

توضیح کوتاهی در باره روش استفاده از نمودارهای ویژگی‌های اجرایی

دکتر وستگارد می‌گوید (نقل به مضمون): "می‌توان روش استفاده از این نمودارها را در کمتر از ۱۵ دقیقه به بچه‌های کلاس پنجم آموزش داد"، پس من هم باید بتوانم آن را در کمتر از ۱۵ خط توضیح بدهم. سعی می‌کنم چنین کنم اما پیش از آن باید کمی در باره کارهایی که لازم است پیش از استفاده از این نمودارها یا هر ابزار دیگری برای پایش کیفیت انجام شود توضیح بدهم (البته توضیحاتی که بیش از ۱۵ خط است!).

در نخستین گام باید بررسی کنیم که آیا روش‌های ما باید قابل قبول هستند یا نه. برای این کار باید:

- بدانیم که "خطای کل مجاز" (Allowable Total Error, ATE) سنجش هر آنالیت چقدر است،
- ویژگی‌های عملکردمان یعنی انحراف معیار (CV) و نامیزانی (Bias) روش‌مان را تعیین کنیم تا بتوانیم "خطای کل" کارمان با فرمول زیر را حساب کنیم:

$$TE = Bias + 2CV$$

- سپس خطای کل کارمان را با خطای کل مجاز مقایسه کنیم. اگر TE ما کمتر از ATE بود روش ما قابل قبول است.

مثال: چهار روش مختلف سنجش گلوکز با ویژگی‌های عملکردی گوناگون را بررسی می‌کنیم:

- ابتدا دانستن خطای کل مجاز: در جدول CLIA خطای کل مجاز برای سنجش گلوکز ۱۰٪ است.
- تعیین ویژگی‌های عملکردی هر یک از روش‌ها: فرض کنیم تکرار پذیری هر روش را در ۲۰ روز بررسی کرده‌ایم. برای تعیین نامیزانی نیز در مدت ۲۰ روز روی هم ۵۰ نمونه‌ی مختلف بیماران را که دارای غلظت‌های گوناگون بوده‌اند و غلظت آن‌ها به وسیله‌ی یک روش مرجع تعیین شده‌است با هر ۴ روش سنجیده‌ایم و با استفاده از معادله‌ی خط رگرسیون، نامیزانی هر روش را در مقابل روش مرجع تعیین کرده‌ایم. حاصل این بررسی‌ها در زیر آمده است:

- الف: $Bias = 3\%; CV = 5\%; TE = 3\% + 2(5\%) = 13\%$

- ب: $B = 2\%; CV = 3.5\%; TE = 9\%$

- پ: $B = 1\%; CV = 2\%; TE = 5\%$

- ت: $B = 0\%; CV = 1\%; TE = 2\%$

- **داوری در باره روش‌ها:** روش الف ناپذیرفته است چون خطای کل آن بیش از خطای کل مجاز است ($13\% > 10\%$). سه روش دیگر پذیرفته هستند و می‌توان آن‌ها را برای سنجش نمونه‌های بیماران به کار بست. روی هم یعنی این که روش الف **کیفیت ندارد** و نباید استفاده شود اما سه روش دیگر **با کیفیت است** و می‌توان از آن‌ها استفاده کرد.

اگرچه سه تا از روش‌ها پذیرفته شدند و با کیفیت هستند، اما برای استفاده از آن‌ها شرط دیگری هم لازم است:

پیش از استفاده از این روش‌های با کیفیت برای سنجش نمونه‌های بیماران، باید راهی برای حفظ کیفیت آن‌ها اندیشید؛ حفظ کیفیت، نگهداری کیفیت، کنترل کیفیت یا "پایش کیفیت" (Quality Control). (حواسمان هم باشد که اصطلاح "کنترل کیفی" درست نیست و استفاده نکنیم.)

پس گام بعدی نهداری کیفیت است. خب! کیفیت را چگونه حفظ کنیم؟ اولین و مهمترین کار آن است که خطراتی را که متوجه کیفیت کار ماست شناسایی کنیم و سدهایی سر راه آن‌ها بگذاریم تا مانع رخدادن آن‌ها شویم. سدهایی که به کار ما می‌آیند چنین است: آموزش روش درست کار به کارکنان، نگهداری معرف‌ها در دمای مناسب، بررسی‌ها و نگهداری دوره‌ای دستگاه‌ها، و غیره. با این کارها تلاش می‌کنیم مانع خراب شدن کیفیت شویم. اما به‌رغم همه‌ی این کارها، گاهی ممکن است کیفیت کار خراب شود، و در این مواقع لازم است که در زنگ هشدار به صدا درآید و ما را از خراب شدن کیفیت با خبر کند. زنگ هشدار ما در این موارد "پایش کیفیت آماری" است، یعنی سنجیدن نمونه‌ی کنترل و کشیدن نمودار و بقیه‌ی کارهای مربوط.

این زنگ‌ها را دکتر وستگارد تهیه کرده است. زنگ هشدارهای وستگارد حساسیت‌های مختلفی دارند و بسته به کیفیت هر روش باید زنگ **مناسبی** را برای پایش کیفیت آن برگزید.

مناسب بودن زنگ بستگی دارد به "عملکرد" روش ما: هر چه یک روش عملکرد بهتری داشته باشد زنگی با حساسیت کمتر لازم دارد و برعکس هر چه کیفیت یک روش پایین‌تر باشد زنگی با حساسیت بیشتر نیاز دارد. همانطور که هر چه نمره‌های سه ماه اول یک دانش‌آموز بهتر باشد، یعنی عملکرد بهتری داشته باشد و کیفیت درس خواندنش بالاتر باشد، برای اطمینان از این که در پایان سال نمره‌ی قبولی (یعنی ۱۰) خواهد گرفت پایش کمتری نیاز دارد، و نیز همانطور که برای دانش‌آموزی که با نمره‌ی مثلاً ۱۱ در سه ماه اول قبول شده است، یعنی کیفیت درس خواندنش پایین است، برای این که مطمئن باشیم در پایان سال قبول خواهد شد مراقبت (پایش کیفیت) سخت‌تری نیاز خواهد بود. بنا بر این، همانطور که اگر ما معلم این بچه‌ها باشیم بسته به عملکردشان برای هر یک دستور مراقبت ویژه‌اش را به خانواده‌اش توصیه خواهیم کرد، برای روش‌هایمان هم باید چنین کنیم.

در مثال بالا اگرچه سه روش پذیرفته شده است و با کیفیت شناخته شده‌اند، اما کیفیت آن‌ها متفاوت است: روش ب با کیفیت است، روش پ با کیفیت‌تر و روش ت خیلی با کیفیت‌تر. مثل پایش کیفیت درس دانش‌آموزان، باید برای پایش کیفیت هر یک از این روش‌ها قانون مناسب عملکردش را توصیه کرد. "نمودارهای ویژگی‌های اجرایی" (نمودارهای OPSpec) همین کار را می‌کند: بسته به میزان خوب بودن عملکرد یک روش، یعنی میزان کوچک بودن انحراف معیار و نامیزانی‌اش (ویژگی‌های عملکردی)، برایش قانون پایش کیفیت مناسب را نشان می‌دهد. اسمش از همین جا آمده است: نمودارهای ویژگی‌های اجرایی (Operational Specification Charts).

بناست با استفاده از این نمودارها برای هر روش زنگ هشدار مناسب آن را انتخاب کرد طوری که وقتی انحراف معیار و/یا نامیزانی یک روش به اندازه‌ای زیاد شده است که نزدیک است خطای کل آن از خطای کل مجاز بیشتر شود، آن زنگ ما را خبر کند تا اشکال را برطرف کنیم.

پرسش: با چقدر افزایش نامیزانی یا انحراف معیار باید زنگ ما به صدا درآید؟ همین پرسش را پدر و مادر دانش‌آموزان ممکن است از ما بپرسند: با چقدر کاهش نمره در آزمون‌های مستمر در طول سال ممکن است بچه‌ی ما در خطر آن باشد که آخر سال رد شود؟ پاسخ در هر دو مورد یکسان است: بستگی دارد به این که جایگاه کنونی روش یا دانش‌آموز شما چقدر تا مرز رد شدن فاصله دارد. برای دانش‌آموز کم‌کیفیتی که نمره‌ی ترم ۱ وی ۱۳ شده است کاهش مثلا ۲ نمره و رسیدن به نمره‌ی ۱۱ به معنای خطر رد شدن است اما همین مقدار و حتا بیش از این مقدار کاهش نمره برای دانش‌آموزی که نمره‌ی ترم اول وی ۱۹ بوده است به معنای خطر نیست و نباید "سر و صدا" راه انداخت. همچنین است برای روش‌های آزمایشگاهی: برای یک روش سنجش گلوکز که خطای کل آن مثلا ۷٪ است، اندکی افزایش نامیزانی یا انحراف معیار به معنای خطر است و باید معیاری برای پایش کیفیت آن برگزید که به در صورت چنین افزایش اندکی ما را خبر کند. اما برای روشی با خطای کل ۲٪ چنین نیست و باید روشی با حساسیت کمتر برگزید که نه تنها با اندکی افزایش نامیزانی یا انحراف معیار اعلام خطا نکند بلکه با افزایش مقادیر متوسط در نامیزانی یا انحراف معیار نیز اعلام خطا نکند. در مورد روش اول، برای این که بتوان با اندکی افزایش نامیزانی یا انحراف معیار از خطر آگاه شد باید مرزهای نمودار کنترل کیفیت را بسته انتخاب کرد یعنی $\pm 2s$ ، در هر دور ۴ تا ۶ کانترا کنترل گذاشت و معیارهای چندقانونی به کار بست. همه‌ی این‌ها برای این است که حساسیت زنگ هشدارمان را افزایش دهیم. با چنین پایش حساسی علاوه بر این که هزینه و سختی پایش کیفیت افزایش می‌یابد هشدارهای کاذب هم افزایش می‌یابد در نزدیک به ۷٪ از دورهایی که هیچ خطایی وجود ندارد نیز هشدار خطا خواهیم داشت؛ که سبب می‌شود بیهوده کار را متوقف کنیم و پس از جست و جوی فراوان به این نتیجه برسیم که هشدار کاذب بوده است. اما برای روش دوم که خیلی با کیفیت است هرگز به زنگی با چنین حساسیتی نیاز نداریم؛ کافی است نموداری بکشیم با مرزهای $\pm 3.5s$ ، در هر دور فقط یک یا دو کانترا بگذاریم و از معیار تک‌قانونی و ساده‌ی $1:3.5s$ استفاده کنیم. یعنی تا وقتی که نتیجه‌ی کانترا درون مرزهای $\pm 3.5s$ می‌افتد اشکالی در کار نیست. در این صورت هزینه و سختی پایش کیفیت بسیار کم خواهد بود و نیز میزان هشدار کاذب نزدیک به صفر خواهد بود.

دکتر وستگارد برای هر دو خطای افزایش نامیزانی و افزایش انحراف به طور جداگانه نمودارهای OPSpec تهیه کرده است. چون با دستگاه‌های امروزی مشکل بیشتر در افزایش نامیزانی است تا افزایش انحراف معیار، وی نیز بیشتر بر روی نمودارهای مربوط به افزایش نامیزانی تاکید دارد. برای شناسایی خطای افزایش نامیزانی هم ۸ نمودار تهیه کرده است: ۴ تا با احتمال خطایابی ۹۰٪ و ۴ تا هم با احتمال خطایابی ۵۰٪.

چرا احتمال خطایابی ۱۰۰٪ نداریم؟ چون اگر خواهیم حساسیت معیارهایمان را تا این حد افزایش دهیم آنگاه درصد هشدار کاذب بسیار بالا خواهد بود و بنا بر این به ۹۰٪ بسنده شده است.

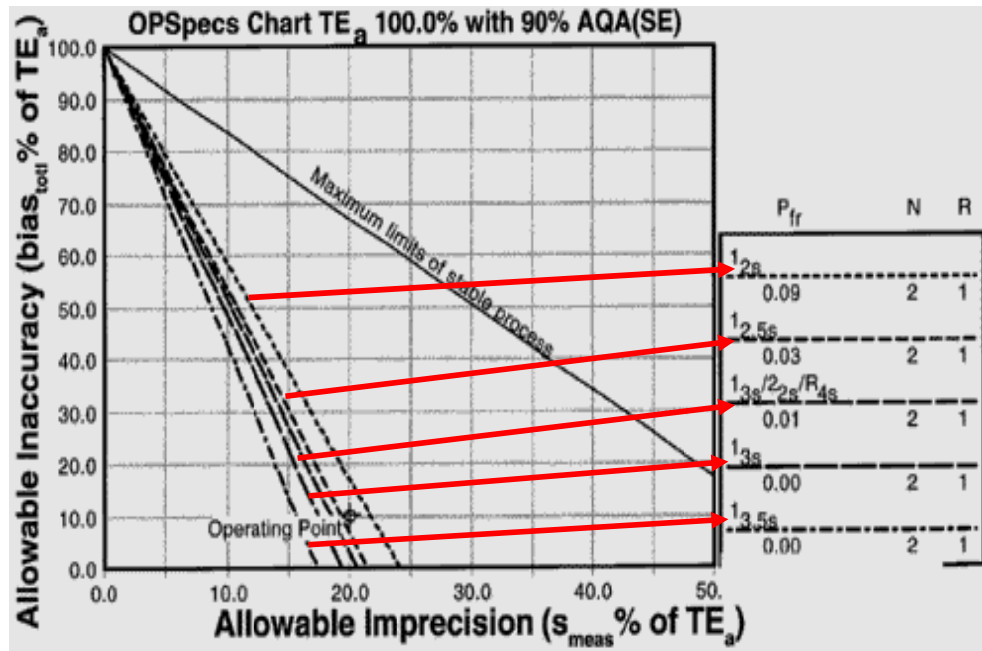
چرا ۵۰٪؟ چون گاهی کیفیت یک روش اگرچه پذیرفته است اما چنان پایین است که نمی‌توان قانونی برای پایش کیفیت آن یافت که بتواند با احتمال ۹۰٪ خطاهایی را که رخ می‌دهد شناسایی کند و مجبوریم برویم سراغ معیارهایی که احتمال خطایابی آن‌ها کمتر است.

خب! وقتی که احتمال خطایابی ۵۰٪ است تکلیف ما با ۵۰٪ خطاهایی که شناسایی نخواهند شد چه می‌شود؟ وقتی که مجبوریم از چنین قانون‌هایی استفاده کنیم باید فعالیت‌های پیشگیری از خطا را بیشتر کنیم و نیز نقش پایش کیفیت غیرآمراری را افزایش دهیم. گاهی حتا ممکن است یک روش اگرچه با کیفیت شناخته شده است اما چنان کم‌کیفیت باشد که حتا با احتمال ۵۰٪ هم نتوان برایش معیار پایش کیفیت پیدا کرد. در این صورت عمده‌ی بار پایش کیفیت بر دوش پیشگیری از بروز خطا و پایش کیفیت غیرآمراری است.

هر کدام از نمودارهای موجود در هر دسته ۴ تایی (احتمال ۹۰٪ و احتمال ۵۰٪) از نظر تعداد کانترا دارای ۲، ۳، ۴ یا ۶ کانترا در هر دور هستند. ۲ و ۴ کانترا برای روش‌هایی که باید برای آن‌ها دو سطح کانترا را گذاشت (یک کانترا سطح طبیعی و یکی از دو سطح پایین یا بالا)، و ۳ و ۶ برای روش‌هایی که باید برای آن‌ها هر ۳ سطح کانترا (پایین، طبیعی و بالا) را گذاشت.

چرا گاهی باید به جای ۲ یا ۳ کانترا، ۴ یا ۶ کانترا در هر دور گذاشت؟ چون گاهی کیفیت روش ما پایین است و نمی‌توانیم در نمودارهای ۲ یا ۳ کانترا برای معیار QC پیدا کنیم و (اگرچه دلمان نمی‌خواهد!) مجبوریم برویم سراغ نمودارهای ۴ یا ۶ کانترا بلکه در آن‌ها بتوانیم معیاری پیدا کنیم.

در زیر نمودار احتمال ۹۰٪ و ۲ سطح کانترا نمایش داده شده است. بر روی این نمودار، ۵ خط نقطه چین می‌بینیم که معیار پایش کیفیت مربوط به هر خط به ترتیب چپ به راست در مستطیل کنار جدول از پایین به بالا نوشته شده است. معیارهای درون مستطیل از پایین به بالا سخت‌تر و پرهزینه‌تر می‌شوند.



حالا پس از این توضیح ها برویم سراغ "کمتر از ۱۵ خط" و توضیح چگونگی استفاده از نمودارهای ویژگی‌های اجرایی با استفاده از مثال زیر برای سنجش ALT و پایش کیفیت با ۲ سطح کنترل:

$$ATE = 20\%; \text{ Bias} = 3\%; \text{ CV} = 2\%$$

(۱) حساب کنید نامیزانی چند درصد خطای کل مجاز است:

$$B / ATE * 100 = 3 / 20 * 100 = 15\%$$

(۲) حساب کنید انحراف معیار چند درصد خطای کل مجاز است:

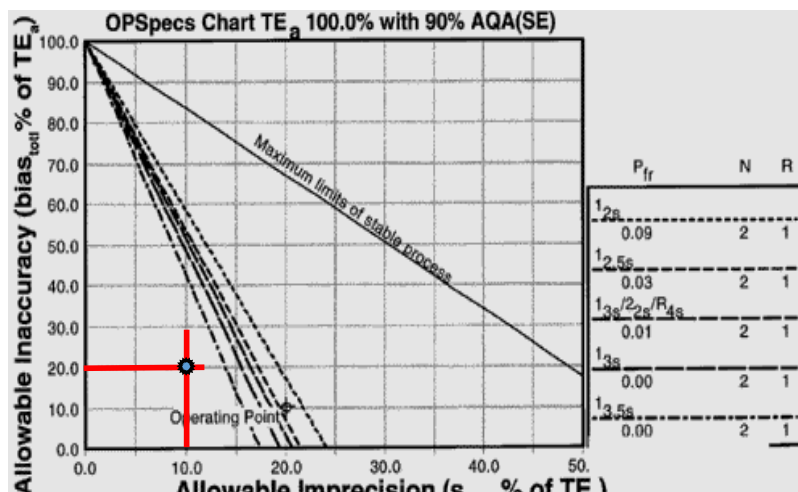
$$CV / ATE * 100 = 2 / 20 * 100 = 10\%$$

(۳) بروید سراغ نمودار با احتمال ۹۰٪ و $N = 3$

(۴) محل B/ATE را روی محور عرض‌ها پیدا کنید از آن جا خطی افقی بکشید

(۵) محل CV/ATE را روی محور عرض‌ها پیدا کنید و خطی بر آن عمود کنید

(۶) محل برخورد دو خط بالا می‌شود "نقطه‌ی عملکردی" روش شما؛ در شکل زیر با نقطه‌ی سیاه رنگ با مرکز آبی نمایش داده شده است.



۷) ببینید کدام یک از خط‌های نقطه‌چین مورب در طرف راست نقطه‌ی عملکردی شما قرار می‌گیرد. همه‌ی معیارهایی که مربوط به آن خط‌ها هستند برای پایش کیفیت روش شما قابل استفاده هستند.

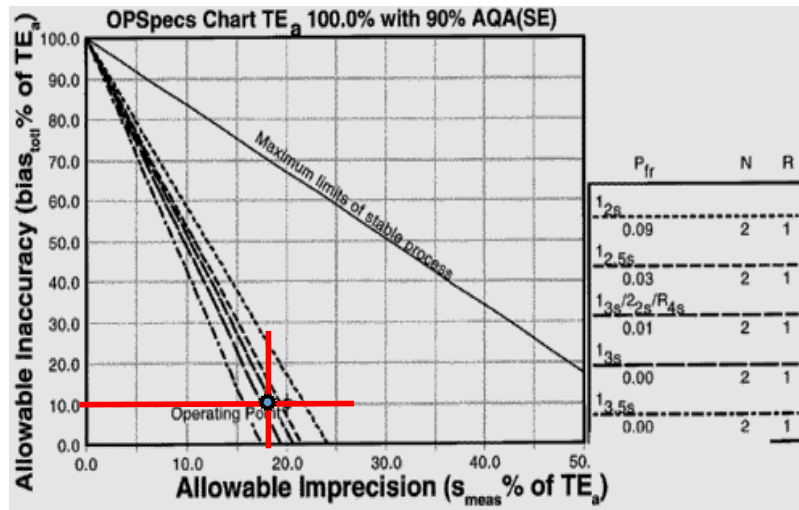
۸) از بین معیارهای قابل استفاده، ساده‌ترین و عملی‌ترین آن را انتخاب کنید. در مورد مثال بالا که یک روش بسیار با کیفیت است می‌توان از معیار ساده‌ی 1:3.5s استفاده کرد.

مثال دیگری برای همین آزمایش:

Bias = 2%; CV = 3.6%

%B/ATE = 10%

%CV/ATE = 18%



با چنین کیفیتی دیگر نمی‌توان از معیارهای 1:3.5s و 1:3s استفاده کرد و باید رفت سراغ معیار چندقانونی و سخت داده شده است). البته می‌توان به جای این معیار سخت از معیار بالای آن که راحت‌تر است یعنی معیار 1:2.5s استفاده کرد اما هشدار کاذب آن ۳٪ است. این که کدام را انتخاب کنیم بستگی دارد به شرایط: اگر کاربر شما می‌تواند معیار چندقانونی را به کار ببندد و تفسیر کند از آن استفاده کنید اما اگر چنین نیست از معیار 1:2.5s استفاده کنید.

ترتیب استفاده از نمودارها:

ابتدا باید رفت سراغ نمودارهای ۹۰٪ با ۲ یا ۳ کنترل. اگر در آن معیار مناسب پیدا نشد باید رفت سراغ نمودارهای ۹۰٪ با ۲ یا ۳ کنترل. اگر باز هم نشد نمودارهای ۵۰٪ با ۲ یا ۳ کنترل و دست آخر باید نمودارهای ۵۰٪ با ۴ یا ۶ کنترل را بررسی کرد. اگر حتی در این نمودارها هم معیار پایش پیدا نشد، آنگاه باید سخت‌ترین معیار یعنی بالاترین معیار در مستطیل را به کار گرفت اما به یاد داشت که توان خطایابی آن کمتر از ۵۰ است و همانطور که پیشتر گفته شد برای پایش کیفیت بر راه سخت پیشگیری و خطایابی غیرآماري تکیه کرد؛ یعنی بهرغم این که ۴ یا ۶ کنترل در هر دور می‌گذاریم باز هم باید کارهای بسیار دیگری انجام دهیم شامل افزایش نگهداری پیشگیرانه، به کار گماردن کاربر با تجربه برای آن سنجش، افزایش ارزیابی‌های دوره‌ای کارکرد دستگاه، توجه به همخوانی نتیجه‌ی سنجش با دیگر آزمایش‌های بیمار، توجه به همخوانی نتیجه با بالین بیمار و و.