

"کنترل کیفی": با اسم و رسم اما...!

حسن بیات - آزمایشگاه سینا (قائم شهر)

پرسش: باتوجه به اتوالیزرهای امروزی با دقت بسیار بالا، بسیار دیده می شود که نتایج سرم کنترل در طول ماه بسیار نزدیک است و بنا بر این، انحراف معیار به دست آمده عددی کوچک و در نتیجه محدوده منحنی رسم شده برای ماه بعد بسیار باریک می شود. حال در ماه جدید اگر سرم کنترل روزانه اندکی بالا/پائین شود به غلط در خارج از محدوده 2SD یا 3SD قرار می گیرد. برای اصلاح منحنی چه باید کرد.

(**هشدار و پوزش!** وقتی من و برادر کوچکترم دانش آموز بودیم، برادرم هر وقت سوال درسی داشت آخر به این نتیجه می رسید که: "آدم وقتی از تو یه سوال می پرسه، اینقدر توضیح می دی که طرف پشیمون می شه!" شاید خیلی هم گناه من نبود؛ پرسش های اون طوری بودن که باید کلی توضیح شامل مقدمه و مثال و تعریف ردیف می کردی تا دست آخر بشه اون پرسش رو پاسخ داد. پرسش بالا هم از اون دسته پرسش هاس؛ یعنی با پوزش خواهی حالا هم قراره اینقدر توضیح بدم که ...)

بیا با یک مثال شروع کنیم: اگر ما وارد یک جاده شویم و در ابتدای جاده سوال کنیم که محدوده سرعت در این جاده چقدر است و کسی جواب بدهد:

"۲۰ کیلومتر ابتدای جاده را رانندگی کنید و سر هر ۱ کیلومتر سرعت خود را به عنوان "سرعت کنترل" یادداشت کنید. بعد میانگین و نوسان سرعت خود را از روی این ۲۰ نتیجه‌ی کنترل حساب کنید و سپس محدوده‌ی نوسان سرعت خود را تعیین کنید. محدوده‌ی به دست آمده در ۲۰ کیلومتر اول، محدوده‌ی مجاز حرکت در این جاده است"،

بی‌گمان ما این گفته را شوخی در نظر خواهیم گرفت. ولی متأسفانه در آزمایشگاه‌ها چنین روش شوخی‌مانندی، روش پایش کیفیت ما را تشکیل می‌دهد.

مثالمان را ادامه دهیم. فرض کنیم که ۳ اتوبوس (الف، ب و پ) به این روش شوخی‌مانند برای خودشان محدوده‌ی سرعت تعیین می‌کنند. میانگین و انحراف معیار رانند این سه اتوبوس در زیر آمده است:

- الف: 80 ± 2

- ب: 140 ± 20

- پ: 77 ± 3

بر پایه‌ی این نتیجه‌ها و با حساب کردن $\pm 2SD$ ، محدوده‌های مجاز حرکت سه اتوبوس به ترتیب چنین می‌شود:

- الف: ۷۶ تا ۸۴
- ب: ۱۰۰ تا ۱۸۰
- پ: ۷۱ تا ۸۳

تا همینجای کار، و بدون در نظر گرفتن محدوده‌ی مجاز واقعی جاده:

- مسافران کدام اتوبوس در آسایش و امنیت کمتری هستند؟ و سلامتی آنها بیشتر در خطر است؟
- و در ادامه‌ی راه؛ رانندگی برای کدام راننده ساده‌تر است؟ و کمتر نسبت به محدوده‌ی مجاز خودش هشدار خطا می‌گیرد؟

در هر دو مورد پاسخ یکی است: اتوبوس ب.

اتوبوس ب به لطف رانندگی پرنوسان در ۲۰ کیلومتر ابتدایی، چنان محدوده‌ی به اصطلاح مجازی برای خود تعیین کرده است که در ادامه‌ی جاده هر چقدر هم بد براند همچنان رانندگی‌اش قابل قبول است. در پاسخ هر پلیس و مسافری هم که به سرعت بالا یا ترمزهای شدید و گازهای نابجای وی خرده بگیرد، نمودار "کنترل کیفی" سرعت را نمایش خواهد داد و دلیل خواهد آورد که رانندگی وی "در کنترل" است (و حتماً پلیس و مسافران نیز خواهند پذیرفت!).

حال بیابید وارد دنیای رانندگی "عقلانه" بشویم؛ جایی که اگر راننده‌ای برای پلیس یا مسافرانچنین دلیل‌هایی بیاورد و چنین نمودار کنترل کیفی‌یی نشان بدهد، در پاسخ، خیلی ساده و سراسر سراسر است به وی خواهند گفت: "محدوده‌ی مجاز سرعت در این جاده (مثلاً) ۷۰ تا ۹۰ کیلومتر است، والسلام!"; و اگر راننده‌ای بگوید "من برای خودم خطای مجاز تعیین کرده‌ام"، پاسخ خواهد شنید که "خطای مجاز جاده‌های گوناگون را سازمان‌های کارشناسی تعیین و بر روی تابلوها می‌نویسند، و رانندگان باید در هر جاده‌ای آن محدوده‌های مجاز را بشناسند و رعایت کنند".

خب، حالا داستان فرق می‌کند. دیگر در فهرست ارزشیابی رانندگان، نمی‌پرسند "آیا راننده خطای مجاز را تعیین می‌کند؟ (سوال شماره‌ی ۱۲۰ چک‌لیست ارزیابی آزمایشگاه‌ها)"; بلکه کارشناسان محدوده‌ی مجاز سرعت در هر جاده را تعیین و روی تابلو می‌نویسند و بازرسان می‌پرسند "آیا راننده محدوده‌ی مجاز سرعت را رعایت می‌کند؟". یعنی دیگر هر راننده با محدوده‌ی شخصی "من درآوردی" خودش مقایسه نمی‌شود، بلکه شیوه‌ی رانندگی همه‌ی رانندگان با محدوده‌ای که به وسیله‌ی کارشناسان، محدوده‌ی امن و مناسب تشخیص داده شده، و محدوده‌ی مجاز نامیده شده است، مقایسه خواهد شد.

با این حساب، پس بررسی رانندگی در ۲۰ کیلومتر اول به چه کاری می‌آید؟

به دو کار:

- ۱) بررسی این که آیا اصلاً یک راننده می‌تواند از پس ویژگی مورد نظر رانندگی در یک جاده برآید یا نه؟ و به اصطلاح "شایستگی" دارد یا نه؟

۲) و اگر راننده‌ای شایسته شناخته شد؛ بسته به ویژگی‌های رانندگی وی، برای او برنامه‌ی "پایش کیفیت"ی که مناسب رانندگی اوست تعیین می‌شود.

حالا مانند پلیس‌های دنیای واقعی و به روش جدی (نه شوخی‌مانند)، ۳ مورد رانندگی بالا را بررسی کنیم. اما پیش از این کار باید به یک پرسش مهم دیگر از سوی رانندگان پاسخ دهیم و آن این است:

- آقا اجازه! قبول که باید در محدوده‌ی ۷۰ تا ۹۰ کیلومتر رانندگی کنیم؛ اما آیا هیچ گذشته‌ی در کار نیست؟ آگه فقط یک بار از محدوده بزنیم بیرون رد می‌شیم؟

پاسخ این است که در حال حاضر و به طور معمول، اگر تا ۵٪ مواقع سرعت شما بیرون از محدوده‌ی مجاز باشد قابل قبول است؛ یعنی دست کم ۹۵٪ مواقع، سرعت شما باید درون محدوده‌ی مجاز باشد.

پس، دو تا معیار داریم:

۱) محدوده‌ی مجاز؛ که بسته به نوع جاده فرق می‌کند،

۲) حداکثر دفعات مجاز خروج از محدوده‌ی مجاز؛ که فعلا ۵٪ در نظر گرفته می‌شود.

محدوده‌ی مجاز که معلوم است؛ سازمان‌های کارشناسی در هر کشوری بسته به شرایط، خطای مجاز برای جاده‌های گوناگون را تعیین کرده و روی تابلوها نوشته‌اند (مثلا CLIA88 برای آزمایشگاه‌های امریکا، RILIBAK در آلمان، و RCPA در استرالیا).

اما از کجا بفهمیم که چند درصد مواقع در محدوده‌ی مجاز هستیم؟ یا باید بنشینیم و تعداد خطاها را یکی/یکی بشماریم، یا اگر خوش‌شانس باشیم (که از قضا هستیم) و توزیع فراوانی سرعت اتوبوس‌ها "گاسی" شکل (توزیع طبیعی سابق!) باشد، کار ما ساده می‌شود و خیلی ساده با به دست آوردن انحراف معیار و میانگین فراوانی، می‌توانیم این بررسی را انجام دهیم؛ چون ۹۵٪ از توزیع گاسی شکل در فاصله‌ی تقریبی ± 2 انحراف معیار از میانگین قرار دارد.

برگردیم سر وقت اتوبوس‌ها. دیدیم که می‌خواهیم همه‌ی این بررسی‌ها را برای دو کار می‌خواهیم: بررسی شایستگی و تهیه‌ی برنامه‌ی مناسب پایش کیفیت. برویم این دو کار را برای اتوبوس‌ها انجام بدهیم:

❖ ارزشیابی برای بررسی شایستگی:

سرعت حرکت اتوبوس الف در ۹۵٪ از مواقع برابر است با ۷۶ تا ۸۴، که کاملا درون محدوده‌ی مجاز ۷۰ تا ۹۰ است.

اتوبوس ب که از بنیان رد است، چون در ۹۵٪ مواقع سرعتی بین ۱۰۰ تا ۱۸۰ دارد که کاملا بیرون از محدوده‌ی مجاز است.

اتوبوس پ با محدوده‌ی ۹۵٪ برابر ۷۱ تا ۸۳ نیز پذیرفته است و اجازه حمل مسافرها (بخوانیم انجام آزمایش‌ها) را دارد. پس اتوکیت‌های الف و پ شایستگی دارند.

نکته‌ی مهم این است که در مدت ارزشیابی، اتوبوس‌ها نباید مسافر جابجا کنند؛ یعنی نباید تا پایان ارزشیابی و تایید شایستگی، و نیز تهیه‌برنامه‌ی پایش کیفیت مناسب، نمونه‌های بیماران با آن روش آزمایش شود. (ارزشیابی روش برای بررسی شایستگی همون چیزیه که به فارس می‌شه متد ولیدیشن!).

❖ تهیه‌ی برنامه‌ی پایش کیفیت مناسب:

گیر قضیه اینجاست. قرار نیست که همه به یک چوب رانده شوند. مثلا اگر قرار باشد ما ۵ کودک را که در حال بازی هستند "بپاییم"، در حالی که می‌دانیم یکی از آن‌ها بسیار شیطان است، سه تای آن‌ها معمولی هستند و پنجمی بسیار آرام اس؛ آیا برنامه‌ی پایش هر ۵ تای آن‌ها را یکسان در نظر خواهیم گرفت؟ و همه را با سختگیرانه‌ترین معیار و بیشترین زحمت، که فقط و فقط مناسب کودک بسیار شیطان است، خواهیم پایید. یا بر عکس، خیلی راحت و بی‌خیال به هر ۵ کودک فقط هر ساعت یک بار نگاهی خواهیم انداخت (برنامه‌ای که فقط و فقط مناسب مراقبت از کودک بسیار آرام است). اگرچه مدیر یک مهد کودک هرگز چنین نخواهد کرد، و به مراقب کودکان برای پاییدن هر کودک سفارش‌هایی مناسب آن کودک خواهد کرد، اما ما در آزمایشگاه این اشتباه را می‌کنیم و همه‌ی روش‌های سنجش را با سختگیرانه‌ترین معیار که همان "معیار چندقانونی وستگارد" است (همان $1:2S/1:3S/2:2S/R4S/4:1S/8:X$ معروف) پایش می‌کنیم.

ببینیم چرا این کار نادرست است؟ یادمان هم نرود که ۲ تا از راننده‌ها که شایسته شناخته شده‌اند هنوز منتظر هستند که برنامه‌ی QC آن‌ها را تعیین کنیم تا بروند سراغ آزمایش نمونه‌های بیماران (ببخشید حمل مسافران).

اگرچه هر دوی این رانندگان شایسته تشخیص داده شده‌اند، ولی وقتی به رانندگی الف نگاه می‌کنیم راننده‌ای را می‌بینیم که میانگین سرعت او دقیقا "سرعت هدف" یعنی ۸۰ است، و علاوه بر این، از چنان نوسان کمی در حرکت برخوردار است که از هر دوسو تا مرزهای مجاز فاصله‌ی زیادی دارد؛ یعنی حاشیه‌ی امنیت خوبی دارد. در طرف مرز پایین، در حالی که حداقل سرعت مجاز ۷۰ است، کمترین سرعت او ۷۶ است که هنوز خیلی با مرز ۷۰ فاصله دارد، و در طرف مرز بالا، در حالی که حداکثر سرعت مجاز ۹۰ است، بیشترین سرعت او ۸۴ است که باز هم خیلی با مرز ۹۰ فاصله دارد.

اما در مورد روش پ چنین نیست. روش پ اگرچه پذیرفته شده است، اما حاشیه‌ی امنیت آن تا یکی از مرزهای مجاز بسیار کم است و با اندکی غفلت یا اشکال در اتوبوس، سرعت وی از مرز پایین (۷۰) کمتر خواهد شد و مسافران (بیماران) را به خطر خواهد انداخت.

خیلی سخت نیست که به این نتیجه برسیم که برای پایش رانندگی الف، برنامه‌ای با حساسیت کمتر نیاز است در حالی که پایش رانندگی ب، برنامه‌ای پرزحمت‌تر و سختگیرانه‌تر نیاز دارد. راه درست هم همین است و بر پایه‌ی آموزش‌های پروفیسور وستگارد، برای روش‌های "خیلی خوب" که نامیزانی (همان "بایاس" فارسی!) و انحراف معیار کوچکی دارند باید از قانون‌های آسان‌تر، مثلا فقط $1:4S$ ، استفاده کرد.

یعنی نموداری رسم کرد که در دو سوی میانگین فقط مرزهای 4SD مشخص شده‌اند. در هر دور (دور به فارسی می‌شود "ران Run") ۲ سطح (و گاهی حتی یک سطح) از کنترل را سنجید، و تا زمانی که نتیجه‌های کنترل درون مرزهای 4S است مطمئن بود که مشکلی در کار نیست و نتیجه‌ی بیماران را گزارش کرد. برای روش‌های "خوب" که نامیزانی و/یا انحراف معیار بزرگتری دارند، از قانون‌های با سختگیری متوسط، مثلاً فقط 1:3S، استفاده کرد، و برای روش‌هایی که خیلی در معرض خطا هستند از معیارهای چندقانونی و افزایش تعداد سنجش کنترل در هر دور به ۴، ۶ و گاهی ۸، استفاده کرد.

گاهی هم شاید یک روش چنان کارکرد ضعیفی داشته باشد که حتی سختگیرانه‌ترین معیارهای "پایش کیفیت آماری" هم برایش کافی نباشد و مجبور باشیم برای پایش کیفیت آن روش، در کنار روش‌های آماری از روش‌های غیر آماری هم استفاده کنیم. گاهی هم ممکن است کارکرد روشی از این اندازه هم ضعیفتر باشد و در آن صورت، هیچ‌یک از روش‌های آماری، حتی با گذاشتن ۶ یا ۸ کنترل در هر دور به هیچ کاری نیاید و از هیچ SD و نموداری هیچ کاری برنیاید و مجبور باشیم برای پایش کیفیت، فقط به روش‌های غیر آماری، و البته پرزحمت، تکیه کنیم.

تازه! معیارهای چندقانونی هم فقط همین معیار یاد شده نیست. مثلاً ترکیب 1:3S/2:2S هم یک معیار چند قانونی است. واقعیت این است که معیار چندقانونی بلن‌بالای "1:2S/1:3S/2:2S/R4S/4:1S/8:X"، یکی از سختگیرانه‌ترین معیارها است و فقط باید برای رانندگی‌های خیلی ضعیف که حاشیه امنیت بسیار کمی دارند به کار گرفته شود.

اشکال دیگری هم هست و آن این است که وقتی قرار است در معیار چندقانونی ما، قانونی باشد که لازم باشد نتیجه‌های کنترل را در بیش از یک بررسی کنیم، آنگاه ممکن است گاهی مجبور شویم جواب بیماران را تا چند دور گزارش نکنیم تا تکلیف آن قانون مشخص شود. مثلاً اگر در معیار ماقانون 4:1S باشد، باید هر بار صبر کنیم تا دور بعدی آزمایش انجام شود تا مطمئن شویم که نتیجه‌ی کنترل در همان طرف نتیجه‌ی دور پیشین نیست، و اگر بود باید تا انجام دور سوم، و گاهی هم چهارم، صبر کنیم و جواب بیماران را گزارش نکنیم تا تکلیف معلوم شود. با حساب این که معمولاً فقط روزی یک بار کنترل می‌گذاریم، این یعنی باید دست کم ۲ روز و گاهی ۳ و حتی ۴ روز صبر کنیم! (با قانون 8X که برو تا هفته‌ای دیگر!).

واقعیت این است که معیار "1:2S/1:3S/2:2S/R4S/4:1S/8:X"، بیشتر به عنوان یک مثال که دربرگیرنده‌ی همه‌ی قانون‌ها باشد به وسیله وستگارد و ۳ نفر دیگر در دهه‌ی هشتاد میلادی برای توضیح چگونگی استفاده از معیارهای چندقانونی در مقاله‌ای در مجله‌ی Clinical Chemistry مطرح شد و پس از آن هم بیشتر کاربرد آموزشی دارد (البته در آن مقاله ته خط این معیار 10:X بود نه 8:X). اگر هم در جایی، پس از تعیین ویژگی‌های اجرایی آزمایش و مقایسه‌ی آن با محدوده‌ی مجاز آن آزمایش، و با استفاده از وسیله‌هایی مانند نمودارهای OPSpec، به این نتیجه برسیم که باید از معیار چندقانونی یاد شده استفاده کنیم، پیش از استفاده باید ببینیم که آیا از نظر زمان جوابدهی و پایداری نمونه و دیگر شرایط، می‌توان گاهی تا ۸ دور صبر کرد تا تکلیف کنترل معلوم شود یا نه، و آنگاه آن را به کار بست. و

اگر نمی‌توان اینقدر صبر کرد باید از خیر آن روش گذشت و گفت "اگرچه این اتوبوس در امتحان ارزشیابی پذیرفته شد اما کنترل آن چنان پردردسر است که بهتر است آن را کنار بگذاریم و دستگاه/کیت بهتری را که پایش کیفیت آن آسان‌تر است و در همان دور تکلیف آن معلوم می‌شود به کار گیریم".

نتیجه‌گیری

(نتیجه‌گیری یعنی این که می‌دونم خیلی پشیمون شده‌اید و دارم سعی کنم تا بیشتر پشیمون نشده‌اید سر و تمش رو هم بیارم)

خب، اشکال کار ما کجاست که با دستگاه‌های خوب، هشدار خطا دریافت می‌کنیم؟

- اشکال اینجاست که نمی‌دانیم محدوده‌ی مجاز حرکت در جاده‌های سنجش آنالیت‌های گوناگون چقدر است؟ محدوده مجاز خطا در سنجش A1C چقدر است؟ خطایمجاز سنجش کلسترول؟ خطای مجاز شمارش WBC؟ خطای مجاز PT؟ ...

- اشکال اینجاست که بدون این که محدوده‌ی مجاز حرکت در جاده تعیین شده باشد، کارکرد هر کیت در ادامه‌ی راه را با ۲۰ روز اول خودش می‌سنجیم. اگر در سنجش گلوکز، انحراف معیار برای دو کیت به ترتیب ۱ و ۱۰ باشد، کیت اول در حالی که انحراف معیار خیلی بهتری دارد، به دلیل روش اشتباه ما در پایش کیفیت، مداوم هشدار خطا می‌گیرد؛ در حالی که کیت دوم که اصلاً نباید با آن کار می‌شد با خیال راحت به آزمایش ادامه می‌دهد و هیچ‌وقت هم نتیجه‌ی کنترل بیرون از مرزهای 2SD ندارد.

اگر فرض کنیم که ما هم خطای مجاز CLIA برای سنجش گلوکز یعنی ۱۰٪ را قبول داریم، روش دوم حتماً اگر هیچ نامیزانی‌ای هم نداشته باشد، با این حال هم فقط ۶۸٪ از جواب‌هایی که تولید می‌کند در فاصله‌ی ۱۰٪± از مقدار واقعی است (یعنی درست است) در حالی که بناست دست کم ۹۵٪ از جواب‌هایی که تولید می‌کند در این فاصله باشد. یعنی روشی که در ارزشیابی باید رد می‌شد و نباید نمونه‌ی هیچ بیماری با آن آزمایش می‌شد، با به‌به/چه‌چه به کار ادامه می‌دهد، جواب‌های کنترل کیفی‌اش هم همیشه "در کنترل" است؛ اگرچه سلامت بیماران بیرون از کنترل است.

اما روش اول که حتماً با وجود نامیزانی ۲٪، باز هم به دلیل نوسان کم، روش بسیار خوبی است و باید با معیار آسان 1:3.5S کنترل شود، به دلیل انتخاب معیار اشتباه "چند قانونی" بی‌جهت و به فراوانی هشدار خطا می‌گیرد.

- اشکال کار اینجاست که بدون در نظر گرفتن خطای مجاز، بدون ارزشیابی روش‌ها، و بدون تعیین برنامه‌ی پایش کیفیت مناسب برای هر روش، فقط و فقط با رسم نمودار "کنترل

کیفی"، همانطور که پروفیسور وستگارد می‌گوید، فکر می‌کنیم که معجزه می‌شود و کار ما با کیفیت انجام می‌شود؛ چیزی که وستگارد آن را "گمان کیفیت و نه تضمین کیفیت" می‌نامد.

- اشکال کار اینجاست که حتا اسمی که ما به کار می‌بریم نادرست است: "کنترل کیفی". در آزمایشگاه هم، مثل هر کار دیگری، ما می‌خواهیم "کار خوب" یعنی "باکیفیت" انجام دهیم. برای این، ابتدا باید ویژگی کیفیت را تعیین کنیم (یعنی خطای مجاز و نیز حداکثر دفعاتی که مجاز هستیم بیرون از این محدوده باشیم)، سپس روش را ارزشیابی کنیم تا معلوم شود که آیا آن روش می‌تواند این ویژگی‌ها را تامین کند و کار با کیفیت ارائه دهد، و سپس برای "حفظ کیفیت"، برنامه‌ی "پایش کیفیت" یا "کنترل کیفیت" مناسب تعیین کنیم. اسم درست "کنترل کیفیت" است (یا بهتر "پایش کیفیت") و نه کنترل کیفی. شاید برخی خرده بگیرند که فرقی ندارد؛ من هم بر این باورم که خیلی وقت‌ها فرقی ندارد و نباید خیلی به اسم اهمیت داد، ولی در همان حال بر این باورم که در برخی موارد مانند این مورد، اسم‌گذاری نادرست نشانه‌ی اشتباه در فهمیدن یک فرآیند است، و بی‌گمان اشتباه در فهمیدن یک فرآیند به اشتباه در اجرای آن می‌انجامد؛ و نیز بر این باورم که اسم درست می‌تواند آموزش دهنده باشد. اگر فرآیند پایش کیفیت درست فهمیده می‌شد آنگاه درست نامگذاری، و مهمتر از آن درست اجرا، می‌شد. اگر ما به یک محصل دبیرستانی بگوییم "پایش کیفیت"، خواهد دانست که پیش نیاز این پاییدن، ایجاد کیفیت است؛ ابتدا باید کیفیت را ایجاد بکنیم، سپس آن را کنترل کنیم که از دست نرود؛ رسم نمودار و باقی کارها، آن هم به شرطی که کاملاً درست و بر پایه‌ی اصول علمی باشد، تنها و تنها برای حفظ کیفیتی است که پیشتر ایجاد شده است و نه برای ایجاد کیفیت.

- اشکال اینجاست که هنوز هم وقتی قرار است از نمودار و SD این جور چیزها صحبت شود، می‌گوییم "کنترل کیفی بیوشیمی"، بدون توجه به اینکه این کارها در واقع استفاده از روش‌های آماری در پایش کیفیت است، یعنی "پایش کیفیت آماری"؛ و چه بیوشیمی باشد چه غیر بیوشیمی، هر جا که با عدد و رقم سر و کار داریم پایش کیفیت آماری به کار می‌آید.

- اشکال اینجاست که (همانطور در پرسش آغاز نوشته دیده می‌شود) فکر می‌کنیم هر ماه باید میانگین و انحراف معیار ماه پیش را برای رسم نمودار ماه بعد به کار بریم.

در حالی که باید در ابتدای کار، خطای مجاز یک آزمایش، یعنی کیفیت موردنظر، را که سازمان‌های کارشناسی صاحب‌نظر تعیین کرده‌اند بشناسیم، با تلاش فراوان سعی کنیم کیفیت موردنظر را ایجاد کنیم، با دقت و وسواس فراوان، کیفیت ایجاد شده را ارزشیابی کنیم، و اگر پذیرفته شد برای پایش و نگهداری این کیفیت "پذیرفته شده"، نمودار بکشیم و بگوییم از این پس باید پیرامون این میانگین و درون این محدوده‌ها حرکت کنیم تا مطمئن باشیم که کیفیتی را که در آغاز کار "ایجاد، بررسی و تایید" کرده بودیم حفظ کرده‌ایم. اگر در آینده توانستیم کیفیت کار را بهبود بخشیم و بهتر کنیم (مثلاً دستگاه را تعمیر کردیم، مهارت کاربران بیشتر شد، کیفیت آب مصرفی بهتر شد و ...) و نمودار پایش کیفیت ما چنین فرضی را مطرح کرد، آنگاه باید دوباره

روش را ارزشیابی کنیم و برنامه‌ی پایش جدیدی که آسان‌تر خواهد بود با نمودار جدید تعیین کنیم؛ و اگر به دلیلی کار ما بدتر شد (دستگاه کهنه شد، کیفیت کیت خراب شد و...) و نمودار پایش کیفیت ما چنین فرضی را مطرح کرد، آنگاه باید دوباره ارزشیابی کنیم که ببینیم آیا هنوز مجاز به استفاده از آن روش هستیم یا نه؟ و اگر روش ماهنوز قابل قبول است آنگاه برایش برنامه‌ی جدید، که البته این بار سختگیرانه‌تر خواهد بود، با نمودار جدید تعیین کنیم.

سرجمع، اشکال کار اینجاست که اگرچه روش "کنترل کیفی" ما، خیلی "اسم و رسم" دارد، اما واقعیت این است که به این شکل که ما کار می‌کنیم هم "اسم"ش نادرست است هم "رسم"ش!

ما دو اشتباه را با هم انجام داده‌ایم: هم کار را از وسط شروع کرده‌ایم: بدون تعیین ویژگی‌های کیفیت و بدون ارزشیابی روش‌ها؛ و هم به شکل اشتباه انجام می‌دهیم: بدون برنامه‌ریزی پایش کیفیت.

راه حل هم آن است که برگردیم و از سرخط شروع کنیم و به شکل درست انجام بدهیم؛ همان چیزی که پروفیسور وستگارد آن را "اجرای پایش کیفیت درست به شکل درست" می‌نامد:

Doing right QC right

(روز اول کنگره‌ی ارتقای کیفیت سال ۹۱، در پایان سخنرانی یکی از مسئولان آزمایشگاه مرجع سلامت با عنوان "مستندسازی؛ چند گام به جلو"، ضمن تشکر از تلاش‌های انجام شده گفتیم لازم است در "پایش کیفیت" چند گام به عقب برگردیم و دوباره شروع کنیم.)

اما پاسخ به سوال

خب، با توجه به این که:

- روش‌های خوب امروزی نباید بی‌دلیل هشدار بگیرند،
- زحمت و هزینه‌ی پایش آن‌ها کمتر است: تعداد سنجش ماده‌ی کنترل برای آن‌ها کمتر است و نیز قانون‌های آسان‌تری باید به کار بست، و به کارهای غیرآماری هم نیاز ندارند،
- توانایی خطایابی برنامه‌های پایش آن‌ها بیشتر است: خطاهای کمتری از چشم این برنامه‌ها دور می‌ماند،
- و نیز هشدارهای خطای آن برنامه‌ها قابل اعتمادتر است: هشدار دروغین کمتری دارند،

برای اصلاح منحنی چه باید کرد؟

برای اصلاح منحنی هیچ کاری نباید کرد! چه روش خوب امروزی به کار بگیریم و چه روش‌های نامیزان‌تر و پرنوسان‌تر قدیمی‌تر، باید از ابتدا بر پایه‌ی کیفیت مورد نظر (یعنی خطای مجاز و حداکثر دفعات مجاز خروج از محدوده‌ی مجاز) و نیز ویژگی‌های روش سنجش (یعنی نامیزانی و انحراف

معیار)، و به کار بستن روش درست، برنامه‌ی پایش کیفیتمناسب را انتخاب کرد طوری کهبر پایه‌ی اطلاعات به دست‌آمده از آن برنامه، اشکال‌های سنجش نمونه‌های بیمار انماین شود و روش آزمایش را اصلاح کرد؛ نه این که برعکس، برنامه‌ای برای پایش کیفیت داشته باشیم که بر پایه‌ی اطلاعات به دست آمده از سنجش نمونه‌های بیمار ان، اشکال آن برنامه نمایان شود و ما نمودار را اصلاح کنیم!

در پایان در راستای روش درست تهیه‌ی برنامه‌ی پایش کیفیت، برنامه‌ای را پیوست کرده‌ام که برای انتخاب روش مناسب پایش کیفیت به کار می‌آید. این برنامه بر پایه‌ی نمودارهای OPSpec و ستگارد طراحی شده است. در زیر نیز توضیح کوتاهی در باره‌ی آن برنامه آمده است.

گزینشگر معیار پایش کیفیت

پروفیسور وستگارد برای انتخاب معیار پایش کیفیت، یک سری نمودار به نام "نمودارهای ویژگی‌های کارکردی (OPSPecs Charts) طراحی کرده است. روی هم ۸ نمودار هستند:

- 90% AQA برای 2 سطح کنترل
- 50% AQA برای 2 سطح کنترل
- 90% AQA برای 4 سطح کنترل
- 50% AQA برای 4 سطح کنترل

- 90% AQA برای 3 سطح کنترل
- 50% AQA برای 3 سطح کنترل
- 90% AQA برای 6 سطح کنترل
- 50% AQA برای 6 سطح کنترل

برای روش‌های قابل قبول، بسته به ویژگی‌های اجرایی روش و خطای مجاز روش، از نمودارهای بالا می‌توان استفاده کرد و معیار مناسب برای پایش کیفیت آن روش را پیدا کرد.

در گزینشگری که پیوست است، همه‌ی ۸ تا نمودار گنجانده شده‌اند. خانه‌های گزینشگر به شرح زیر است:

AQA	N	TE _a	Bias	CV	SN _a	SN
۹۰ یا ۵۰ گذاشته شود.	۳، ۴، ۲ یا ۶ گذاشته شود.	خطای کل مجاز	نامیزانی روش	پراکنندگی روش	به طور معمول ۲ گذاشته شود.	معیار سیگمای روش را نشان می‌دهد

- AQA (Analytical Quality Assurance) توان خطایابی برنامه‌ی پایش کیفیت انتخاب شده را نشان می‌دهد؛ مثلا AQA برابر ۹۰٪ به این معناست که اگر خطایی رخ دهد، ۹۰٪ احتمال دارد که برنامه‌ی پایش کیفیت در اولین دور سنجش، وجود خطا را نمایان کند.
- در خانه‌ی SNa به طور معمول ۲ می‌گذاریم با این فرض که ۵٪ جواب بیرون از محدوده‌ی TEa پذیرفتنی است. اگر خطای کمتری را بخواهیم باید عددهای بزرگتر بگذاریم مثلا برای اینکه حداکثر ۳٪ خطا داشته باشیم باید ۳ بگذاریم و برای این که حداکثر ۳/۴ خطا در یک میلیون جواب داشته باشیم (۰/۰۰۰۳۴٪) باید ۶ بگذاریم (کیفیت ۶ سیگما).
- N نشان دهنده‌ی تعداد کنترل در هر دور است. برای بیشتر سنجش‌های بیوشیمی استفاده از ۲ سطح کافی است. برای سنجش‌های هماتولوژی و ایمونولوژی استفاده از ۳ سطح لازم است. بسته به ویژگی‌های کارکردی روش ممکن است در یک دور، از هر سطح یک کنترل گذاشت (که آنگاه N می‌شود ۲ یا ۳)، و یا اینکه تعداد را دو برابر کرد و از هر سطح دو کنترل گذاشت (که N می‌شود ۴ یا ۶).
- اگر ویژگی‌های کارکردی روشی خیلی خوب باشد و بتوان با ۹۰٪ AQA و N ۲ یا ۳ برای آن معیار پایش پیدا کرد، زحمت و هزینه‌ی پایش کیفیت کم می‌شود: یک کنترل از هر سطح گذاشته می‌شود و کار پایش به عهده‌ی روش آماری گذاشته می‌شود.
- اگر به دلیل متوسط بودن کارکرد روش با N ۲ یا ۳ معیار پیدا نشود، باید N ۴ یا ۶ را امتحان کرد. اگر معیار پیدا شد، باید در هر دور از هر سطح ۲ تا گذاشت؛ یعنی هزینه را افزایش داد. راه دیگر به جای افزودن N امتحان ۵۰٪ AQA است. اما اگر از معیار پایش کیفیت با ۵۰٪ AQA استفاده کنیم آنگاه باید برای پایش کیفیت از روش‌های غیرآماری هم استفاده کرد زیرا برنامه‌ی آماری به تنهایی فقط ۵۰٪ توان خطایابی دارد؛ یعنی زحمت پایش کیفیت افزایش می‌یابد.
- برای روش‌های ضعیفتر، باید هم تعداد N را از ۲ یا ۳ به ۴ یا ۶ رساند و هم از ۵۰٪ AQA استفاده کرد. یعنی هم هزینه‌ی کنترل با می‌رود و هم زحمت آن.
- اما اگر کارکرد روشی چنان ضعیف باشد که حتی با افزایش N و ۵۰٪ AQA هم نتوان برایش معیار پایش کیفیت آماری پیدا کرد، آنگاه یعنی فاتحه‌ی پایش کیفیت آماری خوندس! در این حالت نمودار کشیدن و این کارها هیچ کمکی نمی‌کند و باید برای پایش چنین روش سنجشی به طور کامل به روش‌های پرزحمت و البته پرهزینه‌ی غیرآماری تکیه کرد. البته بهتر است به فکر اصلاح یا جایگزین کردن چنین روشی باشیم چون در دراز مدت زحمت و هزینه‌ی پایش چنین روشی بیشتر از زحمت و هزینه‌ی بهبود یا جایگزینی آن خواهد بود.

مقایسه شود با روش کنترل کیفی! کنونی که بدون آگاهی از خطای کل مجاز و بدون در نظر گرفتن ویژگی‌های اجرایی روش‌ها، برای همه‌ی آن‌ها نمودار می‌کشیم و برای همه یک معیار یکسان به اسم به اصطلاح قانون وستگارد! به کار می‌گیریم.

نکته‌ی مهم دیگری که باید در نظر داشته باشیم این است که پایش کیفیت وقتی معنا دارد که روش قابل قبول باشد. برای روش قابل قبول، یعنی روشی که کیفیت دارد، می‌توان دنبال برنامه‌ی پایش کیفیت گشت، که ممکن است بسته به کیفیت روش، برنامه‌ی آماری ساده‌ای برایش لازم باشد یا برنامه‌ی سختی، و یا اینکه اصلاً آمار به کارش نیاید و مجبور باشیم از راه‌های غیر آماری استفاده کنیم. اگر روشی قابل قبول نباشد، آنگاه نه تنها فاتحه‌ی پایش کیفیت آماری خوندس، و نه این که فاتحه‌ی پایش کیفیت غیر آماری هم خوندس، بلکه از اساس اصلاً فاتحه‌ی خود روش سنجش خوندس!