محاسبه‌ی نامیزانی و میانگین گروه

چندی پیش یکی از همکاران با بیان این که ایشان با استفاده از نتیجه‌ی برنامه‌ی پایش کیفیت خارجی بیوشیمی، به روش زیر نامیزانی (Bias)سنجش‌ها را حساب کرده‌اند:

$$\text{Bias} = \frac{x - \text{mean}}{\text{mean}} \times 100$$

(که در آن x نتیجه‌ی ازمایشگاه و mean میانگین گروه در برنامه‌ی EQAP است)، این پرسش را داشتند که ایا این روش برای حساب کردن نامیزانی درست است‌یا نه‌نابجا ندیدم که پاسخ به این پرسش را برای تالارگفتگو نیز بفرستم شاید برای همکاران جوان سودمند افتد.

اگر ما یک وزنه‌ی ۱۰۰ گرمی مطمئن را روی ترازوی الف بگذاریم و آن ترازو عدد ۱۰۰ را نشان بدهد چه خواهیم گفت؟ آیا می‌توانیم بگوئیم که ترازوی الف کاملا میزان است؛ یعنی نامیزانی آن صفر است؟اگر نتیجه‌ی یکبار سنجش همان‌وزنه با ترازوی ب و پ به ترتیب ۹۸ و ۱۰۲ باشد، ایا می‌توانیم بگوئیم که نامیزانی ترازوهای ب و پ به ترتیب -۲% و ۲% است؟

اگر مطمئن بودیم که نوسان ترازوهای ما صفر است، یعنی هرچند بار هم که یک وزنه را با آن وزن کنیم یک عدد یکسان را خواهیم دید، آنگاه می‌شد به هر سه پرسش بالا با اطمینان پاسخ "آری" داد و مطمئن بود که نامیزانی سه تراوزی بالا به ترتیب ۰%، -۲% و ۲% است.

اما اگر ترازوی ما نوسان داشته باشد انگاه دیگر نمی‌توان چنین ادعایی کرد. در صورت وجود نوسان، دیگر همه‌چیز به بخت و اقبال بسته است: شاید ترازوی الف میزان نباشد و به طور اتفاقی این بار ۱۰۰ را نمایش داده است؛ شاید ترازوی ب‌میزان باشد ویا حتاشایبه طرف بالای ۱۰۰ نامیزان باشد ولی این بار اتفاقی ۹۸ نمایش داده است؟ و ...

راه حل چیست؟راهکار حل مشکل همانطور که می‌دانیم ساده است: هر جایی که یکدستی وجود ندارد و بنا بر این نمی‌توان فقط با یک داده در باره‌ی کل قضاوت کرد، باید مجموعه‌ای مورد بررسی را خوب هم زد و سپس به طور اتفاقی "نمونه"ای را که بیش از یک عضو دارد انتخاب و بررسی کرد و از میانگین بررسی آن نمونه،در باره‌ی وضعیت کل داوری کرد.

بیش از یکی یعنی چندتا؟ پاسخ به این پرسش در خود کلمه‌ی "نمونه" است. در نظر داشتن این نکته که‌این " چندتا" بنسبت نمونه و نماینده‌ی واقعی وضعیت کلی باشند راهنمای ما خواهد بود. و به سادگی روشن است که هرچه یکدستی مجموعه‌ی مورد نظر کمتر باشد، و نیز هرچه بخواهیم برآورد ما قابل اطمینان تر باشد، باید تعداد اعضای نمونه‌ی ما بیشتر باشد. در آزمایشگاه برای این که برآورد مطئمنی داشته باشیم لازم است دست‌کم ۲۰ بار یک نمونه را بسنجیم تا بتوانیم میانگین آن را برآورد قابل قبولی از میانگین کل بدانیم.

مثال: به عنوان مثالی آزمایشگاهی، اگر ما نمونه‌ای را که مطمئن هستیم گلوکز آن ۲۰۰ است فقط یک بار بسنجیم و عدد ۱۹۲ را به دست آوریم نمی‌توانیم به سادگی و بدون در نظر گرفتن نوسان، در باره‌ی مقدار نامیزانی این روش داوری کنیم و نتیجه بگیریم که نامیزانی سنجش ما ۴- % است. تنها اگر CV سنجش ما صفر بود می‌توانستیم چنین کنیم اما چون سنجش‌های ما هنوز ارمانی نیستند و نوسان دارند نمی‌توانیم چنین کنیم.

اگر CV این روش ۵% باشد به ان معنا خواهد بود که اگرچه در حالی که دستگاه ما کاملا میزان باشد، میانگین تکرار سنجش بر روی نمونه‌ای با مقدار حقیقی ۲۰۰، دقیقا ۲۰۰ خواهد شد اما تک تک نتیجه‌ها ۲۰۰ نخواهد بود، بلکه نتیجه‌های دیده شده در اطراف میانگینشان یعنی عدد ۲۰۰ پراکنده خواهند بود طوری که ۹۵% از آن‌ها در فاصله‌ی بین ±2CV از ۲۰۰ خواهد بود:

SD = CV × Mean = 5% × 200 = 10

2 SD= 20

200 – 20 = 180

200 + 20 = 220

یعنی با چنین دستگاهی ۹۵% از نتیجه‌های تکرار سنجش نمونه‌ای با مقدار حقیقی ۲۰۰ بین 180 تا 220 خواهد بود (گستره‌ای برابر ۴۰ واحد)؛ و البته گاهی هم نتیجه‌های پایینتر از ۱۸۰ یا بالاتر از ۲۲۰ دیده خواهد شد.

اگر دستگاه ما نامیزان باشد باز هم گستره‌ی پراکندگی ۹۵% نتیجه‌ها همین ۴۰ واحد است با این تفاوت که این بار پیرامون میانگین دیگری نوسان دارند. مثلا فرض کنیم هنگام میزانکردن (کالیبراسیون) دستگاه به اشتباه به جای عدد ۱۰۰ که عدد درست کالیبراتور بوده است، عدد ۱۰۴ را وارد کرده باشیم. این یعنی این که دستگاه ما به اندازه‌ی ۴% نامیزان است. با چنین دستگاهی، اگر نمونه‌ی بالا را چندین بار بسنجیم میانگین ۲۰۸ را به دست خواهیم اورد و خواهیم دید که ۹۵% از نتیجه‌ها در فاصله‌ی ۱۸۸ تا ۲۲۸ قرار دارند. برعکس، اگر هنگام میزان‌کردن دستگاه به اشتباه عدد ۹۶ را وارد کرده باشیم و نامیزانی دستگاه ما ۴- % باشد؛ آنگاه میانگین حاصل ازتکرار این نمونه ۱۹۲ خواهد بود و ۹۵% از نتیجه‌ها در فاصله‌ی ۱۷۲ تا ۲۱۲ خواهند بود. بیایید عددهای مربوط به این سه وضعیت را مرتب کنیم:

- محدوده‌ی ۹۵% برای نامیزانی ۴%: ۱۸۸ تا ۲۲۸
- محدوده‌ی ۹۵% برای دستگاه کاملا میزان: ۱۸۰ تا ۲۲۰
- محدوده‌ی ۹۵% برای نامیزانی ۴- %: ۱۷۲ تا ۲۱۲

همانطور که می‌بینیم این سه محدوده همپوشانی زیادی دارند و انتظار عددهایی بین ۱۸۸ تا ۲۱۲ از یک بار سنجش نمونه‌ای با مقدار حقیقی ۲۰۰ برای هر سه دستگاه عاقلانه است. با این حساب اگر

یک بار سنجدیم و ۱۹۲ به دست آمد (مثل بالا)، نمی‌توانیم در باره‌ی میزان بودن/نبودن دستگاه هیچ قضاوتی بکنیم و به هیچ عنوان نمی‌توانیم مقدار نامیزانی را حساب کنیم.

به این ترتیب پاسخ به پرسش آغاز این نوشته نیز روشن می‌شود: چون ما نمونه‌های EQAP را فقط یک بار می‌سنجیم و چون روش‌های ما نوسان دارند، این روش برای حساب کردن نامیزانی درست نیست. **اگر** می‌خواهیم خودمان را با میانگین همگروه مقایسه کنیم، باید نمونه‌ی EQAP را دست‌کم ۲۰ بار بسنجیم، از آن میانگین بگیریم و سپس از رابطه‌ی زیر نامیزانی را حساب کنیم:

$$\text{Bias} = \frac{\text{mea}}{}$$

از همه‌ی این‌ها می‌رسیم به تعریف سرراست "نامیزانی" که چنین است: مقدار اختلاف بین **میانگین** حاصل از تکرار سنجش به تعداد زیاد با **یک مقدار حقیقی**.

چند نکته:

- برای حساب کردن نامیزانی، باید میانگین خودمان را با یک **مقدار حقیقی** مقایسه کنیم. این مقدار حقیقی باید با استفاده از روش‌ها و/یا مواد مرجع تعیین شده باشند و قابل ردیابی (Traceable) به روش قطعی و/یا ماده‌ی مرجع اولیه باشند. اگر ماده‌ی مورد استفاده برای ارزیابی نامیزانی دارای چنین ویژگی‌هایی باشد، آنگاه:
  - می‌توانیم روش‌هایمان را نسبت به روش‌های مرجع بین‌المللی میزان کنیم؛
  - جواب‌هایی که برای بیماران گزارش می‌کنیم قابل ردیابی به روش‌های مرجع خواهند بود،
  - جواب‌های به دست آمده از آزمایشگاه‌ها و روش‌های گوناگون قابل مقایسه با یکدیگر خواهند بود و به اصطلاح تبادل‌پذیر (Commutable) خواهند بود؛
  - و در نتیجه پزشکان می‌توانند نتیجه‌های ما را بر پایه‌ی انچه که در منابع مرجع بین‌المللی می‌خوانند تفسیر کرده و به کار برند.
- اگر خودمان را در برابر میانگین گروه بررسی کنیم و نسبت به ان میزان باشیم، فقط می‌توانیم مطمئن باشیم که "هماهنگ" با اکثریت گروه هستیم (Harmonized). ولی این که آیا میانگین گروه همان "مقدار حقیقی" است یا نه، نامشخص است.

- اگر پیش از شروع به شرکت در برنامه‌ی پایش کیفیت خارجی، ابتدا همه‌ی آزمایشگاه‌ها در برنامه‌ی "یکسان‌سازی" شرکت کنند (Standardization) و در آن برنامه، همه در برابر نمونه‌هایی با مقدار حقیقی میزان‌شوند، و پس از آن نیز چنین نمونه‌هایی با تناوب مشخص مثلا سالی یک بار برای آزمایشگاه‌ها فرستاده شود تا میزان‌بودن آن‌ها بررسی و نگهداری شود؛ آنگاه می‌توان در برنامه‌های پایش کیفیت خارجی شرکت کرد تا در فاصله‌ی بین فرستادن نمونه‌های با مقدار حقیقی، از میانگین گروه برای بررسی نامیزانی استفاده کرد. **البته به شرط آنکه * و با توجه به این که **:**

  * **به شرط آنکه** همانطور که در بالا آمد نمونه‌ی EQAP بارها سنجیده‌شوند و از نتیجه‌ها میانگین گرفته شود؛

  ** **و با توجه به این که** نامیزانی بررسی شده فقط مربوط محدوده‌ی نزدیک به سطح آن ماده‌ی کنترل است و نمی‌توان آن را به همه‌ی گستره‌ی سنجش تعمیم داد. مثلا اگر میانگین گروه EQAP برای گلوکز برابر ۲۴۰ باشد و میانگین تکرار سنجش ما برروی آن ۲۴۲ بشود، فقط می‌توانیم نسبت به‌مقدار نامیزانی (۰٫۸%) در محدوده‌ی جواب‌های بالا مطمئن باشیم و نمی‌توانیم با استناد به آن در باره‌ی نامیزانی این روش در سنجش نمونه‌های دارای گلوکز طبیعی یا پایین قضاوت کنیم.

- برای دانستن نامیزانی در سراسر گستره‌ی سنجش، باید دستکم در یک فاصله‌ی زمانی ۵ روزه، روزانه حدود ۱۰ تا از نمونه‌های تازه‌ی بیماران را انتخاب کنیم و آن‌ها را با روش مرجع و روش خودمان به فاصله‌ی زمانی کمی از یکدیگر بسنجیم. جمع تعداد نمونه‌ها باید دست‌کم ۴۰ تا بشود و همچنین طوری انتخاب شوند که در سراسر گستره‌ی سنجش از مقادیر غیرطبیعی پایین تا طبیعی و تا مقادیر غیرطبیعی بالا گسترده باشند. سپس همبستگی نتیجه‌های حاصل از دو روش را با استفاده از محاسبه‌ی رگرسیون به دست آوریم. چنانچه ضریب همبستگی بیش از ۰٫۷۵ بود از معادله‌ی خط همبستگی برای حساب کردن نامیزانی در سطوح تصمیم‌گیری بالینی استفاده کنیم. اگر همبستگی مناسب نبود باید برای هر سطح تصمیم‌گیری بالینی جداگانه حدود ۲۰ نمونه‌ی نزدیک به آن سطح را بسنجیم و چنانچه آزمون t وجود اختلاف را تایید کرد آنگاه با استفاده از اختلاف میانگین‌های دو روش، نامیزانی هر سطح را حساب کنیم.

- به عنوان راهکاری عملی، اگر بنا داریم که نامیزانی یک یا چند تا از روش‌هایمان را نسبت به گروه‌مان بسنجیم، می‌توانیم سنجش آن آنالیت‌ها را روی نمونه‌ی EQAP بیش از یک بار انجام دهیم؛ مثلا اگر می‌خواهیم نامیزانی کلسترول را بررسی کنیم و روش ما CV کوچکی دارد، مثلا ۱%، می‌توانیم با ۱۰ بار سنجش کلسترول، میانگین قابل قبولی داشته و باشیم و از اختلاف آن با میانگین گروه برآورد قابل قبولی از نامیزانی خودمان داشته باشیم.

- با ترکیب نتیجه‌ی یک بار سنجش ماده‌ی کنترل EQAP و CV روش، فقط می‌توانیم گستره‌ای را که نامیزانی ما با یک احتمال معین در آن قرار دارد تعیین کنیم.

مثلا اگر نتیجه‌ی یک بار سنجش سدیم نمونه‌ی EQAP در آزمایشگاه ما ۱۵۰ شود و CV روش ما ۱۰% باشد آنگاه می‌توانیم با احتمال ۹۵% مطمئن باشیم که اگر این نمونه را به تعداد زیاد تکرار می‌کردیم میانگین سنجش ما جایی در گستره‌ی 2CV قرار می‌گرفت:

SD = CV ×X = 10% ×150 = 15

2 SD=30

150 – 30 = 120

150 + 30 = 180

با فرض این که میانگین گروه ۱۶۰ شده است؛ اگر میانگین ما کمترین مقدار یعنی ۱۲۰ باشد آنگاه نامیزانی ما برابر -۲۵% خواهد شد و اگر میانگین ما بیشترین مقدار یعنی ۱۸۰ باشد نامیزانی ما ۱۲٫۵% خواهد بود. یعنی با یک بار سنجش نمونه می‌توانیم چنین نتیجه بگیریم که نامیزانی ما برای این روش به احتمال ۹۵% مقداری است بین -۲۵% تا ۱۲٫۵%.

اگر روش بهتری داشتیم که CV آن ۵% بود و همین نتیجه‌ی ۱۵۰ را در مقابل نتیجه‌ی ۱۶۰ برای میانگین گروه به دست آورده بودیم، آنگاه می‌توانستیم چنین نتیجه بگیریم که نامیزانی ما برای این روش به احتمال ۹۵% مقداری است بین -۱۵٫۶% تا ۳٫۱%. بسته‌ترشدن بگستره به دلیل کوچک بودن CV است.

اگر CV روش ما ۱% باشد و به جای یک بار، نمونه را ۱۰ بار بسنجیم و میانگین بگیریم بگستره‌ی ما بسیار کوچکتر می‌شود چون با n بار تکرار و میانگین گرفتن، CV روش به رادیکال n تقسیم می‌شود؛ در مورد این مثال CV برابر ۱% به رادیکال ۱۰ تقسیم می‌شود و می‌شود ۰٫۳۲%. چنانچه میانگین ۱۰ بار سنجش بشود ۱۴۸ می‌توانیم چنین نتیجه بگیریم که نامیزانی ما برای این روش به‌احتمال ۹۵% مقداری است بین -۸٫۱% تا -۶٫۹%. همانطور که می‌بینیم این بار به دلیل این که هم CV روش کم بود و هم از میانگین ۱۰ بار سنجش استفاده کردیم، می‌توانیم اطمینان بالایی داشته باشیم که روش بین ۷ تا ۸% ما به طرف پایین نامیزان است و باید برای میزان کردن آن اقدام کنیم.

- **نکته‌ی آخر** این که: برای این که یک برنامه‌ی پایش کیفیت خارجی بتواند در ارزیابی عملکرد ما کمک کند باید:
  - در هر بار بیش از یک نمونه بفرستد؛ مثلا همانندCLIA هر بار ۵سطح بفرستد؛
  - به جای استفاده از انحراف معیار گروه برای قضاوت در باری چگونگی کارکرد آزمایشگاه‌ها، از مرزهای پذیرشی که بر اساس خطای کل مجاز تعیین شده‌اند استفاده کند؛ و
  - دستکم هر چند گاه یکبار و دستکم برای انالیت‌های مهمی مانند A1C، به جای نمونه‌ی کنترل تجاری نمونه‌ی تازه‌ی بیمار که با روشهای مرجع قابل ردیابی به

روش‌ها مواد مرجع بین‌المللی تعیین مقدار شده‌اند بفرستند تا بتوان کارکرد همه‌ی روش‌ها و کیت‌ها را در مقابل یک مرجع معتبر ارزیابی کرد.

مثلا کراتینین سرم تازه‌ی بیمار را با روش مرجع IDMS تعیین کند (فرض کنیم نتیجه‌ی سنجش، یعنی مقدار حقیقی شبود ۳)؛ بعد ان نمونه‌ها را به شکل فریز شده و بدون هیچگونه پردازش دیگری به سرعت به ازماشگاهها برساند؛ و بعد مثلا با ملاک قرار دادن TEa جدول CLIA برای کراتینین که ۱۵% برای مقادیر بالای ۲ است نتیجه‌های بین ۲٫۵۵ تا ۳٫۴۵ را پذیرفته بداند؛ و برای پذیرفتن عملکرد آزمایشگاه هیچ کاری به این که دستگاه و یا کیت ما چه و چه است، و میانگین گروه و CV گروه چند است نداشته باشد.

با احترام،

حسن بیات