگر یک سکه سالم سه بار پرتاب شود، احتمال این که دو شیر مشاهده شود چقدر است؟

برای حل مسئله، ابتدا باید در نظر داشته باشیم که قرار است ۲ پرتاب از ۳ پرتاب شیر باشند که می‌تواند در سه حالت مختلف زیر رخ دهد (H نمادی برای شیر و T نمادی برای خط است):

H H T
H T H
T H H

بنابراین در هر یک از سه حالت فوق، ممکن است دو شیر وجود داشته باشد، ولی چون احتمالات مربوط برای هر سه حالت یکسان است یعنی اگر احتمال آوردن شیر در حالت کلی با p و احتمال آمدن خط با (1-p) نشان داده شود، آنگاه برای سه حالت فوق، احتمالات به صورت زیر خواهد بود:

$$p\times p\times (1-p)$$
$$p\times(1-p) \times p$$
$$(1-p) \times p\times p$$

علت این که بین احتمالات از عملگر ضرب استفاده شده این است که در محاسبات فوق فرض استقلال پرتاب‌ها از یکدیگر در نظر گرفته شده است یعنی فرض شده است که از روی پرتاب اول نمی‌توان نتیجه پرتاب دوم را حدس زد. به هرحال همان طور که مشاهده شد (با توجه به این که برای عملگر ضرب جا به جایی امکان پذیر است) در هر سه حالت فوق احتمال یکسان $p^2\times(1-p)$ به دست می‌آید. بنابراین کافی است یکی از سه احتمال حساب شود و در سه (تعداد کل حالت‌ها) ضرب گردد. برای یافتن یک قاعده کلی برای تعداد کل حالت‌ها، ابتدا آن را در همین مثال بررسی می‌کنیم. در این مثال، هدف یافتن تعداد حالاتی بود که از ۳ پرتاب ۲ شیر رخ می‌دهد که در ریاضیات آن را ترکیب دوتایی از سه تایی می‌نامند و با نماد $(\frac{۳}{۲})$ نشان داده می‌شود که فرمول آن به صورت زیر است: $(\frac{3}{2}) = \frac{3!}{2!\,(3-2)!}$

که اگر به جای ۳ (یعنی تعداد کل) حرف n و به جای ۲ (یعنی تعداد مورد نظر برای رخداد) حرف k آورده شود، خواهیم داشت: $(\frac{n}{k}) = \frac{n!}{k!(n-k)!}$

بنابراین، احتمال مورد نظر به صورت زیر نتیجه خواهد شد: $(\frac{3}{2}) \times p^2 \times (1-p)^1$

یعنی یکی از احتمالات ضرب در تعداد کل حالات. در توان عبارت (1-p) عدد ۱ و در واقع از تفاضل ۳ از ۲ به دست می‌آید که برای دستیابی به قاعده کلی به جای ۱ از (۲–۳) استفاده می‌شود، یعنی: $(\frac{3}{2}) \times p^2 \times (1-p)^{(3-2)}$

حال اگر به جای ۳ و ۲ به ترتیب حرف‌های n و k به کار برده شود، خواهیم داشت: $(\frac{n}{k})p^k \times (1-p)^{(n-k)}$

بنابراین، قاعده کلی فوق برای محاسبه احتمال k پیشامد از n آزمایش به دست می‌آید. رابطه فوق، قاعده‌ی احتمال دوجمله‌ای است.

حال اگر در مثال فوق، سکه سالم باشد که $p=\frac{1}{2}$ را نتیجه می‌دهد، احتمال مورد نظر برابر است با $3 \times (\frac{1}{2})^2 \times (\frac{1}{2}) = \frac{3}{8}$

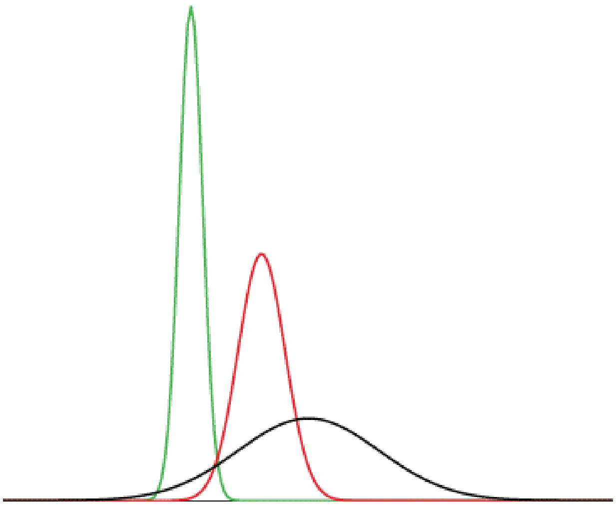
۲- توزیع احتمال پواسن

یکی دیگر از توزیع‌ها و قواعد پرکاربرد احتمال، قاعده احتمال پواسن است که برای بررسی فرآیندهای شمارشی به کار می‌رود. به عنوان مثال، برای بررسی این احتمال که در یک ساعت مشخص یک تعداد (شمارش) مشخص از تعداد ذرات α از یک شیء رادیواکتیو خارج می‌شود، به شرط آن که نرخ متوسط خارج شدن آن معلوم باشد، چقدر است. به عنوان یک مثال دیگر، فرض کنید که در ژاپن به طور متوسط در هر روز سه زلزله می‌آید که حسب تجربیات گذشته در یک مدت زمان طولانی این نتیجه حاصل شده باشد. حال برای به دست آوردن احتمال این که در یک روز خاص، یک تعداد (شمارش) معلوم از زلزله رخ دهد چقدر است. در هر دو مثال فوق، هدف بررسی احتمال تعداد مشخصی از وقایع است که در یک واحد زمانی مشخص روی می‌دهند. لازم به ذکر است که واحد اندازه‌گیری مورد بررسی لزوماً واحد زمانی نیست و می‌تواند واحد مکانی و یا هر واحد دیگری از اندازه نیز باشد. به عنوان مثال دیگر، اگر هدف محاسبه احتمال رخداد تعداد مشخصی تصادف در یک محدوده معین جغرافیایی باشد نیز از توزیع احتمال پواسن استفاده می‌شود.

مورد کاربرد این نوع احتمال در مطالعاتی نظیر همگروهی یا مورد – شاهدی و یا مقطعی، مشاهده می‌شود که در آنها پیامد مورد نظر در دو یا چند سطح و در ارتباط با یک مواجهه در دو یا چند سطح بررسی می‌گردد. مثلاً در یک مطالعه مورد شاهدی برای بررسی رابطه مصرف (یا عدم مصرف) سیگار با سرطان ریه (داشتن یا نداشتن)، پیامد و مواجهه هرکدام دو حالتی هستند. در عمل، برای ارزیابی نتایج باید مشخص باشد که چه تعداد (شمارشی) از افراد در هر کدام از گروه‌های مراجعه یا عدم مراجعه با سیگار دچار سرطان ریه می‌شوند و چه تعدادی از افراد سالم باقی می‌مانند که به منظور ساختن شاخص‌های مناسب برای بررسی رابطه در آنها، از توزیع احتمالی پواسن استفاده می‌شود.

۳- توزیع نرمال

یکی دیگر از توزیع‌های مهم مورد استفاده در بحث آمار، توزیع نرمال است که تقریباً مهم‌ترین توزیع مورد استفاده در مباحث آمار و مخصوصاً آمار کاربردی است. این منحنی زنگوله‌ای شکل و متقارن است و بسته به مکان میانگین و میزان پراکندگی (انحراف معیار) می‌تواند شکل‌های مختلفی داشته باشد (شکل ۱):



شکل ۱- توزیع نرمال با میانگین ها و انحراف معیارهای متفاوت

نام دیگر این توزیع "منحنی گاوسی" است که به دلیل تلاش‌های کارل فردریش گاوس در این زمینه می‌باشد. توزیع نرمال، توسط ابراهام دموآور کشف شد؛ بدن ترتیب که او حین مشاوره دادن به قماربازان متوجه شد که با افزایش تعداد پرتاب‌های سکه، شکل احتمالی توزیع آنها به یک منحنی هموار میل می‌کند. دموآور این منطق را جستجو نمود که اگر بتوان برای این منحنی هموار معادله ریاضی پیدا کرد می‌توان احتمالاتی نظیر مشاهده تعداد ۶۰ یا بیشتری از شیرها را در پرتاب ۱۰۰ سکه به دست آورد. برای روشن شدن موضوع مثال زیر را ببینید.

مثال ۷: در پرتاب یک سکه سالم اگر شیر آوردن سکه ویژگی مورد نظر باشد، آنگاه با احتمال ۰/۵ شیر (H) و با احتمال ۰/۵ سکه خط (T) می‌آید که به ترتیب با p و q نشان داده می‌شوند (q=1-p). حالت‌های مختلف برای پرتاب دو سکه به صورت زیر خواهد بود:

HH ۲ شیر ٠/٥ × ٠/٥ = ٠/٢٥

TH ۱ شیر ٠/٥ × ٠/٥ = ٠/٢٥

HT ۱ شیر ٠/٥ × ٠/٥ = ٠/٢٥

٠/٥

TT ۰ شیر ٠/٥ × ٠/٥ = ٠/٢٥

حال فرض کنید احتمال بهبود یافتن (R) فردی از بیماری برابر ۰/۴ باشد (P=۰/۴) در نتیجه احتمال عدم بهبودی (NR) برابر ۰/۶ (q=۰/۶) خواهد بود. و در نتیجه احتمالات مرتبط با بهبودی دو نفر به صورت زیر خواهد بود:

RR ۲ بهبودی ٠/٤ × ٠/٤ = ٠/١٦

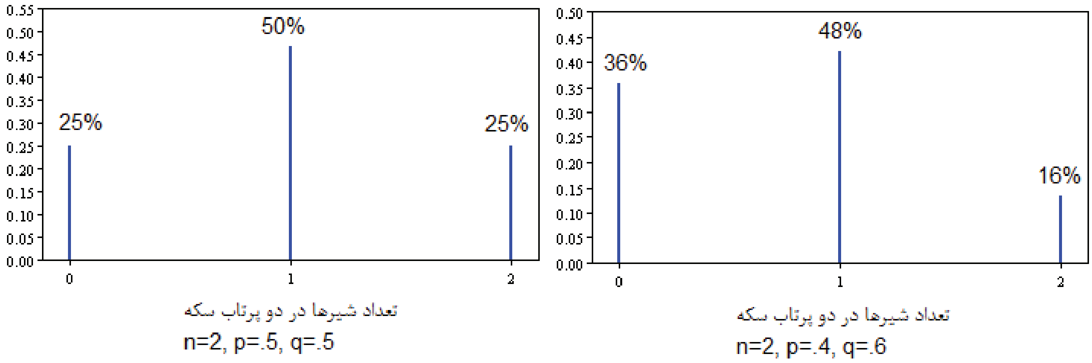
NRR ۱ بهبودی ٠/٤ × ٠/٦ = ٠/٢٤

RNR ۱ بهبودی ٠/٦ × ٠/٤ = ٠/٢٤

٠/٤٨

NRNR ۰ بهبودی ٠/٦ × ٠/٦ = ٠/٣٦

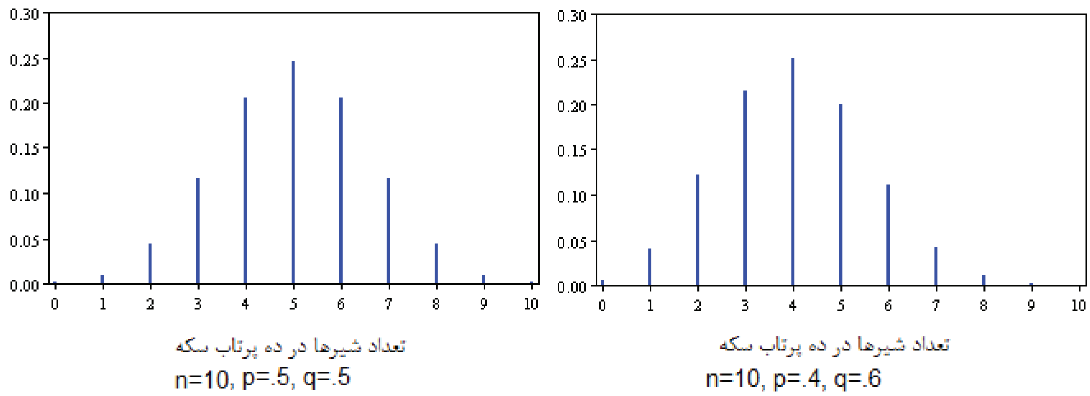
نحوه توزیع شدن احتمال به ازای ۰، ۱ و ۲ شیر یا ۰، ۱ و ۲ بهبودی در مثال‌های فوق، یعنی اختصاص ۰/۲۵، ۰/۵، ۰/۲۵ یا ۰/۳۶، ۰/۴۸ و ۰/۱۶ را توزیع احتمال می‌نامند. توزیع احتمال مثال‌های فوق، در شکل زیر ارائه شده است (شکل ۲):



شکل ۲- توزیع احتمال دوجمله‌ای برای n=۲ و در دوحالت P= ۰/۵ (چپ) و P= ۰/۴ (راست)

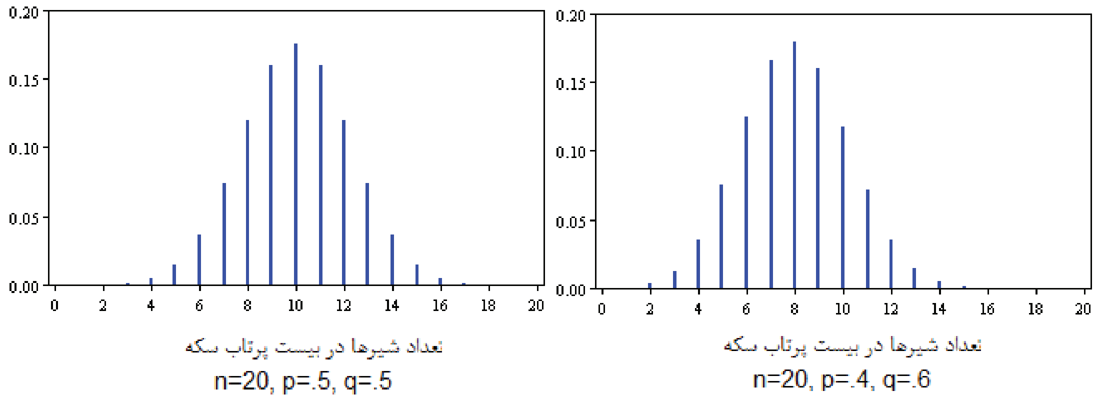
دو نمودار فوق از احتمالات که به ازای دو نمونه ارائه شده است، مثال‌هایی از مجموعه بزرگتری از توزیع‌های احتمال هستند که آنها را توزیع‌های نمونه‌گیری احتمال می‌نامند. بنابراین توزیع نمونه‌گیری دو پرتاب یک سکه (۰/۵ = 1−p و ۰/۵ =p و ۲ =n) مشخص می‌نماید که نمونه‌ای تصادفی از پرتاب‌های دو سکه ۲۵٪ شانس رخداد صفر شیر ۵۰٪ شانس رخداد یک شیر و ۲۵٪ شانس رخداد دو شیر را دارند (و می‌توان این احتمالات را بسط داد مثلاً احتمال رخداد حداقل یک شیر (یعنی یک یا دو شیر) برابر ۷۵٪ = ۲۵٪ +۵۰٪ می‌باشد). این تعابیر را می‌توان به گونه‌ای دیگر نیز بیان کرد که با عنوان «توزیع نمونه‌گیری» نیز همخوانی داشته باشد. به عبارت دیگر، در یک نمونه بزرگ از پرتاب‌های دو سکه انتظار می رود ۲۵، ۵۰ و ۲۵ درصد از موارد، به ترتیب رخداد صفر شیر، یک شیر و دو شیر وجود داشته باشد. یا در مورد مثال بهبودی، می‌توان در یک نمونه بزرگ از بیماران با شرایط این مثال، انتظار داشت در ۳۶، ۴۸ و ۱۶ درصد از موارد، به ترتیب دو، یک و صفر بهبودی وجود داشته باشد.

همان‌طور که مشاهده می‌شود، نمودار مربوط به P=۰/۵ متقارن است، در حالی که نمودار مربوط به P=۰/۴ کمی نامتقارن است. عدم تقارن نمودار، زمانی که مقدار P به ۱ یا ۰ نزدیک می‌شود، بیشتر و بیشتر می‌گردد. زمانی که حجم نمونه افزایش یابد، به عنوان مثال، اگر از n=۲ به n=۱۰، افزایش در حجم نمونه وجود داشته باشد، در این صورت منحنی مزبور را می‌توان در نمودار زیر مشاهده نمود (شکل ۳):



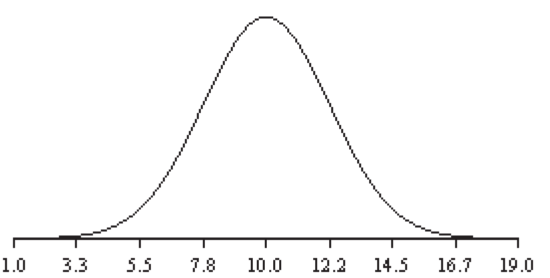
شکل ۳- توزیع احتمال دوجمله‌ای برای n=۱۰ و در دوحالت P= ۰/۵ (چپ) و P= ۰/۴ (راست)

در این نمودارها شکستگی‌های منحنی نسبت به منحنی قبل کمتر و کمتر شده است، به طوری که در این نمودار منحنی به سمت یک منحنی هموار میل می‌کند. زمانی که حجم نمونه به ۲۰ = n افزایش می‌یابد منحنی‌های آنها را می‌توان در نمودار زیر مشاهده کرد (شکل ۴):



شکل ۴- توزیع احتمال دوجمله‌ای برای n=۲۰ و در دو حالت P= ۰/۵ (چپ) و P= ۰/۴ (راست)

در این نمودار زیر شکستگی‌ها هموارتر شده‌اند و شبیه به منحنی شده‌اند. مثال‌های فوق نشان می‌دهد که با افزایش حجم نمونه، منحنی‌های توزیع دوجمله‌ای هموارتر می‌شوند و مخصوصاً در حالت p = ۰/۵ شکل متقارنی دارند و توزیع احتمال دیگری را پیشنهاد می‌کنند که آن را توزیع احتمال نرمال می‌نامند. برای حالت ۰/۵ در شکل فوق می‌توان تقریب نرمال آن را به صورت زیر مشاهده کرد (شکل ۵):



شکل ۵ - توزیع نرمال با میانگین ۱۰ و انحراف معیار ۲/۲۴ (برای n=۲۰ و P= ۰/۵ در شکل ۴)

توزیع نرمال یک توزیع احتمال متقارن است که در آن میانگین و میانه و مد روی یک نقطه، یعنی نقطه وسط منحنی قرار دارد. ارتفاع منحنی در نقطه میانگین بیشترین مقدار را دارد و در دو دنباله توزیع به کمترین مقدار خود می‌رسد. در این توزیع کل سطح زیر منحنی برابر ۱ است.

برای انجام تقریب برای توزیع دو جمله‌ای، زمانی که حاصلضرب np و n× (۱− p) بزرگ‌تر از ۵ یا مساوی آن باشند، در این صورت توزیع دوجمله‌ای به نرمال نزدیک می‌شود و این تقریب مناسب است. به عنوان مثال، در مثال پرتاب سکه (P=۰/۵)، تعداد ۱۰ نمونه برای رسیدن به این تقریب کافی است (۵ = ۰/۵ ×۱۰) یا برای مثال بهبودی بیماران، تعداد ۱۳ نمونه لازم است (۷/۸=۱۳×۰/۶ و ۵/۲=۱۳×۰/۴). تقریب فوق که به ازای حجم نمونه بزرگ امکان‌پذیر می‌شود و براساس آن شکل توزیعی متغیر دوجمله‌ای (نظیر بهبودی یا عدم بهبودی بیمار) را می‌توان توسط یک توزیع احتمال دیگر، یعنی نرمال، تقریب زد، در قالب قضیه‌ای به نام **قضیه حد مرکزی** توجیه می‌شود. قضیه حد مرکزی، بیان می‌کند که توزیع اولیه داده‌ها هر چه که باشد، وقتی حجم نمونه بزرگ انتخاب شود (در مثال های فوق np>۵ و n(۱−p)>۵ و در حالت کلی براساس توافق اکثریت n≥۳۰)، در این صورت میانگین یا نسبت توزیع نرمال خواهند داشت. توزیع نرمال، توزیع احتمالی مخصوص متغیرهای کمی پیوسته است که بر اساس مشخصه‌های میانگین و انحراف معیار، شکل دقیق و توزیعی آن مشخص می‌شود. زمانی که از آن برای تقریب توزیع دوجمله‌ای استفاده می‌شود، میانگین و انحراف معیار آن بر اساس فرمول‌های ارائه شده برای توزیع دوجمله‌ای محاسبه می‌شوند. در توزیع احتمال دوجمله‌ای میانگین، واریانس و انحراف معیار به صورت زیر به دست می‌آیند:

$$\text{میانگین} = n\times p$$

$$\text{واریانس} = n \times p \times (1-p)$$

$$\text{انحراف معیار} = \sqrt{p \times (1-p) \times n}$$

در نتیجه در مثال‌های فوق به صورت زیر میانگین و انحراف معیار به صورت زیر به دست می‌آیند:

برای مثال شیر و خط:

۱= ۰/۵ × ۲ = میانگین

۰/۷۱ = (۲ × ۰/۵ × ۰/۵) √ = انحراف معیار

برای مثال بهبودی بیماران:

۰/۸ = ۰/۴ × ۲ = میانگین

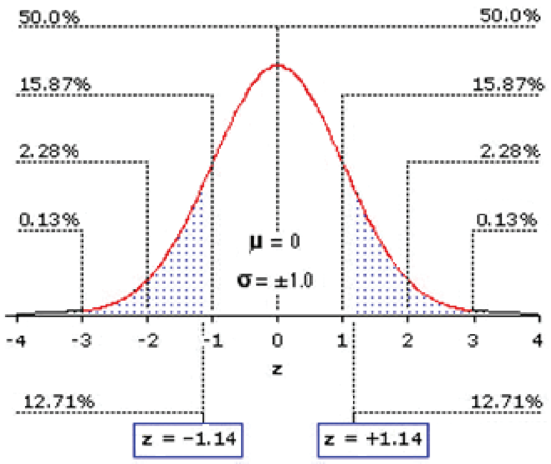
۰/۶۹ = (۲ × ۰/۴ × ۰/۶) √ = انحراف معیار

به عبارت دیگر، در مثال پرتاب سکه (به طور متوسط) انتظار می‌رود در دو پرتاب یک سکه سالم ۱ شیر مشاهده شود و در مثال بهبودی بیماران (به طور متوسط) انتظار می‌رود به ازای هر ۲۰ مریض ۸ مورد بهبودی مشاهده شود (هر دو کمیت تعداد بیماران و تعداد بهبودی در عدد ۱۰ ضرب شدند تا تفسیر ساده‌تری داشته باشند).

توزیع نرمال استاندارد

با توجه به این که تعداد تقریب‌ها و ویژگی‌های زیادی را می‌توان یافت که از توزیع نرمال تبعیت می‌کنند و هر کدام میانگین‌ها و همچنین انحراف معیارهای مختلفی دارند، بنابراین در عمل بیشمار توزیع نرمال وجود دارد که کار کردن با همه آنها دشوار است. راه حل مناسبی که برای حل این مسئله وجود دارد استاندارد کردن توزیع‌های مختلف نرمال است. در فرآیند استاندارد کردن، هر یک از اعداد خام منهای میانگین خود شده و بر انحراف معیار خود تقسیم می‌شوند که بدین ترتیب اعداد استاندارد شده به دست می‌آیند که میانگین و انحرف معیار متغیر جدید، که آن را با Z ($Z= (X-\mu)/\sigma$) نشان می‌دهند، به ترتیب برابر صفر و یک می‌شود.

شکل توزیع احتمالی این متغیر و میزان پوششی را که این توزیع به ازای مقادیر ممکن خود فراهم می‌کند در شکل زیر می‌توان مشاهده نمود (شکل ۶):



شکل ۶- سطح خارج منحنی به ازای مقادیر مختلف در توزیع نرمال استاندارد

توزیع نرمال استاندارد، توزیع نرمال با میانگین صفر و انحراف معیار ۱ است. سطح زیر منحنی بالای ۲ انحراف معیار (یعنی عدد ۲)، تقریباً ۲/۲۸٪ از سطح زیر منحنی را در بر می‌گیرد و چون منحنی متقارن است پایین دو انحراف معیار (یعنی عدد ۲-)، تقریباً ۲/۲۸٪ از سطح زیر منحنی را می‌پوشاند که مجموعاً ۴/۵۶٪ از سطح زیر منحنی است و در نتیجه ۹۵/۴۶٪ از سطح زیر منحنی بین دو عدد ۲- و ۲+ (۲±) قرار می‌گیرد. در این توزیع، تقریبا ۶۸/۲۷ درصد از زیر منحنی در دامنه ۱± قرار می‌گیرد.

در عمل برای محاسبات مربوط به سطح زیر منحنی نرمال، جداولی نظیر جدول زیر، وجود دارد که با مراجعه به آن می‌توان با مشخص بودن اعداد مورد نظر، احتمالات را به دست آورد (شکل ۷):

±z	Area Beyond ±z	±z	Area Beyond ±z
0.00	.5000	1.80	.0359
0.20	.4702	**1.90**	**.0287**
0.40	.3446	2.00	.0228
0.60	.2743	2.20	.0139
0.80	.2119	2.40	.0082
1.00	.1587	2.60	.0047
1.14	**.1271**	2.80	.0026
1.20	.1151	3.00	.0013
1.40	.0808	3.50	.0002
1.60	.0548	4.00	.00003

برای مثال سطح خارج منحنی برای عدد ۱/۱۴ ± برابر ۰/۱۲۷۱ است.

شکل ۷- سطح خارج منحنی به ازای مقادیر مختلف در توزیع نرمال استاندارد

به عنوان مثال، برای مسئله‌ای در مورد بهبودی بیماران با شرایط مثال قبل، احتمال بهبودی ۴۳۰ نفر و یا بیشتر از تعداد کل ۱۰۰۰ نفر چقدر است، اگر از فرمول دو جمله‌ای استفاده شود، در این صورت باید ابتدا انتخاب ۴۳۰ از ۱۰۰۰ یعنی ترکیب $\binom{1000}{430}$ را حساب کرد و آنرا در $(۰/۶)^{۵۷۰}\times(۰/۴)^{۴۳۰}$ ضرب نمود. ولی با استفاده از تقریب نرمال می توان به صورت زیر عمل کرد:

میانگین=۱۰۰۰×۰/۴=۴۰۰

انحراف معیار=√(۱۰۰۰×۰/۴ ×۰/۶)=۱۵/۴۹

z=((۴۳۰−۴۰۰)−۰/۵)÷۱۵/۴۹=۲۹/۵÷۱۵/۴۹=۱/۹۰

بنابراین با مراجعه به جدول نرمال (شکل ۷)، سطح زیر منحنی بالای ۱/۹۰ برابر ۲/۸۷٪ و در نتیجه سطح زیر منحنی پایین ۱/۹۰ برابر ۹۷/۱۳٪ به صورت زیر به دست می‌آید. بنابراین احتمال مشاهده بیشتر یا مساوی ۴۳۰ بهبودی در ۱۰۰۰ نفر برابر ۹۷/۱۳٪ است.

نکته:

۱) توزیع نرمال کاربرد گسترده‌ای در مسایل استنباط آماری دارد، به طوری که در بسیاری از مطالعات برای انجام تحلیل‌های آماری ابتدا فرض می‌شود که توزیع متغیرهای پیامد مورد بررسی نرمال است و بر اساس آن روش‌های مناسب برای تحلیل داده‌ها استفاده می‌شود [۸].

۲) علاوه بر توزیع‌های نمونه‌گیری معرفی شده، توزیع‌های نمونه‌گیری دیگری نظیر t، کای‌دو و F نیز در بحث استنباط آماری و روش‌های آن نقش زیادی دارند که در سری‌های آتی معرفی خواهند شد.

مأخذ

1. Hill AB. *Principles of Medical Statistics*. London: Lancet 1937.
2. Fisher M. *The Instant Millionaire: A Tale of Wisdom and Wealth*. NY: Amazon Pgw 1990.
3. Asghari Jafarabadi M, Soltani A, Mohammadi SM. Statistical Series: Summarizing and Displaying Data. *Journal of Diabetes and Metabolic Disorders* 2013 under Press, [In Persian].
4. 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.
5. Asghari Jafarabadi M, Allahverdipour H, Bashirian S, Jannati A. Modeling the Underlying Predicting Factors of Tobacco Smoking among Adolescents. *Iranian J Publ Health* 2012 41(5):46-57.
6. Mirfeizi M, Toorzai ZM, Asghari-Jafarabadi M, Shoghi M, Mohammad JG, Tekmehdash AM. Examining Diagnostic Value of the Fasting Plasma Glucose in Screening Gestational Diabetes. *Iranian Journal of Diabetes and Lipid Disorders* 201110:1-5.
7. Asghari-Jafarabadi M, Hajizadeh E, Kazemnejad A, Fatemi SR. A comparative study on the prognostic impact of concurrent smoking and alcohol drinking on colon and rectal cancers: A frailty competing risks survival analysis. *Gastroenterology and Hepatology from Bed to Bench* 2010 3(1):19-26.
8. 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.