

این نوشته توی شماره ۲۰۵ (شهریور ۱۴۰۰) نشریه اخبار آزمایشگاهی چاپ شده. متن زیر نسبت به متنی که توی اخبار آزمایشگاهی چاپ شده، به دلیل بهرهمندی از ویرایش وسیع املائی و محتوایی آقای دکتر مهدی صابونی، بسیار روانتر و رساتر شده که سبب می‌شه صمیمانه سپاسگزار آقای دکتر صابونی باشم.

«خطای کل مجاز تعدیل شده» چیه؟!

نقدی به مفهوم «تعدیل خطای کل مجاز» به دنبال سنجش تکراری نمونه‌ها، ارائه شده توی کتاب «مدیریت کیفیت در بیوشیمی» نوشته آقای دکتر رضا محمدی.

حسن بیات؛ درس‌خونده علوم آزمایشگاهی

توضیح:

این نوشته با انشای گفتاری نوشته شده. قصد خاصی از اینطور نوشتن ندارم جز این که با نوشتن به این شکل هم مطلب رو راحت‌تر بیان می‌کنم و هم سریع‌تر می‌نویسم؛ که هر دو مورد، سبب صرفه‌جویی توی وقت کمی می‌شه که برای مطالعه و نوشتن دارم. از طرف دیگه، توی سال‌های اخیر رد و بدل کردن پیام‌های متنی با زبان گفتاری توی فضای مجازی سبب شده دیگه مطالعه مطالب به این شکل خیلی نامأنوس نباشه. به هر حال، از دست‌اندرکاران نشریه اخبار آزمایشگاهی خواهش کردم حداقل به عنوان یه تجربه، این نوشته به این شکل منتشر بشه.

مثال‌های ساده‌ای از فعالیت‌های غیر آزمایشگاهی آوردم که به باور من به فهم مطلب کمک می‌کنه. از اونجایی که «اندازه‌گیری کمی» بخش عمده‌ای از کار آزمایشگاه و محصول کار ما عمدتاً تولید «نتایج عددی‌یه»، می‌شه به آزمایشگاه به چشم یه «کارگاه تولید نتایج کمی» نگاه کرد؛ و بنا بر این، میشه از مثال‌ها و تشبیه‌هایی از کارهای تولیدی و بخصوص مؤسسات اندازه‌گیری برای کمک به توضیح مباحث آزمایشگاهی استفاده کرد. امیدوارم مثال‌های ارائه شده از جنس مثال‌هایی باشه که پروفیسور وستگارد در باره‌شون می‌گه «تو مثال‌هایی می‌زنی که از هر سدی در مقابل فهمیدن و یادگیری عبور می‌کنه». به هر حال، امیدوارم سادگی مثال‌ها و همینطور ساده‌نویسی و مفصل‌نویسی من، سبب رنجش خوانندگان گرامی نشه.

توی خیلی از فرمول‌هایی که توی متن هست، علامت % رو نذاشتم تا از شلوغ شدن فرمول‌ها جلوگیری بشه؛ مثلاً $TAE = \%Bias + 2CV$ به صورت $TAE = Bias + CV$ نوشته شده. مطمئناً خوانندگان به خوبی آگاه هستن که کجاها باید به شکل درصد باشه؛ مثلاً این که هر جا توی فرمول CV هست، بقیه اجزا باید به شکل درصد باشه.

مقدمه

این نوشته، نقدی به مفهومی با اسم «خطای کل مجازِ تعدیل شده» (Modified Allowable Total Error; mTEa) که توسط آقای دکتر رضا محمدی توی کتابشون با عنوان «مدیریت کیفیت در بیوشیمی» (نشر آییز؛ سال ۱۳۹۵) ارائه شده. تو صفحه ۵۶ کتاب، نویسنده مدعی شده اگه برای کوچیک کردن CV، نمونه‌ها رو بیشتر از ۱ بار اندازه‌گیری و میانگین نتایج رو گزارش بکنیم، اون وقت باید خطای کل مجاز (خکم) (Allowable Total Error, TEa) رو هم تعدیل، یا به عبارت بهتر، کم بکنیم. نویسنده یه فرمول هم برای این کار ابداع و ارائه کرده:

$$\%mTEa = [(2+\sqrt{n})/3\sqrt{n}]*TEa$$

این پیشنهاد به طور خلاصه این می‌شه که وقتی ما با عملکرد معمول‌مون، نمی‌تونیم درون محدوده مجاز باشیم و از حد مجاز رد می‌شیم، و بنا بر این با صرف تلاش بیشتر، هزینه بیشتر، و وقت بیشتر، عملکردمون رو بهتر می‌کنیم تا بیفتیم توی محدوده مجاز، همزمان باید حد مجاز هم برای ما کمتر بشه؛ کاری که امکان داره نه تنها پیامد بهتری به دنبال نداشته باشه و سبب نشه یه عملکرد بیرون از محدوده مجاز بیفته توی محدوده مجاز، حتا ممکنه سبب بشه یه عملکرد قابل قبول بیفته بیرون از محدوده مجازو مردود بشه. قصد من از این نوشته اینه که نشون بدم همچین تعدیلی درست نیست. خطای مجاز برای هر عملکردی تو هر سامانه‌ای، از جمله خطای مجاز سنجش‌های آزمایشگاهی، مقدار ثابتیه که با توجه به شرایط و نیازهای اون سامانه تعریف و تعیین می‌شه و تلاش‌هایی که برای بهبود عملکرد انجام می‌شه سبب نمی‌شه که خطای مجاز رو کم بکنیم. برای بیان موضوع، لازمه اول یه مرور کوتاه (البته نه خیلی کوتاه) داشته باشیم به بعضی از جنبه‌های مبحث پایش کیفیت (پک).

خطای مجاز «محصول» در مقابل خطای مجاز «تولید»

خطای مجاز «محصول» به ما می‌گه که یه دونه محصول، چه مشخصاتی باید داشته باشه تا اون رو خوب یا قابل قبول به حساب بیاریم. در مقابل، خطای مجاز «تولید» در باره مشخصات عملکرد خوب یا قابل قبول صحبت می‌کنه.

مثلاً اگه به یه تولیدکننده بگن خطای مجاز برای قطر توپ‌های بازی برابر با ۲۰ درصده، این می‌شه مشخصات یه دونه محصول خوب یا اختلاف مجاز. این خطای مجاز به ما می‌گه اگه مثلاً یه مشتری توپ‌هایی با قطر ۳۰ سانت (قطر هدف برابر ۳۰) بخواد، تا ۶ سانت کم و زیاد اشکال نداره و توپ‌هایی به قطر ۲۴ سانت تا ۳۶ سانت، خوب، درست، سالم یا قابل قبول به حساب می‌آن و توپ‌هایی که بیشتر از ۶ سانت کم و زیاد دارن یعنی توپ‌هایی با قطر کمتر از ۲۴ سانت یا بیشتر از ۳۶ سانت، بد، نادرست، ناسالم یا غیر قابل قبول محسوب می‌شن.

اما علاوه بر دونستن خطای مجاز محصول، لازمه یه تولیدکننده خطای مجاز تولید رو هم بدونه؛ یعنی بدونه چند درصد از توپ‌هایی که تولید می‌کنه باید خوب باشن: همه‌شون (۱۰۰٪)؟ ۹۵٪؟ ۷۰٪؟ چند درصد؟ به بیان دیگه، قطر چند درصد از توپ‌ها باید بیفته توی محدوده مجاز 20 ± 30 (۲۴ سانت تا ۳۶ سانت). فرض کنیم توی قرارداد با یه توپ‌ساز، علاوه بر «اختلاف مجاز، برابره با ۲۰٪»، بنویسن «خطای مجاز تولید، برابره با ۸۵٪». این یعنی: باید قطر حداقل ۸۵٪ از توپ‌ها بیفته توی محدوده

مجاز $20 \pm 30\%$ ؛ یا قطر حداکثر 15% از توپ‌ها می‌تونه بیفته بیرون از فاصله $20 \pm 30\%$. حالا تکلیف اون توپ‌ساز روشننه و می‌دونه که اگه یه سفارش 10000 تایی رو تحویل می‌ده، باید قطر حداقل 8500 تا از توپ‌ها قابل قبول (24 تا 36 سانت) باشه. از اون طرف اگه قطر حداکثر 1500 تاشون نادرست باشه (بیفته بیرون از 24 تا 36)، اشکال نداره. قرارداد بدی نیست. چون این معیارها خیلی سخت به نظر نمی‌آید: (الف) خطای مجاز $= 20\%$ ، (ب) خرابی مجاز $= 15\%$. البته همه مشتری‌ها خواسته‌های یکسان ندارن، ممکنه یه مشتری دیگه بیاد بگه ما 1000 تا توپ با قطر 20 سانت می‌خواهیم با معیارهای مثلاً «الزامات بهبود کیفیت فوتبال»: (الف) خطای مجاز $= 5\%$ ، (ب) خرابی مجاز $= 1\%$. از منظر نیازهای این مشتری، (الف) یه دونه محصول قابل قبول، تویی‌یه که قطرش بین 19 تا 21 سانت باشه و (ب) تولید قابل قبول یعنی خیالش راحت باشه توی 1000 تا تویی که تحویل می‌گیره حداکثر 10 تاشون خرابه و حداقل 990 تاشون خوبه.

حالا «توپ‌بازی» رو بذاریم کنار، ببخشید «توپ‌سازی» رو، بریم سراغ چند تا مثال درباره «اندازمگیری» و «نتیجه اندازمگیری». کار اصلی ما آزمایشگاهی‌ها اندازمگیری‌یه و مثل همه شغل‌های مرتبط با اندازمگیری، متر کردن کوچه‌ها باشه یا وزن کردن اشیا یا هر اندازمگیری دیگه، محصول ما می‌شه «نتیجه اندازمگیری» و تولید ما می‌شه «کار اندازمگیری»؛ به بیان بهتر: «نتیجه سنجش» و «عملکرد سنجشی».

«نتیجه سنجش» قابل قبول. به نتیجه‌ای می‌گیم قابل قبول (درست یا خوب) که اختلاف اون با مقدار واقعی کمتر از مقدار مجاز باشه. اگه کار ما اندازمگیری طول کوچه‌هاست و به ما گفتن «اگه نتیجه اندازمگیری‌تون حداکثر 10% با اندازه واقعی اختلاف داشته باشه، می‌گیم جوابتون خوبه (قابل قبوله)»، این می‌شه اختلاف مجاز، یا انحراف مجاز، یا خطای مجاز (یا هر اسم مناسب دیگه). حالا اگه ما بریم یه کوچه با طول واقعی 100 متر رو اندازه بگیریم، اختلاف مجاز واسه طول این کوچه می‌شه 10 متر و محدوده مجاز نتایج می‌شه از 90 متر تا 110 متر:

$$10\% * 100 = 10 \text{ m}$$

$$100 - 10 = 90 \text{ m}$$

$$100 + 10 = 110 \text{ m}$$

تا وقتی که جوابمون تو فاصله 90 و 110 باشه، جوابمون یه محصول سالم به شمار می‌آید؛ یعنی درست، خوبه، یا قابل قبوله؛ مثلاً اگه گزارش بکنیم 108 متر یا 91 متر، کنار جوابمون علامت «درست» می‌زنن؛ اما اگه گزارش بکنیم 111 متر یا 86 متر، علامت «نادرست» می‌زنن.

در مورد اندازمگیری‌های آزمایشگاهی هم همینطوره. مثلاً با معیار CLIA، یه جواب گلوکز درست، جوابی‌یه که حداکثر 10% با مقدار واقعی‌ش اختلاف داشته باشه. اگه مقدار واقعی گلوکز توی یه نمونه 100 mg/dL باشه، جواب‌های بین 90 و 110 درست به حساب می‌آید و جواب‌های بیرون از این محدوده، نادرست به حساب می‌آید.

«عملکرد سنجشی» قابل قبول. حالا فرض کنیم ما یه شرکت نقشه‌برداری داریم و رفتیم با شهرداری یه شهر قرارداد بستیم براشون کوچه‌های شهر رو متر کنیم. موقع قرارداد بستن باید ازشون پرسیم: (الف)

خطای مجاز چقدره؟ دست‌اندرکاران شهرداری هم بسته به کاربرد نتایج، تعیین می‌کنن که چقدر خطا براشون مهمه و اون رو توی قرارداد می‌نویسن؛ مثلاً همون ۱۰٪ مثال بالا. اما حواسمون باید باشه که فقط دوندستن خطای مجاز کافی نیست، بلکه باید در باره درصد مجاز خرابی هم بیرسیم: **(ب) چند درصد از نتایج ما باید درست باشه؟** این پرسش یعنی: از نظر شما «عملکرد سنجشی» قابل قبول چه جور عملکردیه؟ معلوم کنین حداکثر چند درصد از نتایج ما می‌تونه بیفته بیرون از محدوده مجاز ۱۰٪±؟ فرض کنیم شهرداری توی قرارداد می‌نویسه ۵٪.

حالا تکلیف ما معلومه: اگه اون منطقه ۱۰۰۰۰ تا کوچه داره، برای این که آخر کار شهرداری پول ما رو بده، طبق قرارداد، باید حداقل اندازه ۹۵۰۰ تا از کوچه‌ها رو درست گزارش کرده باشیم (یعنی اندازه گزارش شده برای حداقل ۹۵۰۰ تا از کوچه‌ها، بیشتر از ۱۰٪ از مقدار واقعی‌شون کمتر یا بیشتر نباشه). با این معیار ما فقط ۵٪ جای خطا داریم یعنی اندازه گزارش شده برای حداکثر ۵۰۰ تا از کوچه‌ها می‌تونه بیشتر از ۱۰٪ از مقدار واقعی‌شون کوچیکتر یا بزرگتر باشه.

چرا دوندستن خطای مجاز عملکرد مهمه؟ چون در عمل ما نمی‌خواهیم یا نمی‌تونیم تک تک محصولات رو واریسی کنیم ببینیم کدوم سالمه کدوم ناسالم. مثلاً می‌خواهیم ۱۰۰۰ تا پرتقال بخریم، توی هر بسته پذیرایی یه دونه بذاریم بدیم دست شرکت‌کننده‌های کنفرانس. نمی‌تونیم تک تک پرتقال‌ها رو باز کنیم ببینیم کدوم سالمه و کدوم ناسالم، تا ناسالم‌ها رو دور بریزیم و سالم‌ها رو بدیم دست مهمون‌ها. اما اگه بدونیم پرتقال‌ها مثلاً کمتر از ۱٪ خرابی دارن، می‌گیم «خرابی کل این بار کمه، می‌خریم می‌ذاریم تو بسته‌ها. اینجوری پرتقالی که نصیب هر کسی می‌شه حداقل ۹۹٪ احتمال داره سالم باشه و حداکثر ۱٪ احتمال داره که خراب باشه».

توی آزمایشگاه هم داستان همینه؛ مثلاً وقتی یه اتوآنالایزر در ماه ۵۰۰۰ تا جواب گلوکز تولید می‌کنه، راهی نداریم که بدونیم کدوم جواب درسته و کدوم نادرست. مگر این که مقادیر واقعی اون ۵۰۰۰ تا نتیجه رو بدونیم؛ که اگه می‌دونستیم دیگه نیازی نبود نمونه‌ها رو بذاریم روی این اتوآنالایزر! تنها اطلاعاتی که می‌تونیم درباره درستی/ نادرستی نتایج بدیم اینه که تعیین کنیم هر جواب چقدر احتمال داره درست باشه. همه‌مون می‌دونیم که احتمال رو می‌شه از روی فراوانی اون حساب کرد. مثلاً اگه توی یه کیسه ۱۰۰ تا توپ باشه که فقط ۹۵ تا شون سالم باشه، احتمال این که توپی که به طور تصادفی از کیسه در می‌آیم سالم باشه برابر با ۹۵ درصده. اگه یه خط تولید داشته باشیم که حداقل ۹۵٪ از تولیدش خوب باشه، احتمال خوب بودن هر محصول برابره با حداقل ۹۵٪. به همین شکل، اگه یه اتوآنالایزر داشته باشیم که حداقل ۹۵٪ از جواب‌های گلوکزش کمتر از ۱۰٪ با مقدار واقعی اختلاف داشته باشن، احتمال این که هر جواب کمتر از ۱۰٪ با مقدار واقعی اختلاف داشته باشه برابره با حداقل ۹۵٪.

اینجاس که ارزیابی عملکرد سنجشی و اطمینان از پذیرفته بودن اون اهمیت پیدا می‌کنه. در مثال اندازه‌گیری‌های کوچه‌ها، اگه عملکرد سنجشی ما قابل قبول باشه، اون وقت کاربر نتایج، در مورد طول هر کوچه، حداقل ۹۵٪ مطمئنه که حداکثر ۱۰٪ از مقدار واقعی کمتره یا بیشتره. مثلاً اگه بخش عمرانی شهرداری بخواد کوچه‌های اون شهر رو سنگفرش بکنه و از ما بیرسه که توی حساب کتابشون بر اساس نتایجی که ما گزارش کردیم، چقدر خطا رو باید در نظر بگیرن (یا به قول مهندس‌های عمران، چند درصد تolerانس در نظر بگیرن) می‌تونیم توضیح بدیم «طول این کوچه‌ها با خطای مجاز ۱۰٪ اندازه‌گیری شده، پس شما حداکثر ۱۰٪ کم و زیاد رو در نظر بگیرین». بعد می‌پرسن: «یعنی خیالمون

صد درصد راحت باشه در مورد هیچ کوجه‌ای بیشتر از ۱۰٪ کم و زیاد نداریم؟» ما هم جواب می‌دیم: «صد درصد که نه! خیالتون حداقل ۹۵٪ راحت باشه»؛ یعنی حداقل توی ۹۵٪ از کوجه‌ها، حساب کتابتون درست از کار درمید اما توی حداکثر ۵٪ از کوجه‌ها، با بیشتر از ۱۰٪ کم و زیاد روبرو می‌شین. **کدوم کوجه‌ها خطاشون از ۱۰٪ بیشتره؟! (کدوم پرتقال‌ها خرابه؟!)** نمی‌دونیم! به خاطر همینه که می‌گیم به محاسباتتون برای هر کوجه، حداقل ۹۵٪ اطمینان داشته باشین. اگه این کیفیت سنجشی برای تصمیمات عمرانی‌تون مناسبه، با ما قرارداد ببندین کوجه‌ها رو اندازه‌گیری کنیم.

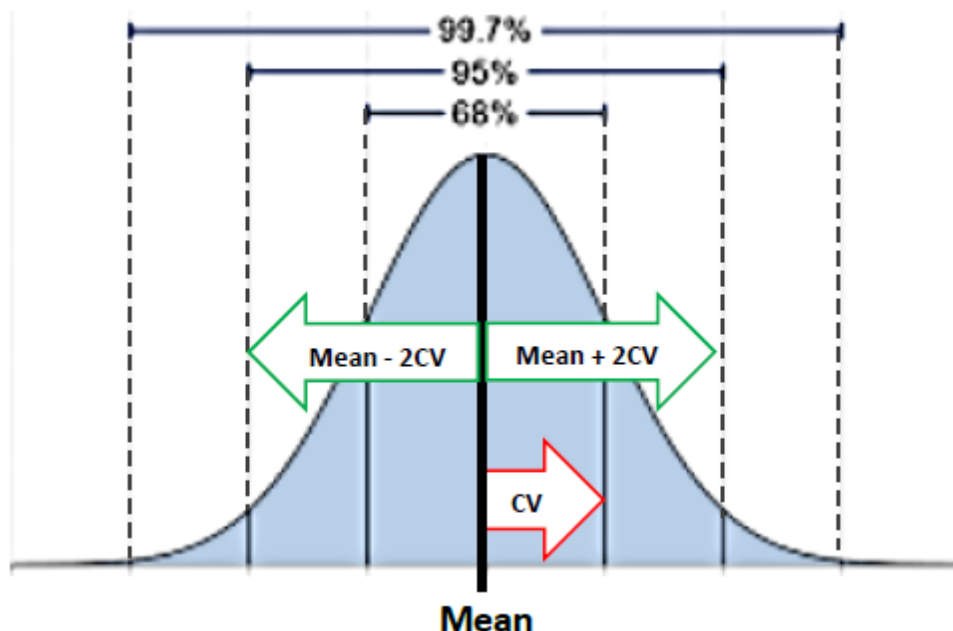
در مورد اندازه‌گیری‌های آزمایشگاهی هم همینطوره؛ یعنی هم لازمه بدونیم اختلاف مجاز واسه هر نتیجه چقدره و هم این که بدونیم چند درصد از جواب‌های ما می‌تونه نادرست باشه (بیفته بیرون از محدوده مجاز). مورد اول، یعنی اختلاف مجاز رو به شکل سراسر است خیلی جاها نوشتن. توی جدول‌های خطای مجاز CLIA و RCPA می‌شه خطاهای مجاز برای سنجش خیلی از آنالیت‌ها رو پیدا کرد؛ برای بعضی‌ها شون به شکل درصد نوشتن (مثلاً CLIA خطای مجاز کلاسترول رو ۱۰٪ در نظر گرفته) و برای بعضی‌ها شون به شکل یه مقدار ثابت نوشتن (باز هم مثلاً CLIA خطای مجاز کلاسیم رو mg/dL ۱ در نظر گرفته). اما در مورد دومی، یعنی این که درصد مجاز جواب‌های نادرست چقدره، راستش رو بخواهین جایی به شکل سراسر است چیزی ننوشتن. از حدود ۵۰ سال پیش که پروفیسور وستگارد مدل "خطای کل" رو ارائه کرده، یه جور توافق نانوشته بوده که اجازه داریم حداکثر ۵٪ از جواب‌ها مون بیفته بیرون از محدوده مجاز (یعنی درصد مجاز خرابی یا تولید خطا یا هر اصطلاح مناسب دیگه، برابره با ۵٪).

مثال: فرض کنیم واسه سنجش گلوکز، اختلاف مجاز رو برابر ۱۰٪ و درصد مجاز خرابی رو برابر ۵٪ درصد در نظر گرفتیم و با یه مرکز قرارداد بستیم گلوکز ۱۰۰۰۰ تا از کارکنانش رو اندازه‌گیری کنیم. طبق این الزامات، برای این که عملکردمون قابل قبول باشه (و آخر کار پولمون رو بدن) باید حداقل ۹۵۰۰ تا از جواب‌ها مون درست باشه، یعنی از مقدار واقعی، حداکثر ۱۰٪ کمتر یا بیشتر باشن و حداکثر ۵۰۰ تا از جواب‌ها مون بیفته بیرون از محدوده مجاز $\pm 10\%$. اگه عملکرد سنجشی آزمایشگاه در مورد گلوکز خوب باشه و کمتر از ۵٪ جواب‌ها مون نادرست باشه، می‌تونیم به یه پزشک در مورد کیفیت کارمون اینجوری توضیح بدیم: «حداقل ۹۵٪ از نتایج گلوکز ما درست هستن یعنی حداکثر ۱۰٪ از مقدار واقعی شون کمتر یا بیشتر هستن و حداکثر ۵٪ از جواب‌ها نادرست هستن و بیشتر از ۱۰٪ خطا (اختلاف) دارن». **کدوم نتیجه‌ها بیشتر از ۱۰٪ خطا داره؟** (کدوم پرتقال‌ها خرابه؟ اندازه کدوم کوجه‌ها بیشتر از ۱۰٪ خطا داره؟) نمی‌دونیم! به خاطر همینه که می‌گیم در مورد هر جواب گلوکزی که از آزمایشگاه ما به دستتون می‌رسه، حداقل ۹۵٪ اطمینان داشته باشین که دیگه خیلی... خیلی... خیلی... خطا داشته باشه، حداکثر ۱۰٪ از مقدار واقعی‌ش کمتره یا بیشتره. اگه این کیفیت رو واسه تصمیمات بالینی‌تون مناسب می‌دونین، مریض‌هاتون رو بفرستین آزمایشگاه ما.

اما...، چجوری تعیین کنیم که چند درصد از نتایج ما درسته؟ در مورد مثال اندازه‌گیری کوجه‌ها، یه راه غیرعقلانه اینه که شهرداری اندازه واقعی اون ۱۰۰۰۰ تا کوجه رو با یه ابزار اندازه‌گیری مرجع تعیین بکنه؛ تکتک اندازه‌های ما رو با اندازه‌های واقعی مقایسه بکنه و ببینه هر کدوم چند درصد اختلاف

داره؛ اون‌هایی رو که بیشتر از ۱۰٪ اختلاف دارن بشماره و حساب کنه چند درصد از جواب‌های ما درسته؛ و بعد تصمیم بگیره که عملکرد ما درسته یا نه (و این که دستمزد ما رو بده یا نه). این راه، غیرعقلانه‌س. چون اگه شهرداری اندازه واقعی کوچه‌ها رو داشت که دیگه نیازی به قرارداد با شرکت ما نداشت. همینطور در مورد اندازه‌گیری گلوکز، واسه بررسی این که عملکرد سنجشی ما خوبه یا بد، نمی‌شه مثلاً همه جواب‌های گلوکز اون ۱۰۰۰۰ تا کارمند رو با مقدار واقعی‌شون مقایسه کرد. تا معلوم بشه چند درصد از جواب‌های ما درسته و چند درصد نادرسته. **پس چی کار کنیم؟** اینجاس که محاسبه «خطای سنجشی کل؛ خسک» (Total Analytical Error; TAE) به کارمون می‌آد.

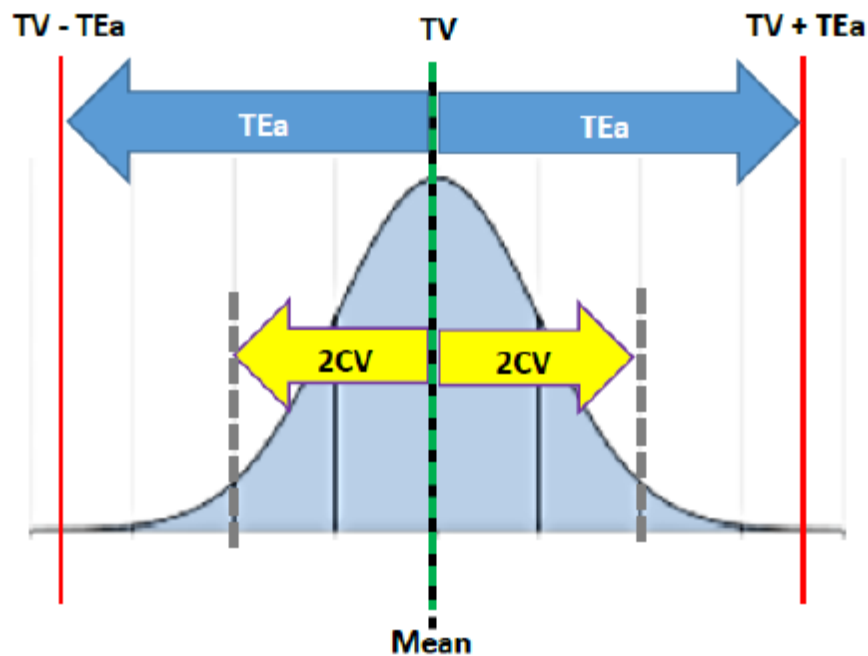
خطای سنجشی کل. بدون وارد شدن تو جزئیات، اگه عدم صحت و عدم دقت (Bias و CV) رو داشته باشیم، می‌تونیم به راحتی تعیین بکنیم که چند درصد از نتایج می‌افته توی محدوده خطای مجاز. حتا این هم لازم نیست که تعیین کنیم دقیقاً چند درصد توی محدوده مجازه، فقط کافیه معلوم کنیم آیا حداقل ۹۵٪ درصد از نتایج می‌افته داخل محدوده مجاز یا نه. همونطور که می‌دونیم، بر اساس توزیع زنگوله‌ای شکل یا گوسی (توزیع طبیعی)، ۹۵٪ از سطح میانی زیر منحنی می‌افته توی فاصله $1.96 \times SD$ کمتر از میانگین تا $1.96 \times SD$ بیشتر از میانگین (یا $1.96 \times CV$ کمتر و بیشتر از میانگین). البته برای راحتی، از ۲ به جای ۱.۹۶ استفاده می‌کنیم (شکل زیر).



اگه Bias و CV طوری باشه که حداقل ۹۵٪ از سطح زیر منحنی (نتایج آزمایش) بیفته توی محدوده مجاز، عملکرد سنجشی ما پذیرفته‌س. توی شکل زیر، میانگین سنجش کاملاً با مقدار هدف (True Value; TV) برابره و بنا بر این عدم صحت یا Bias برابر صفره.

$$\text{Bias} = \text{Mean} - \text{TV} = 0$$

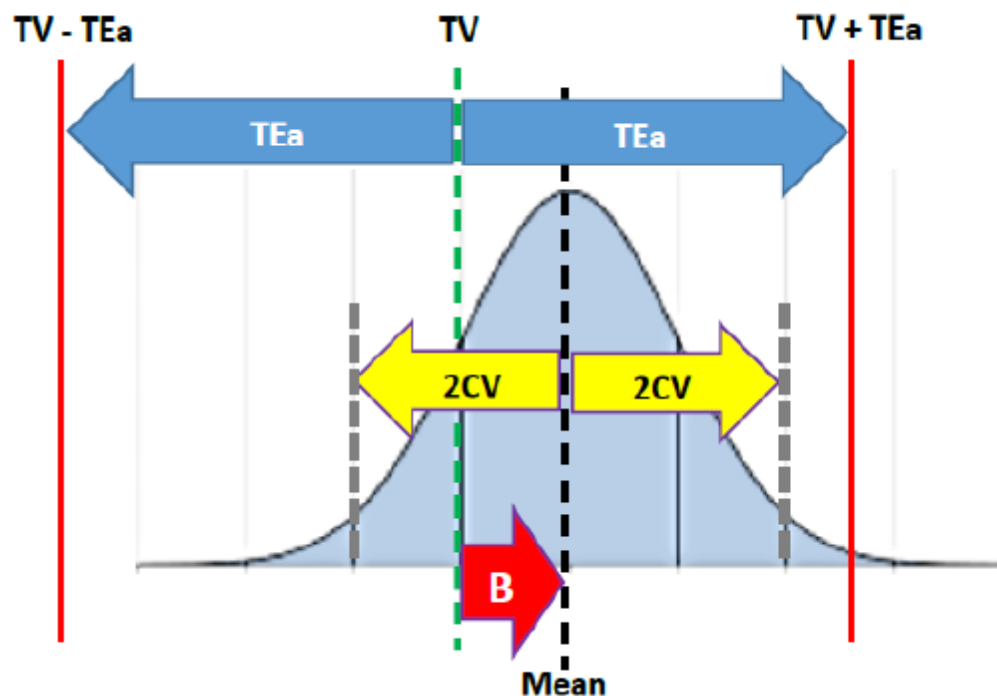
همونطور که تو شکل زیر دیده می‌شه، چون CV از نصف TEa کمتره، بنا بر این $\text{Mean} \pm 2\text{CV}$ که ۹۵٪ از نتایج رو در بر می‌گیره افتاده داخل $\text{TV} \pm \text{TEa}$ و این یعنی حداقل ۹۵٪ از نتایج درسته.



شکل بعدی حالتی را نشون می‌ده که به رغم وجود مقداری عدم صحت (Bias) باز هم چون CV به اندازه کافی کوچیکه، $\text{Mean} \pm 2\text{CV}$ افتاده توی $\text{TV} \pm \text{TEa}$. در نتیجه حداقل ۹۵٪ از نتایج داخل محدوده مجازه و درست به حساب میاد (توی شکل زیر Bias با پیکان قرمز رنگ با حرف B مشخص شده):

$$|\text{Mean} - \text{TV}| > 0$$

$$|\text{Bias}| > 0$$



این دو تا شکل چیزی رو نشون می‌دن که همه‌مون به خوبی می‌دونستیم: تا وقتی که $|Bias| + 2CV$ کوچیکتر از اختلاف مجاز بشه، می‌تونیم مطمئن باشیم که حداقل ۹۵٪ از نتایج می‌افتن توی محدوده مجاز؛ و درست هستن. بنا بر این، عملکرد سنجشی خوبه یا قابل قبوله. این مفهوم، اساس مدل «خطای کل» پروفیسور وستگارد. توی این مدل، حاصل جمع $|B| + 2CV$ رو می‌گن «خطای سنجشی کل» (Total Analytical Error; TAE):

$$TAE = |Bias| + 2CV$$

و اختلاف مجاز (همون ارقامی که توی جداولی مثل CLIA و RCPA و غیره هست) رو می‌گن «خطای کل مجاز» (Allowable Total Error; TEa). با این حساب اگه $|Bias| + 2CV \leq TEa$ باشه یا به طور خلاصه‌تر $TAE \leq TEa$ باشه، اون وقت عملکرد سنجشی قابل قبوله؛ چون درصد جواب‌های قابل قبول مون ۹۵٪ یا بیشتره و درصد خطامون ۵٪ یا کمتره.

وقتی سر و کارمون با یه دونه نتیجه‌اس، حداکثر اختلاف قابل قبول با مقدار واقعی‌ش رو می‌گیم «اختلاف مجاز» یا «انحراف مجاز» یا «خطای مجاز» یا هر اصطلاح مناسب دیگه‌ای که این مفهوم رو برسونه. مثلاً CLIA توی جداول قبلی‌ش می‌گفت «عملکرد قابل قبول» (Acceptable Performance) اما توی جدول پیشنهادی ۲۰۱۹ می‌گه «حدود قابل قبول» (Acceptable Limits)؛ CAP می‌گه «بازه قابل قبول» (Acceptable Range)؛ RCPA می‌گه «خطای مجاز» (Allowable Error)؛ و جدول RiliBak آلمان می‌گه «انحراف قابل قبول» (Acceptable Deviation)؛ منظور همه‌شون هم اینه که در مورد یه دونه نتیجه آزمایش، چقدر اختلاف با مقدار واقعی قابل قبوله. توی مدل خطای

کل، چون قراره حساب کنیم به خاطر مجموع تأثیرات Bias و CV (با هم)، چند درصد از جواب‌هامون نادرست به حساب میان، به همین «انحراف قابل قبول یا اختلاف مجاز» می‌گن «خطای کل مجاز».

خطای سنجشی کل برای «یه دونه» جواب یا برای «عملکرد»؟

چون مدل خطای کل، روی قابل قبول بودن «عملکرد» تمرکز داره نه تک-جواب‌ها، بنا بر این توی این مدل اصطلاح «خطای سنجشی کل» اساساً برای عملکرد استفاده می‌شه، نه تک-جواب‌ها. چون همونطور که گفتیم قرار نیست ما بررسی کنیم هر کدوم از جواب‌ها چقدر با مقدار واقعی اختلاف دارن. با استفاده از این مدل، بر اساس Bias و CV معلوم می‌کنیم آیا حداقل ۹۵٪ از نتایج درست هست یا نه تا بدونیم آیا می‌شه به هر تک-جواب حداقل ۹۵٪ اطمینان کرد که اختلافش با مقدار واقعی کمتر از خطای مجازه. به کار بردن واژه «کل» توی این مدل برای همین بوده که بگن به جای تمرکز روی تأثیر هر کدوم از Bias و CV به تنهایی، تمرکز گذاشته شده روی «کل» خطایی که Bias و CV با هم ایجاد می‌کنن. البته اصطلاح «خطای سنجشی کل» برای اختلاف یه تک-جواب با مقدار واقعی هم به کار برده می‌شه؛ به این معنی که اختلاف با مقدار واقعی ناشی‌یه از تأثیر کل Bias و CV روی هم؛ و البته تأثیر خطاهای پیش‌سنجش، مداخله‌گرها و گاهی هم پس‌سنجش رو هم باید در نظر گرفت. به هر حال، باید یادمون بمونه که توی آزمایشگاه در مورد درستی/نادرستی تک-جواب‌ها قضاوت نمی‌کنیم، بلکه تمرکزمون روی درستی/نادرستی عملکرده و مدل خطای کل در این مورد به کارمون می‌آد.

تکرار سنجش برای کاهش CV

اگه جمعیتی با توزیع گوسی داشته باشیم، با تعداد اعضای N، میانگین μ و انحراف معیار σ ، و از اون جمعیت تمام نمونه‌های n تایی ممکن رو بگیریم و میانگین‌هاشون رو تعیین بکنیم (میانگین نمونه با \bar{X} نشون داده می‌شه)؛ توزیع میانگین‌های اون نمونه‌های n تایی هم گوسی می‌شه، میانگین اون میانگین‌ها یعنی $\mu_{\bar{x}}$ برابر با میانگین جمعیت می‌شه، و انحراف معیار اون میانگین‌ها یعنی $\sigma_{\bar{x}}$ برابر با σ/\sqrt{n} می‌شه:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \sigma/\sqrt{n}$$

مثلاً اگه از یه جمعیت ۱۰۰۰ تایی با میانگین ۵۰ و انحراف معیار ۵، همه نمونه‌های ۲ تایی ممکن رو بگیریم، میانگین و انحراف معیار توزیع نمونه‌های ۲ تایی به ترتیب می‌شه ۵۰ و ۳.۵۴:

$$\mu_{\bar{x}} = \mu = 50$$

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 5/\sqrt{2} = 3.54$$

و اگر همه نمونه‌های ۱۰ تایی رو بگیریم، میانگین و انحراف معیار توزیع نمونه‌های ۱۰ تایی به ترتیب می‌شه ۵۰ و ۱.۵۸:

$$\mu = \mu = 50$$

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 5/\sqrt{10} = 1.58$$

همونطور که می‌بینیم، انحراف معیار توزیع نمونه‌های n تایی از انحراف معیار جمعیت کوچکتره؛ هر چی n بزرگتر، انحراف معیار توزیع نمونه کوچکتر. موقع بررسی‌های آماری، برای این که یافته‌های حاصل از نمونه، برآورد قابل اطمینانی از جمعیت باشه، لازمه n به اندازه کافی بزرگ باشه. در صورتی که n به اندازه کافی بزرگ باشه، می‌شه یافته‌های آماری نمونه رو با اطمینان بالایی به جمعیت تعمیم داد و در باره جمعیت قضاوت کرد. مثلاً اگر برای برآورد میانگین وزن یه جمعیت ۱۰۰۰۰ نفری از بچه‌های ۱۰ ساله، فقط ۱ بچه رو وزن کنیم و وزن اون بچه ۲۷ کیلوگرم، مطمئناً نمی‌تونیم با اطمینان بگیم وزن متوسط این ۱۰۰۰۰ تا بچه تقریباً ۲۷ کیلوگرمه؛ اما اگر ۵۰ تا بچه رو وزن کرده باشیم و میانگین ۲۷ رو به دست آورده باشیم، با اطمینان بیشتری می‌تونیم بگیم وزن متوسط اون جمعیت حدود ۲۷ کیلوگرمه؛ و اگر میانگین وزن یه نمونه ۵۰۰ تایی ۲۷ شده باشه، با اطمینان خیلی بیشتری می‌تونیم میانگین وزن جمعیت رو حدود ۲۷ در نظر بگیریم.

از این موضوع که با افزایش n ، انحراف معیار یا CV توزیع کوچکتر می‌شه، می‌تونیم برای کاهش CV توی اندازه‌گیری‌ها استفاده بکنیم. مثلاً اگر یه ترازو داشته باشیم که CV اون برای کارمون زیاد باشه، می‌تونیم برای کاهش CV، به جای ۱ بار وزن کردن اشیاء، هر چیز رو چند بار (n بار) وزن کنیم و از n تا نتیجه به دست اومده میانگین بگیریم. اگر CV ترازوی ما ۱۵٪ باشه و هر چیز رو ۵ بار وزن کنیم و میانگین بگیریم، CV اندازه‌گیری‌های ۵ تایی برابر ۶.۷٪ می‌شه. اگر CV نتایج تکراری رو با CV_{repeat} نشون بدیم:

$$CV_{repeat} = CV/\sqrt{n} = 15/\sqrt{5} = 6.7\%$$

اگر ۶.۷٪ هم برای کارمون زیاد باشه می‌تونیم تکرار رو بیشتر بکنیم، مثلاً ۲۰ بار:

$$CV_{repeat} = 15/\sqrt{20} = 3.5\%$$

در مورد اندازه‌گیری‌های آزمایشگاهی هم همینطوره. مثلاً اگر یه روش سنجش کراتینین داشته باشیم که CV اون ۵٪ باشه و به جای ۱ بار سنجش، نمونه‌ها رو ۲ یا ۳ بار اندازه‌گیری کنیم و میانگین بگیریم، CV روش ما به ترتیب به ۳.۵٪ و ۲.۹٪ کاهش پیدا می‌کنه:

$$CV_{repeat} = 5/\sqrt{2} = 3.5\% \quad \text{CV حاصل از ۲ بار سنجش:}$$

$$CV_{repeat} = 5/\sqrt{3} = 2.9\% \quad \text{CV حاصل از ۳ بار سنجش:}$$

اگر عدم صحت این روش برابر ۱٪ باشه، خطای سنجشی کل می‌شه:

$$\%TAE = 1\% + (2 \times 5\%) = 11\% \quad \text{۱ بار سنجش:}$$

$$\%TAE = 1\% + (2 \times 3.5\%) = 8\% \quad \text{۲ بار سنجش:}$$

$$\%TAE = 1\% + (2 \times 2.9\%) = 6.8\% \quad \text{۳ بار سنجش}$$

همونطور که در مثال بالا می‌بینیم، با تکرار سنجش‌ها، CV و در نتیجه TAE کم می‌شه.

عدم صحت مجاز، عدم دقت مجاز

تو مدل خطای کل، چیزی به اسم عدم صحت مجاز یا عدم دقت مجاز نداریم. داستان اینه که تا اواسط دهه ۷۰ میلادی، مفهومی به اسم خطای کل سنجش مطرح نشده بود و صحبتی از این نبود که عدم صحت و عدم دقت روی هم چه نقشی در تولید نتایج خطا بازی می‌کنن. سال ۱۹۷۴، پروفیسور وستگارد با ۳ نفر از همکارانش، یه مقاله به کنفرانس AACC ارائه کردن و برای اولین بار مفهوم «خطای کل» رو (که حالا به اون می‌گیم «خطای سنجشی کل») معرفی کردن. بحث وستگارد و همکارانش این بود که به جای این که عدم صحت و عدم دقت رو به طور جداگانه بررسی کنیم، بیائیم «کل» خطای ناشی از Bias و CV رو با هم رو در نظر بگیریم (به صورت $TAE = B + 2CV$). ایده خطای سنجشی کل کم‌کم جا افتاد و تو کنفرانس AACC سال ۱۹۷۶ به عنوان شیوه ارزیابی و پایش کیفیت توی امریکا پذیرفته شد. از اون موقع تا حالا، علاوه بر امریکا، تو کشورهای دیگه مثل آلمان و استرالیا و البته کشور خودمون از این مدل استفاده می‌شه. البته خیلی از کشورهای اروپایی از مدل خطای کل استفاده نمی‌کنن و همچنان عدم صحت و عدم دقت رو به صورت جداگانه در نظر می‌گیرن و برای اون‌ها «عدم صحت مجاز» و «عدم دقت مجاز» تعیین می‌کنن.

توی مدل خطای کل کاری ندارن به این که CV چقدره یا Bias چقدره، بلکه مهم اینه که خطای ناشی از ترکیب این دو تا با هم، یعنی «خطای سنجشی کل»، چقدره و این که «خطای سنجشی کل» از «خطای کل مجاز» بیشتر نشه تا مطمئن بشیم که بیشتر از ۵٪ خطا تولید نمی‌کنیم. توی این مدل، انواع ترکیب‌های عدم صحت و عدم دقت قابل تصویره؛ یه آزمایشگاه ممکنه عدم صحتش بزرگ باشه اما عدم دقتش کوچیک باشه؛ یکی دیگه برعکس، ممکنه عدم دقتش بزرگ باشه اما عدم صحتش کوچیک باشه؛ در مورد یکی دیگه ممکنه هر دو تا متوسط باشه و غیره. اصلاً مهم نیست Bias و CV هر کدوم به تنهایی چقدر هستن؛ چیزی که مهمه اینه که کل درصد جواب‌های نادرست که ناشی می‌شه از Bias و CV روی هم، بیشتر از ۵٪ نشه. به همین دلیل، توی منابع مربوط به مدل خطای کل مثل CLIA، CAP و RCPA و غیره چیزی به عنوان CV مجاز یا Bias مجاز ارائه نشده و فقط خطای کل مجاز (با اسامی مختلف) ارائه شده.

خطای کل مجاز مثل یه بودجه ثابت، مثلاً یه اعتبار خرید صد هزار تومنی برای ما در نظر گرفتن بریم سیب و پرتقال بخریم. اینجا، «صورتحساب کل واسه خرید سیب و پرتقال روی هم»، معادل خطای سنجشی کله؛ و «اعتبار صد هزار تومنی» معادل خطای کل مجازه (بگیم «خرید کل» و «خرید کل مجاز»). گفتن برین واسه خودتون تا سقف صد هزار تومن سیب و/یا پرتقال بخرین؛ کاری ندارن از هر کدوم چقدر می‌خریم، فقط مهمه که خرید کل ما از خرید کل مجاز بیشتر نشه. یه نفر ممکنه بیشتر پرتقال بخره؛ یکی دیگه ممکنه برعکس عمل بکنه؛ یکی ممکنه فقط پرتقال بخره؛ یکی دیگه برعکس؛ یکی ممکنه ته صد هزار تومن رو دربیاره؛ یکی ممکنه متعادل‌تر خرید بکنه؛ یکی هم ممکنه صورتحسابش روی هم خیلی کم بشه و غیره. هر ترکیبی قابل قبوله به شرط این که وقتی می‌ریم پای صندوق،

صورتحسابمون حداکثر صد هزار تومن بشه. به همین شکل، توی مدل خطای کل هر ترکیبی از Bias و CV قابل قبوله به شرط این که خطای سنجشی کل (TAE) از خطای کل مجاز (TEa) بیشتر نشه.

سؤال: پس چرا موقع ارزشیابی روش می‌گن ، CV بیست روزه نباید از یک سوم خطای کل مجاز بیشتر بشه؟

قبل از توضیح در این باره، لازمه به این نکته توجه کنیم که نمی‌شه همه فضای TEa رو به CV و Bias اختصاص بدیم (یعنی نمی‌شه $B + 2CV$ میزون اندازه TEa بشه). چون در اون صورت، دیگه فضایی برای پایش کیفیت (پک) باقی نمی‌مونه. واسه این که نریم تو جزئیات، یه مثال ساده رو در نظر بگیریم: اگه یه ظرف با ظرفیت کل ۱ لیتر داشته باشیم (یعنی TEa برابر ۱ لیتر) مطمئناً میزون ۱ لیتر شیر (یعنی TAE برابر ۱ لیتر) نمی‌ریزیم تو اون گرم کنیم، چون در اون صورت دیگه هیچ فضای خالی‌ای نمی‌مونه که ما بتونیم سر نرفتن شیر رو «پایش» بکنیم. باید یه مقدار از حجم ۱ لیتری ظرف رو بذاریم واسه پاییدن (پایش)؛ هر چی فضای باقیمونده بیشتر باشه، خیال ما راحت‌تره و ساده‌تر پایش می‌کنیم (با فواصل زمانی بیشتر به شیر روی اجاق سر می‌زنیم). تو آزمایشگاه هم همین طوره: هرچه CV و Bias کوچیکتر باشن، TAE از TEa کوچیکتر می‌شه و بنا بر این فضای بیشتری واسه پک باقی می‌مونه و در نتیجه می‌شه از یه برنامه پک ساده‌تر با هزینه‌تر و زحمت کمتر استفاده کرد. این موضوع رو می‌شه با رابطه زیر نشون داد:

$$TAE + QC = TEa$$

رابطه بالا تأکید می‌کنه که از فضای TEa، یه سهمی هم واسه پک لازمه (یه بخش از ظرف TEa رو خالی بذاریم واسه پایش سر نرفتن خطا!). با در نظر داشتن این موضوع، بریم سراغ این که اگه CV بیشتر از یک سوم خطای کل مجاز بشه چطور می‌شه. اگه CV یک سوم خطای کل مجاز بشه:

$$TAE = B + 2CV$$

$$TAE = B + (2 \times 1/3TEa)$$

$$TAE = B + 2/3TEa$$

و در بهترین حالت، اگه Bias برابر صفر بشه:

$$TAE = 0 + 2/3TEa$$

$$TAE = 2/3TEa$$

در این صورت، با در نظر گرفتن سهم پک:

$$TAE + QC = TEa$$

$$2/3TEa + QC = TEa$$

$$QC = 1/3TEa$$

یعنی با CV که یک سوم خطای کل مجازه و $Bias$ برابر صفر، دو سوم از فضای TEa به TAE اختصاص پیدا می‌کند و فقط $1/3$ از فضای TEa واسه پک باقی می‌مونه. با این فضای کم، باید از یه برنامه پک چندقانونی با حداقل ۴ تا کنترل توی هر دور استفاده کرد. که خیلی پرزحمت و پرهزینه‌س. حالا فرض کنیم CV از یک سوم خکم بیشتر باشه و/یا کمی هم $Bias$ داشته باشیم؛ اون وقت فضای خیلی کمتری می‌مونه واسه پک و هزینه و زحمت خیلی بیشتری می‌مونه واسه ما.

البته به طور نظری، اگه یه روش داشته باشیم که کاملاً پایدار باشه طوری که مطمئن باشیم در طول مدتی که از اون استفاده می‌کنیم هرگز پایداریش به هم نمی‌خوره، اون وقت دیگه به پک نیازی نیست و می‌شه همه فضای TEa رو به TAE اختصاص داد. با همچین روش همیشه-پایدار و البته یافت ناشدنی‌ای، اگه از قضا عدم صحت هم برابر صفر بشه، دیگه شانس CV زده و همه فضای TEa می‌مونه واسه اون. در این صورت، CV می‌تونه تا اندازه نصف TEa بزرگ باشه:

$$TAE + QC = TEa$$

$$(B + 2CV) + QC = TEa$$

$$(0 + 2CV) + 0 = TEa$$

$$CV = 1/2TEa$$

به همین شکل، اگه به طور نظری یه روش هم همیشه-پایدار با CV برابر صفر داشته باشیم، اون وقت همه TEa می‌مونه واسه $Bias$ و عدم صحت می‌تونه تا اندازه TEa بزرگ باشه:

$$TAE + QC = TEa$$

$$(B + 2CV) + QC = TEa$$

$$(B + 2*0) + 0 = TEa$$

$$B = TEa$$

از اونجایی که در عمل همچین روش‌های همیشه-پایداری نداریم و بنا بر این همیشه به پک احتیاج داریم، پس موقع ارزشیابی باید مطمئن باشیم که یه سهمی از فضای TEa می‌مونه واسه پک. واسه همین که موقع ارزشیابی یه روش، می‌گن CV از یک سوم TEa بزرگتر نشه به این امید که $Bias$ هم صفر یا خیلی کوچیک از کار درمی‌آد و حداقل نزدیک به یک سوم از TEa می‌مونه واسه پک. توجه داشته باشیم که این گفته به اون معنا نیست که CV مجاز، برابره با یک سوم TEa یا $Bias$ مجاز، برابره با یک سوم TEa .

یه سوال دیگه: پس چرا تو جدول خطاهای مجاز بیولوژیک، هم TEa داریم هم CV مجاز و هم $Bias$ مجاز؟

اول یادآوری نشانه‌ها: توی جدول خطاهای بیولوژیک، CV مجاز و Bias مجاز و خطای کل مجاز به ترتیب با علائم $I(\%)$ ، $B(\%)$ و $TE(\%)$ نشون داده شده.

همونطور که قبلاً گفته شد، $I(\%)$ و $B(\%)$ برای استفاده تو بعضی کشور های اروپاییه که مقادیر CV و Bias رو به طور جداگانه ارزیابی و پایش می‌کنن و کاری هم به خطای سنجشی کل ندارن. اما استفاده از خطاهای مجاز جداگانه وقتی می‌رسه به برنامه پایش کیفیت خارجی (پکخ) با یه اشکال روبرو می‌شه. روش پکخ تو اون کشور ها اینه که تو هر دوره، نمونه‌هایی رو با مقدار هدف مشخص می‌فرستن آزمایشگاه (معمولاً مقدار هدف رو با روش‌های مرجع تعیین می‌کنن). آزمایشگاه هم هر آنالیت رو ۱ بار آزمایش می‌کنه و جواب رو برمی‌گردونه. مشکل اینجاس که با یه تک-جواب نمی‌شه Bias و CV آزمایشگاه رو حساب کرد تا معلوم بشه که از $B(\%)$ و $I(\%)$ بزرگترن یا نه تا بشه در مورد عملکرد آزمایشگاه توی اون دوره تصمیم گرفت. البته عملی هم نیست که از آزمایشگاه‌ها بخوان که تو هر دوره، ماده کنترل رو واسه هر آنالیت، مثلاً ۲۰ بار اندازه بگیرن تا مجری پکخ بتونه CV و Bias اون‌ها رو محاسبه و بررسی بکنه. واسه حل این مشکل، پروفیسور فریزر و پروفیسور پترسن^۱ یه راهکار ارائه دادن با این استدلال که وقتی مجری‌های پکخ فقط به یه تک-جواب از آزمایشگاه محدود هستن بیائیم یه «حد مجاز اختلاف با مقدار هدف» تعیین کنیم با اسم $TE(\%)$ که مجری‌های پکخ بتونن تک-جواب آزمایشگاه رو با اون مقایسه بکنن. تصمیم هم گرفتن مقدار $TE(\%)$ به اندازه‌ای باشه که اگه یه آزمایشگاه با Bias برابر با $B(\%)$ و CV برابر با $I(\%)$ کار می‌کنه (یعنی با حداکثر مقادیر مجاز واسه Bias و CV)، ۹۵٪ احتمال داشته باشه که جوابش بیفته توی محدوده خطای مجاز و فقط ۵٪ احتمال داشته باشه که جوابش به طور کاذب بیفته بیرون از محدوده خطای مجاز. البته هر چی Bias و CV آزمایشگاه از $B(\%)$ و $I(\%)$ کوچیکتر باشه، احتمال این که جوابش به طور کاذب بیفته بیرون از محدوده مجاز از ۵٪ کمتر می‌شه. با این استدلال، اون‌ها فرمول زیر رو واسه محاسبه «حداکثر اختلاف مجاز» برای استفاده توی برنامه‌های پکخ پیشنهاد کردن:

$$TE(\%) = B(\%) + 1.65 * I(\%)$$

اگه اختلاف جواب آزمایشگاه با مقدار هدف یا TV، بیشتر از $TE(\%)$ باشه جواب آزمایشگاه رد می‌شه. بنا بر این، محدوده پذیرش می‌شه:

$$\text{Acceptable Range} = TV \pm TE(\%)$$

مثلاً توی جدول بیولوژیک ۲۰۱۴، $B(\%)$ برای گلوکز برابر با ۲.۳۴٪ و $I(\%)$ برابر با ۲.۸٪. با این مقادیر، $TE(\%)$ می‌شه ۶.۹۶٪:

$$TE(\%) = 2.34\% + 1.65 * 2.8\% = 6.96\%$$

$$\text{Acceptable Range} = TV \pm 6.96\%$$

اگه یه نمونه گلوکز با مقدار واقعی ۱۰۰ رو بفرستن واسه آزمایشگاه‌ها:

$$\text{Acceptable Range} = 100 \pm 6.96\%$$

¹ Callum G. Fraser; Per Hyltoft Petersen

Acceptable Range: From 93 to 107

انتظار دارن جواب آزمایشگاه‌ها بیفته تو محدوده $6.96\% \pm 1.00\%$ ؛ یعنی از ۹۳ تا ۱۰۷. حالا اگه یه آزمایشگاه با Bias برابر 2.34% و CV برابر 2.8% یعنی دقیقاً برابر B(%) و I(%) کار بکنه، 95% احتمال داره جوابش بیفته توی محدوده قابل قبول و 5% احتمال داره که جوابش بیفته بیرون از محدوده قابل قبول. همونطور که گفته شد، بیرون افتادن جواب این آزمایشگاه «رد کاذب» (False Rejection) یا «مثبت کاذب» به حساب می‌آد؛ اما مثل هر برنامه پایش دیگه‌ای، برای این که حساسیت خطایابی خیلی کم نشه، باید مقداری رد کاذب رو تحمل کرد.

یکی از انتقادهای بجایی که به این پیشنهاد وارد شده اینه که ممکنه Bias یه آزمایشگاه بزرگتر از B(%) باشه اما CV اون به اندازه‌ای کوچیکتر از I(%) باشه که جوابش از محدوده $TV \pm TE(\%)$ بیرون نیفته؛ یا برعکس، CV آزمایشگاه بزرگتر از I(%) باشه اما Bias اون به اندازه‌ای کوچیکتر از B(%) باشه که جوابش بیرون نیفته. به عبارت دیگه یه Bias بزرگتر از B(%) با یه CV کوچیک جبران بشه و برعکس. یادمون بیاد که- بر خلاف «مدل خطای کل» که اندازه Bias و CV به تنهایی مهم نیست بلکه تأثیر مشترک اون‌ها به صورت TAE اهمیت داره- توی مدل «خطاهای مجاز جداگانه برای Bias و CV»، باید هم Bias از B(%) بزرگتر نباشه و هم CV از I(%) بزرگتر نباشه. بنا بر این، اگه یکی از Bias یا CV بزرگتر از مقدار مجاز باشه اما توسط کوچیک بودن اون یکی جبران بشه طوری که جواب آزمایشگاه نیفته بیرون از $TE(\%)$ ، سبب می‌شه عملکرد آزمایشگاه به طور کاذب پذیرفته بشه؛ یعنی «پذیرش کاذب» (False Acceptable) و یا «منفی کاذب».

البته خیلی وقت‌ها می‌بینیم که مقادیر $TE(\%)$ بیولوژیک، کنار خطاهای مجاز CLIA و RCPA و غیره به عنوان یکی از منابع TEa برای مدل خطای کل استفاده می‌شه. اگرچه مقادیر $TE(\%)$ در اصل واسه این کار درست نشدن اما استفاده از اون‌ها به این شکل اشکالی نداره. نکته مهم اینه که حواسمون باشه این دو تا رویکرد، یعنی «خطاهای مجاز جداگانه» و «خطای کل»، رو با هم قاطی نکنیم. اشتباهی که توی چند سال اخیر توی ایران دیده می‌شه؛ در حالی که مراجع نظارتی از آزمایشگاه‌ها می‌خوان از مدل خطای کل و قوانین پک و ستگارد- که برای مدل خطای کل طراحی شده- استفاده بکنن، همزمان از آزمایشگاه می‌پرسن: «CV مجازت چقدره؟»!

نکته مهم دیگه اینه که اگر چه $TE(\%)$ بیولوژیک از ترکیب B(%) و I(%) به دست می‌آد، اما همونطور که در بالا توضیح داده شده، این که برای منظور خاصی (یعنی استفاده توسط مؤسسات پکخ تو اروپا) انجام شده و هرگز به این معنی نیست که توی مدل خطای کل هم، CV مجاز و Bias مجاز داریم و از ترکیب اون‌ها، TEa حساب می‌شه. به هیچ عنوان توی توضیحات همراه جدول‌های CLIA و RCPA و غیره خبری از این حرف‌ها نیست. توی هیچ کدوم نه حرفی از CV مجاز و Bias مجاز به صورت جدا هست و نه هیچ فرمولی و راهکاری برای به دست آوردن TEa از CV و Bias به اصطلاح مجاز. مثلاً نویسندگان جدول RCPA توضیح می‌دن که هدفشون این بوده که (نقل به مضمون) «مقادیری برای خطای مجاز ارائه بدن که هم نیاز بالین رو در نظر گرفته باشه، هم سطح فعلی فناوری رو و هم امکانات و منابع آزمایشگاه‌ها رو. برای رسیدن به این هدف، افراد صاحب‌نظر از طرف‌های مختلف شامل پزشک‌ها، تولیدکننده‌های محصولات آزمایشگاهی، آزمایشگاه‌های، برنامه‌های

پک‌خ، نهادهای نظارتی و نمایندگان نهادهای قانونی، دور هم نشستند، بحث و گفتگو کردند و روی مقادیر ارائه شده تو جدول RCPA توافق کردند.»

تأثیر تکرار سنجش بر خطای کل مجاز

مژده: بالاخره بعد از خوندن این توضیح و تفصیل (امیدوارم مربوط)، کم کم داریم می‌رسیم به نقد «خطای کل مجاز تعدیل شده».

خطای کل مجاز، مثل هر حد مجاز دیگه‌ای، یه خط قرمزیه که ما نباید از اون رد بشیم و سعی کنیم تا جایی که می‌شه عملکردمون از اون بهتر باشه تا پایش عملکرد، ساده‌تر و ارزونتر بشه. مثلاً اگه حداکثر دمای مجاز برای سردخونه نگهداری میوه ۱۰ درجه سلسیوس باشه، معنیش اینه که اگه دمای یه سردخونه بیشتر از ۱۰ درجه‌س به اون مجوز فعالیت نمی‌دن چون میوه‌ها توی دمای بالای ۱۰ درجه خراب می‌شن. فرض کنیم یه سردخونه‌دار با دستگاه خنک‌کننده‌ش همه زورش رو می‌زنه و حداقل دمایی که می‌تونه ایجاد بکنه می‌شه ۱۴ درجه. دست آخر تصمیم می‌گیره ۲ تا دستگاه بذاره و از قضا تصمیمش جواب می‌ده و ۲ تا دستگاه می‌تونن با هم دما رو به ۹ درجه برسونن. بعد هم تصمیم می‌گیره برای این که دما پائین‌تر بره تا خیالش از جهت پایش دما راحت‌تر بشه، ۳ تا دستگاه بذاره و اینجوری دما رو می‌رسونه به ۵ درجه. پرسش اینه که آیا، حالا که به عوض ۱ دستگاه خنک‌کننده از ۲ تا یا ۳ تا دستگاه استفاده می‌کنه، باز هم دمای مجاز همون ۱۰ درجه‌س؟ آیا باید دمای مجاز رو برای استفاده از ۲ تا دستگاه، کمی پائین‌تر از ۱۰ درجه در نظر بگیریم و برای ۳ تا دستگاه خیلی کمتر؟ آیا باید به اون سردخونه‌دار مثلاً بگیم با ۲ تا دستگاه، دمای مجاز واسه تو می‌شه ۹ درجه و اگه ۳ تا دستگاه بذاری دمای مجاز برات می‌شه ۴ درجه؟ البته که نه! دمای یه سردخونه باید زیر ۱۰ درجه باشه تا میوه‌ها سالم باشن. ۱ دستگاه بذاره یا ۲ تا یا ۳ تا یا بیشتر، دمای مجاز همون ۱۰ درجه‌س، چون توی دمای کمتر از ۱۰ درجه میوه‌ها سالم می‌مونن؛ فقط همین! اگه به این دلیل که از بیشتر از ۱ دستگاه استفاده می‌شه دمای مجاز رو کمتر از ۱۰ درجه در نظر بگیریم، اون سردخونه‌دار حق داره اعتراض بکنه که کلی هزینه کرده دمای سردخونه‌ش بره زیر خط قرمز ۱۰ درجه، نه این که ما هم از اون طرف، خط قرمز رو پائین‌تر بیاریم، و البته حق داره به ما بگه: «شما گفتین اگه دما از ۱۰ درجه کمتر باشه میوه‌ها سالم می‌مونن. به چه دلیل اگه دما رو با ۱ دستگاه به زیر ۱۰ درجه برسونم میوه‌ها سالم می‌مونن اما اگه با بیشتر از ۱ دستگاه این کار رو بکنم میوه‌ها سالم نمی‌مونن؟ مگه واسه میوه‌ها فرقی داره من با چند تا دستگاه دمای سردخونه‌م رو می‌برم زیر ۱۰ درجه؟»

یه مثال دیگه: فرض کنیم برای کارمندان یه بانک، تأخیر مجاز موقع ورود ۲۰ دقیقه‌س. البته که این مرز مجاز، ربطی نداره به این که یه کارمند با چه وسیله‌ای می‌ره سر کار: پیاده، با دوچرخه، با موتوسیكلت و غیره. به هر شکلی که یه کارمند بره سر کار، مهم اینه که بیشتر از ۲۰ دقیقه تأخیر نداشته باشه. فرض کنیم یه کارمند چند روز پیاده می‌آد و می‌بینه تأخیرش بیشتر از ۲۰ دقیقه می‌شه. واسه همین از روزهای بعد با دوچرخه می‌آد و از قضا مشکلش برطرف می‌شه و تأخیرش حدود ۱۴ دقیقه می‌شه. قطعاً به اون نمی‌گیم حالا که با دوچرخه میایی، پس تأخیر مجاز برای تو تعدیل می‌شه و از ۲۰ دقیقه می‌رسه به ۱۵ دقیقه. یا اگه بعد از مدتی یه موتوسیكلت خرید که هم خیالش از بابت به موقع رسیدن راحت‌تر بشه

و هم کمی زودتر برسه تا با آرامش بیشتر کارش رو شروع بکنه، و تأخیرش شد حدود ۵ دقیقه؛ صداش نمی‌کنیم دفتر به اون بگیم «چون با موتوسیکلت می‌آی تأخیر مجاز برات می‌شه ۷ دقیقه». اگه مرز تأخیر مجاز رو به خاطر استفاده از دوچرخه یا موتوسیکلت کمتر بکنیم (یا به اصطلاح تعدیل بکنیم)، اون کارمند حق داره اعتراض بکنه که برای خرید دوچرخه هزینه کرده و بعد برای خریدن موتوسیکلت هزینه بیشتری کرده که زمان تأخیرش رو کم بکنه نه این که ما هم از اون طرف، تأخیر مجاز رو کمتر بکنیم، و حق داره به ما بگه: «شما گفتین چون اول صبح مشتری‌ها تک و توک هستن اگه تأخیرم کمتر از ۲۰ دقیقه باشه مشکلی پیش نمی‌آد. حالا به چه دلیلی اگه با دوچرخه بیام، بعد از ۱۵ دقیقه مشکل پیش می‌آد و اگه با موتوسیکلت بیام بعد از ۷ دقیقه؟ مگه تعداد مشتری‌های اول صبح به این بستگی داره که من با چه وسیله‌ای خودم رو به بانک می‌رسونم؟»

توی مثال‌های بالا و البته در موارد دیگه، مرز خطای مجاز باید ثابت باقی بمونه؛ فارغ از این که افراد به چه صورتی عملکردشون رو اصلاح می‌کنن تا بتونن معیار خطای مجاز رو برآورده بکنن. ما زحمت بیشتری می‌کشیم، وقت بیشتری صرف می‌کنیم و هزینه بیشتری متحمل می‌شیم که از پس خطای مجاز بر بیائیم؛ نه این که به نسبت زحمت، وقت و هزینه بیشتری که ما متحمل می‌شیم مرز خطای مجاز هم تعدیل بشه. البته که فرمول‌هایی وجود داره برای ارتباط دادن دمای سردخونه با تعداد دستگاه سردکننده و برای ارتباط دادن وسیله‌ای که با آن سر کار می‌ریم با مدت تأخیر؛ اما این فرمول‌ها برای اینه که حساب کنیم در صورت افزایش هزینه و تلاش، عملکرد ما چقدر بهتر می‌شه و خطای ما چقدر کمتر می‌شه. نه این که از این فرمول‌ها استفاده کنیم واسه کاهش و به اصطلاح تعدیل خطای مجاز.

یه بار دیگه بریم سراغ مثال اندازه‌گیری طول کوچه‌ها که به اندازه‌گیری غلظت مواد تو آزمایشگاه نزدیک‌تره. اول معیارهای شهرداری یادمون بیاد: (الف) اختلاف مجاز = ۱۰٪؛ (ب) درصد مجاز برای اندازه نادرست = ۵٪. این نکته هم یادمون بیاد که چه در مورد متر زدن کوچه‌ها و چه در مورد آزمایش نمونه‌ها توی آزمایشگاه، تمرکزمون روی خطای تکتک جواب‌ها نیست، بلکه تمرکزمون روی اینه که مطمئن بشیم **عملکردمون** خوبه؛ به این معنی که اختلاف حداقل ۹۵٪ از جواب‌هامون با اندازه واقعی، کمتر از خطای کل مجازه؛ یا به عبارت دیگه، حداقل ۹۵٪ از جواب‌هامون می‌افته توی محدوده قابل قبول ($TV \pm TEa$).

از منظر یه **جواب تکی**، اگه طول واقعی یه کوچه ۱۰۰ متر باشه و ما ۱۰۸ متر گزارش بکنیم، جواب‌مون درسته چون اختلافش با مقدار واقعی از ۱۰٪ کمتره؛ همین و بس! دیگه کاری ندارن ما ۱ بار متر زدیم، یا واسه کم کردن CV اندازه‌گیری‌مون، چند بار متر زدیم و میانگین گرفتیم. در هر صورت، خطای مجاز واسه اندازه‌گیری طول کوچه‌ها برابره با ۱۰٪ و به هیچ عنوان بسته به دفعات سنجش، کم یا تعدیل نمی‌شه. با هر تعداد تکرار که ما نتیجه رو به دست آورده باشیم، اگه اختلافش با مقدار واقعی کمتر از ۱۰٪ باشه واسه کار شهرداری‌چی‌ها کافی‌یه و اون‌ها رو توی تصمیمات عمرانی‌شون گمراه نمی‌کنه. اگه این جواب ۱۰۸ متر، حاصل مثلاً میانگین گرفتن از ۳ بار اندازه‌گیری باشه، این دلیلی نمی‌شه که شهرداری بگه: «متأسفیم! چون شما ۳ بار متر زدی و میانگین گرفتی، خطای کل مجاز با فرمول mTEa تعدیل می‌شه و می‌شه ۷.۲٪؛ و بر اساس این mTEa، جواب ۱۰۸ شما نادرسته». اگه شهرداری همچین ادعایی بکنه ما حق داریم بگیم «شما گفتین اگه اختلاف جواب ما با مقدار واقعی کمتر از ۱۰٪ باشه، شما تو محاسباتتون واسه تصمیمات عمرانی دچار اشتباه نمی‌شین.

حالا به چه دلیل با ۳ بار متر زدن، اختلاف مجاز باید کمتر از ۷.۲٪ باشد تا شما تو تصمیم‌هاتون دچار اشتباه نشین؟ مگه «حداکثر اختلاف مجاز» واسه اشتباه نکردن شما، به دفعات اندازه‌گیری ما ربط داره؟»

از منظر **عملکرد** هم که به موضوع تعدیل خطای کل نگاه کنیم، با فرض TEa برابر ۱۰٪؛ اگر Bias متر ما ۱٪ و CV متر کردمون ۶٪ باشد، خطای سنجشی کل می‌شه ۱۲٪. که از TEa بزرگتره و عملکردمون قابل قبول نیست. اگر مترمون رو بدیم برامون کالیبر بکنن و Bias صفر بشه، عملکرد بهتر می‌شه اما مشکل کاملاً حل نمی‌شه. چون خطای سنجشی کل می‌شه ۱۱٪ که هنوز هم از TEa بزرگتره. پس تصمیم می‌گیریم هر کوچه رو ۲ بار متر بزنین و میانگین بگیریم. با این کار CV ما می‌شه ۴.۲٪ و خطای سنجشی کل هم می‌شه ۸.۴٪. با این خطای سنجشی کل، بیشتر از ۹۵٪ از نتایج ما درسته و طبق قراری که با شهرداری داشتیم عملکردمون خوبه. مطمئناً قبول نمی‌کنیم که شهرداری بگه «نه، قبول نیست! چون شما ۲ بار متر زدین، پس TEa با فرمول mTEa براتون تعدیل می‌شه و می‌شه ۸٪». اگر همین چیزی بگه ما حق داریم اعتراض کنیم که «چه فرقی می‌کنه ما ۱ بار متر زدیم یا ۲ بار یا چند بار؟ مهم اینه که طبق الزامات کیفیت که توی قرارداد نوشته شده، باید حداقل ۹۵٪ از نتایج ما کمتر از ۱۰٪ با مقدار واقعی اختلاف داشته باشه که با Bias برابر صفر و CV برابر ۴.۲٪ این شرط شما برآورده شده.»

در مورد نتایج آزمایشگاهی هم همین طوره؛ فارغ از این که ۱ بار نمونه رو سنجش می‌کنیم یا چند بار، خطای کل مجاز در هر حالت یکسانه. از منظر **یه جواب تک**، مهم اینه که اون جواب بیشتر از TEa از مقدار واقعیش دور نباشه تا موجب تصمیمات اشتباه نشه. به عنوان مثال، از مثالی که توی صفحه ۵۷ کتاب طرح شده استفاده می‌کنم. با فرض TEa برابر ۷٪ برای سنجش کراتینین؛ مقدار هدف (مقدار واقعی) کراتینین تو یه نمونه ۲.۵ mg/dL بوده و آزمایشگاه ۲.۶ mg/dL گزارش کرده. آیا جوابش درسته؟ اختلاف جواب آزمایشگاه با مقدار هدف برابره با ۰.۱ mg/dL که نسبت به مقدار هدف می‌شه ۴٪ و چون ۴٪ اختلاف از ۷٪ اختلاف مجاز کمتره، این جواب درسته؛ همین! دیگه کاری نداریم به این که جوابی که آزمایشگاه داده حاصل ۱ بار سنجشه یا ۲ بار یا ۱۶ بار (تکرارهای مطرح شده در مثال کتاب). در هر حال جواب ۲.۶ کمتر از ۷٪ با ۲.۵ اختلاف داره و بنا بر این موجب اشتباه در تصمیم‌گیری‌های بالینی نمی‌شه.

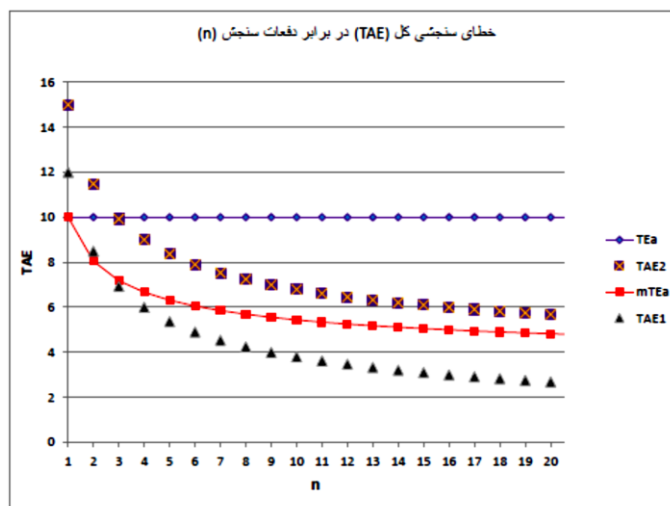
مطمئناً این طور نیست که اگر این جواب ۲.۶ رو با ۱ بار سنجش به دست آورده باشیم، جواب خوبی‌یه، چون ۴٪ اختلاف موجب اشتباه در تصمیم‌های بالینی نمی‌شه؛ اگر با ۲ بار سنجش به دست آورده باشیم، بد نیست چون هنوز هم ۴٪ اختلاف سبب اشتباه بالینی نمی‌شه؛ اما اگر همین ۲.۶ حاصل ۱۶ بار سنجش باشه به درد نمی‌خوره چون دیگه ۴٪ اختلاف خیلی زیاده و شدیداً منجر به اشتباهات بالینی می‌شه. از قضا سرکه انگبین صفرا فزود! این همه تکرار کردیم که کار بهتر بشه، بدتر شد. اینجاس که آدم ساده‌ای مثل من هاج و واج می‌پرسه: «یعنی چی؟ مگه حد و مرز ۷٪ واسه این نیست که اگر اختلاف جواب با مقدار واقعی کمتر از این حد باشه، موجب اشتباهات بالینی نمی‌شه؟ پس چرا ۴٪ اختلاف با ۱ بار سنجش خوبه، با ۲ بار سنجش بدک نیست و با ۱۶ بار سنجش خیلی بده؟» چون mTEa واسه ۲ تکرار شده ۵.۶٪ و برای ۱۶ بار تکرار شده ۳.۵٪؟ خوب موضوع اینه که اساساً نباید همین محاسبه‌ای انجام بشه.

به همین دلیل، یعنی این که هر جور که آزمایشگاه جواب رو به دست بیاره در هر حال خطای مجاز ثابت و نباید بسته به دفعات تکرار آزمایش تعدیل بشه، برنامه‌های پایش کیفیت خارجی مثل CAP، CLIA، RiliBak، UK-NEQAS (بریتانیا) و SKLM (هلند) که برنامه‌های «درستی-بنیان» (Accuracy-Based) هستن و نتایج آزمایشگاه رو با مقدار واقعی حاصل از روش‌های مرجع مقایسه می‌کنن، از آزمایشگاه‌ها نمی‌پرسن چند بار نمونه‌ها رو سنجش می‌کنن تا خطای مجاز رو براشون تعدیل بکنن. در حالی که به آزمایشگاه‌ها گوشزد میکنن و تأکید می‌کنن که فقط در صورتی اجازه دارن نمونه کنترل رو ۲ بار یا ۳ بار یا بیشتر سنجش بکنن و میانگین بگیرن که این کار رو به طور معمول برای نمونه‌های بیمارها هم انجام می‌دن، اما هرگز ادامه نمی‌دن که مثلاً «در صورت تکرار سنجش، تعداد دفعات تکرار رو گزارش بکنین که خطای مجاز برای آزمایشگاه شما تعدیل بشه».

توی مثال بالا، از منظر درستی/نادرستی یه تک-جواب به موضوع نگاه کردیم؛ اما همونطور که پیشتر تکرار شد، توی «مدل خطای کل» روی درستی/نادرستی **عملکرد** تمرکز می‌شه نه تک-جواب‌ها. به عنوان مثال، اگه یه روش سنجش گلوکز داشته باشیم با Bias برابر ۲٪ و CV برابر ۶.۵٪، خطای سنجشی کل می‌شه ۱۵٪ که با معیار CLIA از TEa برابر با ۱۰٪ بزرگتره و این عملکرد قابل قبول نیست. برای بهبود عملکرد، دستگاه رو خوب کالیبر می‌کنیم و Bias رو کاملاً حذف می‌کنیم و در نتیجه TAE می‌شه ۱۳٪ که هنوز هم از TEa بزرگتره. بنا بر این، تصمیم می‌گیریم نمونه‌ها رو ۳ بار تکرار کنیم و میانگین بگیریم. با این ترفند، CV از ۶.۵٪ به ۳.۸٪ کم می‌شه و در نتیجه TAE ما می‌شه ۷.۶٪. حالا می‌تونیم بیشتر از ۹۵٪ جواب‌ها مون رو درست تولید بکنیم و عملکردمون قابل قبوله. خلاصه این که با تحمل زحمت و هزینه ۳ بار سنجش، موفق شدیم از پس تولید حداقل ۹۵٪ جواب درست بریبائیم. حالا هیچ دلیلی نداره به ما بگن «چون ۳ بار سنجش میکنین، TEa واسه شما تعدیل می‌شه و می‌شه ۷.۲٪ و چون خطای کل شما شده ۷.۶٪ پس عملکرد شما قابل قبول نیست». معلومه که اعتراض می‌کنیم و می‌گیم «شما گفتین اگه حداقل ۹۵٪ جواب‌ها مون درست باشه عملکرد ما خوبه، چون پزشک می‌تونه نسبت به هر جواب گلوکزی که ما گزارش می‌کنیم حداقل ۹۵٪ مطمئن باشه که اون جواب درسته. خطای کل ما نشون می‌ده که ما در بیشتر از ۹۵٪ موارد، جواب درست تولید می‌کنیم؛ به چه دلیل اگه ۱ بار سنجش می‌کردیم، پزشک می‌تونست حداقل ۹۵٪ به درستی جواب‌های ما اطمینان بکنه، اما حالا که با ۳ بار سنجش (یا هر چند بار دیگه) کار می‌کنیم، پزشک نمی‌تونه حداقل ۹۵٪ به درستی جواب‌ها اطمینان بکنه؟»

شکل زیر چکیده بحث بالا رو نشون می‌ده. از دو روش ۱ و ۲ برای آزمایشی با خطای کل مجاز ۱۰٪ استفاده شده. CV هر دو روش برابره با ۶٪، عدم صحت روش شماره ۱ صفره، اما عدم صحت روش شماره ۲ برابره با ۳٪. با این داده‌ها، خطای سنجشی کل به ترتیب می‌شه ۱۲٪ و ۱۵٪؛ که برای هر دو روش غیر قابل قبوله. خسک برای روش اول فقط با ۲ تکرار می‌رسه به حد مجاز ۱۰٪ اما برای روش دوم، ۳ تکرار لازمه تا به ۱۰٪ برسه. با تکرارهای بیشتر، خسک هر دو روش به سرعت از ۱۰٪ کمتر می‌شه؛ یعنی با ثابت در نظر گرفتن خطای کل مجاز (لوزی- خط بنفش تو شکل زیر)، افزایش تکرار سبب می‌شه TAE به اندازه کافی از TEa کمتر بشه و فضای بیشتری برای پک باقی بمونه. فضای بیشتر برای پک، به این معنی‌یه که می‌تونیم کیفیت کارمون رو با معیارهای ساده‌تر و با هزینه و زحمت کمتری پایش بکنیم. اما اگه قرار باشه به دلیل تکرار سنجش، خطای کل مجاز رو هم با فرمول mTEa کم بکنیم و از mTEa برای قضاوت در باره کیفیت سنجش استفاده بکنیم (مربع-خط

قرمز)، TAE روش دوم حتماً با ۲۰ تکرار هم از mTEa بالاتر می‌مونه و بنا بر این کیفیت اون غیر قابل قبول به حساب می‌آد. TAE روش اول با ۳ تکرار می‌ره زیر mTEa اما حتماً تا ۲۰ تکرار هم فاصله چندانی از mTEa نمی‌گیره و فضای چندانی برای پاک ایجاد نمی‌شه. نداشتن فضای کافی برای پاک به این معنی‌یه که برای پایش کیفیت باید از معیارهای چند-قانونی سخت استفاده کرد و هزینه و زحمت بیشتری رو متحمل شد.



این موضوع که خطای مجاز ثابت و بسته به عملکرد آزمایشگاه تعدیل نمی‌شه فقط مختص مدل خطای کل نیست، بلکه توی جاهایی هم که به عوض خطای کل مجاز، از CV مجاز و Bias مجاز به طور جداگانه استفاده می‌شه موضوع همینطوره. I(%) و B(%) (CV مجاز و Bias مجاز) رو به آزمایشگاه‌ها اعلام می‌کنن و می‌گن حواستون باشه CV و Bias روش شما از این مقادیر مجاز بیشتر نشه؛ فقط همین! دیگه کاری ندارن شما با چه ترفندی CV و Bias رو می‌برید زیر مرز مجاز تا به اصطلاح مرز مجاز رو براتون تعدیل نکنن.

یه کم توضیح در باره نوسان بیولوژیک. به نظر می‌آد توی وضعیت سلامتی، غلظت مواد داخل مایعات بیولوژیک مثل خون و ادرار و غیره، اطراف یه «نقطه تنظیم» (Set-point) نوسان می‌کنه. به این نوسان می‌گیم «CV درون-فردی» (Intra-individual CV; CV_I). علاوه بر این، نقطه تنظیم همه افراد سالم یکسان نیست و افراد مختلف، نقطه-تنظیم‌های مختلف دارن. به پراکندگی نقطه-تنظیم‌ها بین افراد می‌گیم «CV گروه» (Group CV; CV_G). برای تعیین CVهای بیولوژیک، بسته به ماده مورد نظر، از تعدادی آدم سالم چند بار در زمان‌های مختلف نمونه می‌گیرن و میانگین و CV هر فرد رو تعیین می‌کنن. از CVهای افراد، میانگین می‌گیرن که می‌شه CV_I و از میانگین‌های افراد (که نماینده نقطه-تنظیم‌های افراد)، CV می‌گیرن که می‌شه CV_G .

استدلال پشت CV مجاز بیولوژیک، اینه که وقتی برای پیگیری تغییر در وضعیت یه بیمار، دو تا جواب متوالی رو با هم مقایسه می‌کنیم، CV آزمایشگاه باید اینقدر بزرگ نباشه که نتونیم با اطمینان، تفاوت دو

تا نتیجه رو بذاریم به حساب تغییر در وضعیت بیمار و شک کنیم «نکنه این اختلاف بین دو تا نتیجه به خاطر بزرگ بودن CV آزمایشگاه باشه». پراکندگی نتایج متوالی یه بیمار، در حالی که وضعیت بیمار پایداره و نقطه-تنظیم ثابت، ناشی می‌شه از CV درون-فردی (CV_I) و CV سنجش. با این استدلال، پیشنهاد کردن برای این که CV سنجش خیلی به پراکندگی نتایج اضافه نکنه، مقدار CV سنجش حداکثر نصف CV درون-فردی باشه:

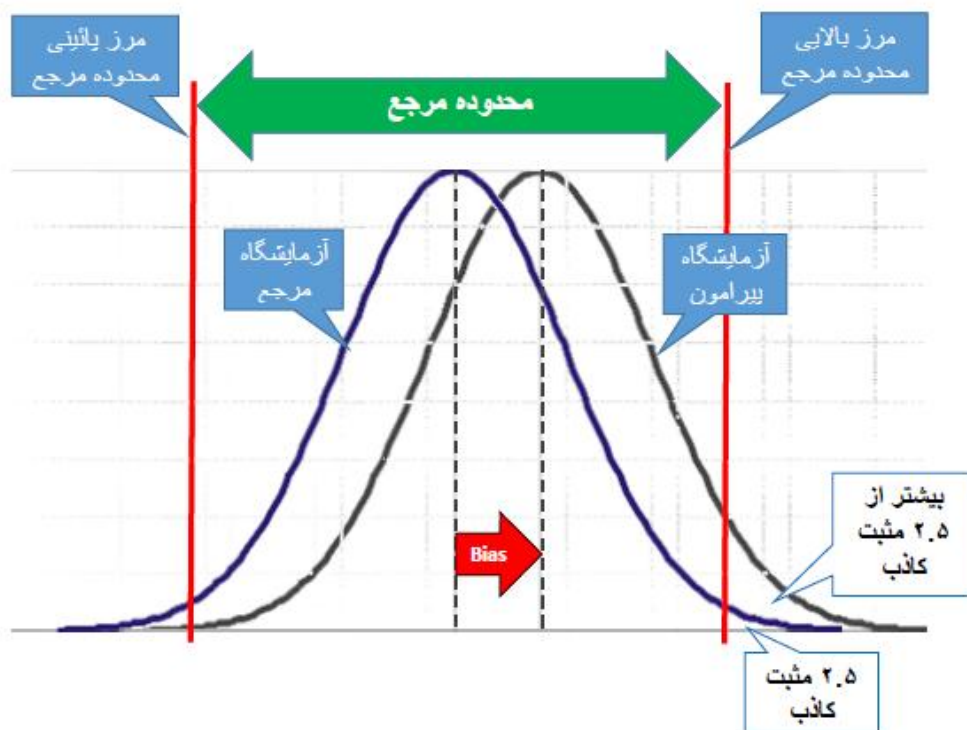
$$I(\%) = 0.5 * CV_I$$

$I(\%)$ رو به آزمایشگاه‌ها اعلام می‌کنن و می‌گن CV تون از این بیشتر نباشه؛ فقط همین. هیچ وقت به آزمایشگاه‌ها مثلاً نمی‌گن «این CV های مجاز واسه وقتی‌یه که نمونه‌ها رو ۱ بار آزمایش می‌کنین؛ اگه بیشتر از ۱ بار سنجش کردین و میانگین گرفتین، باید CV مجاز رو در $1/\sqrt{n}$ ضرب بکنین تا تعدیل بشه».

مثلاً اگه $I(\%)$ برای یه آزمایش ۳% باشه معنیش اینه که تا وقتی که CV آزمایشگاه کمتر از ۳% باشه، می‌شه با اطمینان جواب‌های متوالی بیمار رو با هم مقایسه کرد. حالا اگه CV یه آزمایشگاه بیشتر از ۵% بود و اون آزمایشگاه همه تلاشش رو کرد (شامل تعمیر دستگاه، حذف نوسانات برق، آموزش کاربر، غیره) و نتونست CV رو از ۴% کمتر بکنه، ممکنه به عنوان یکی از آخرین ترفندها تصمیم بگیره نمونه‌ها رو بیشتر از ۱ بار سنجش بکنه تا CV ش از ۴% هم کمتر بشه بره زیر ۳%. مثلاً اگه نمونه‌ها رو ۲ بار سنجش بکنه، CV ش از ۴% می‌شه ۲.۸% و قابل قبول می‌شه. اگه CV یه آزمایشگاه دیگه ۵% باشه و راه دیگه‌ای غیر از سنجش تکراری واسه کم کردن CV نداشته باشه، باید نمونه‌ها رو حداقل ۳ بار تکرار بکنه تا CV بشه ۲.۹% و قابل قبول بشه. کاملاً معلومه که فارغ از این که CV آزمایشگاه به چه شکلی کمتر از CV مجاز شده، بدون تکرار، ۲ بار تکرار، ۳ بار تکرار، یا بیشتر، در هر حال تا وقتی که آزمایشگاه بتونه با CV کمتر از ۳% کار بکنه می‌شه با اطمینان نتایج متوالی بیمار رو با هم مقایسه کرد. اگه بنا باشه آزمایشگاه با سنجش تکراری نمونه‌ها، زحمت و هزینه بیشتری رو متحمل بشه که CV رو پایین بباره، اما بعد به اون بگیم «خوب، حالا باید $I(\%)$ رو هم بسته به دفعات تکرار با استفاده از ضریب $1/\sqrt{n}$ تعدیل بکنی»، عملاً زحمت آزمایشگاه رو دود کردیم. اگه در مورد مثال‌های بالا این کار رو بکنیم؛ با ۲ بار تکرار، CV آزمایشگاه از ۴% می‌شه ۲.۸% اما $I(\%)$ هم کم می‌شه و به عوض ۳% می‌شه ۲.۱%؛ و با ۳ بار تکرار، CV آزمایشگاه از ۵% می‌شه ۲.۳% اما $I(\%)$ هم تعدیل می‌شه و به عوض ۳% می‌شه ۱.۷%. اگه همچین تعدیلی رو پیشنهاد بکنیم، آزمایشگاه حق داره بگه «شما گفتین اگه CV زیر ۳% باشه مشکلی واسه مقایسه نتایج متوالی بیمار پیش نمی‌آد. من هم از راه تکرار سنجش، CV رو بردم زیر ۳%؛ به چه دلیل اگه با ۱ بار سنجش جواب می‌دادم، CV کمتر از ۳% واسه مقایسه نتایج متوالی خوب بود؛ اما با جواب حاصل از بیشتر از ۱ بار سنجش، CV کمتر از ۳% دیگه کافی نیست و باید CV مجاز تعدیل بشه؟ آخه CV مجاز که اون طرف داستان بر اساس نیاز بالین تعیین شده و ربطی به دفعات تکرار من نداره».

استدلال پشت تعیین Bias مجاز بیولوژیک، اینه که آزمایشگاه‌ها هر کدوم به صورت فردی واسه آزمایش‌ها «محدوده مرجع» (Reference Range) (به اصطلاح رایج «رنج نرمال») تعیین نکنن، بلکه یه «آزمایشگاه مرجع» با استفاده از روش مرجع، این کار رو بکنه و بقیه آزمایشگاه‌ها (به اصطلاح «آزمایشگاه‌های پیرامون») اون محدوده مرجع رو تو گزارش جواب‌هاشون استفاده بکنن.

موقع تعیین محدوده مرجع، بعد از نمونه‌گیری از یه تعداد آدم سالم (افراد مرجع) و اندازه‌گیری نمونه‌ها، از نتایج میانگین و انحراف معیار می‌گیرن و بازه‌ای رو که ۹۵٪ از نتایج میانی توی اون قرار داره، به عنوان محدوده مرجع گزارش می‌کنن. با این حساب، محدوده‌های مرجع طوری تعیین می‌شن که در هر طرف، ۲.۵٪ از نتایج آدم‌های سالم می‌افته بیرون و به اصطلاح از هر طرف ۲.۵٪ «مثبت کاذب» داریم. اگه روش آزمایشگاه پیرامون نسبت به روش آزمایشگاه مرجع Bias داشته باشه، اون وقت توی طرفی که Bias هست بیشتر از ۲.۵٪ مثبت کاذب ایجاد می‌شه؛ هر چی Bias بزرگتر، مثبت کاذب تو طرف Bias بیشتر (شکل زیر).



پیشنهاد شده Bias آزمایشگاه پیرامون اون قدر زیاد نباشه که در صد مثبت کاذب توی طرف Bias از ۴.۶٪ بیشتر بشه. چون پراکندگی نتایجی که آزمایشگاه مرجع از نمونه‌های افراد مرجع به دست میاره به منابع نوسان بیولوژیک، یعنی CV_1 و CV_G ، بستگی داره، Bias مجاز $B(\%)$ رو از روی CV_1 و CV_G تعیین کردن:

$$B(\%) = 0.25 \times (CV_1^2 + CV_G^2)^{0.5}$$

در مورد $B(\%)$ هم فقط به آزمایشگاه‌ها می‌گن حواستون باشه عدم صحت‌تون از این بیشتر نشه و دیگه کاری ندارن آزمایشگاه با چه ترفندی عدم صحتش رو می‌بره زیر در صد مجاز تا بر اساس اون $B(\%)$ رو تعدیل بکنن. به هر شکلی که آزمایشگاه بتونه Bias رو برسونه زیر $B(\%)$ ، پزشک‌ها می‌تونن با خیال راحت نتایج اون آزمایشگاه رو با محدوده‌ها مرجعی که از طرف یه مرکز مرجع تعیین شده مقایسه بکنن بدون این که تصمیم‌های بالینی اشتباه واسه بیمار هاشون بگیرن.

نقش تکرار سنجش در عدم قطعیت

باز هم بدون وارد شدن تو جزئیات؛ عدم قطعیت اندازه‌گیری (Measurement Uncertainty, MU) بازه‌ای رو تعیین می‌کنه که با یه احتمال معین، مقدار واقعی توی اون بازه قرار می‌گیره. مثلاً اندازه یه کوچه ۲۰۰ متر با عدم قطعیت ۹۵٪، برابر ۲٪ گزارش شده، این یعنی به احتمال ۹۵٪ طول واقعی کوچه یه مقداری به بین ۲٪±۲۰۰ متر (۱۹۶ تا ۲۰۴ متر) و فقط ۵٪ احتمال داره که طول واقعی کوچه کمتر از ۱۹۶ متر یا بیشتر از ۲۰۴ متر باشه. حالا اگه یه مهندس عمران بخواد مصالح و هزینه‌های لازم واسه سنگفرش کردن اون کوچه رو حساب بکنه، نباید حساب کتابش برای دقیقاً ۲۰۰ متر باشه، بلکه باید عدم قطعیت اندازه‌گیری رو هم در نظر بگیره. هر چی عدم قطعیت اندازه‌گیری‌ها کوچیکتر باشه، خیال کاربری که از نتایج استفاده می‌کنه راحت‌تره و با اطمینان بیشتری تصمیم‌گیری می‌کنه. بر عکس، هر چی عدم قطعیت بزرگتر باشه، اطمینان به نتایج کمتر می‌شه و با شک و تردید بیشتری از اون نتایج استفاده می‌شه؛ مثلاً اگه توی این مثال گفته بودن عدم قطعیت با احتمال ۹۵٪ برابر با ۲۰ درصد، اون وقت اون مهندس عمران باید در نظر می‌گرفت که با احتمال ۹۵٪ طول واقعی کوچه بین ۱۶۰ متر و ۲۴۰ متره (و البته ۵٪ هم احتمال داره که از ۱۶۰ متر کمتر یا از ۲۴۰ متر بیشتر باشه). حالا اگه بخواد واسه سنگفرش اون کوچه حساب کتاب بکنه، باید به قول عمرانی‌ها «تولرانس» زیادی رو در نظر بگیره؛ اگه تولرانسی که باید در نظر گرفته بشه خیلی زیاد باشه، اون وقت ممکنه از تولرانس خود مهندس/پزشک بیشتر بشه و از خیر جواب‌های اون شرکت/آزمایشگاه بگذره و کوچه/مریض رو بفرسته یه شرکت/آزمایشگاه دیگه (به طور کلی، یه «اندازه‌گیری‌کده» دیگه).

داستان عدم قطعیت توی آزمایشگاه هم همینطوره. اگه یه جواب گلوکز رو ۱۱۰ mg/dL گزارش بکنیم با عدم قطعیت ۲۰٪ با احتمال ۹۵٪، این گزارش می‌گه غلظت واقعی گلوکز به احتمال ۹۵٪ یه مقدار یه بین ۸۸ تا ۱۳۲؛ یعنی جواب واقعی ممکنه توی محدوده طبیعی باشه (کمتر از ۱۰۰)، توی محدوده بینابینی باشه (۱۰۰ تا ۱۲۵)، یا توی محدوده دیابتی باشه (بیشتر از ۱۲۵). اگه این رو به یه پزشک بگیریم، مسلماً نمی‌تونه با خیال راحت در باره بیمار تصمیم بگیره؛ چه تشخیصی بذاره: این بیمار از نظر دیابت، مشکلی نداره؟ توی مرحله پیش‌دیابت؟ یا این که دیابت داره؟ اما اگه عدم قطعیت برابر ۲٪ باشه، غلظت واقعی گلوکز به احتمال ۹۵٪ یه مقداری یه بین ۱۰۸ تا ۱۱۲ و پزشک با خیال خیلی راحت‌تری تصمیم‌گیری بالینی می‌کنه. حتا اگه پزشک بیشتر از ۹۵٪ اطمینان بخواد، عدم قطعیت این روش برای احتمال ۹۹.۷٪ برابر با ۳ در صده و این یعنی ۹۹.۷٪ مطمئن باش که غلظت واقعی گلوکز یه مقدار یه بین ۱۰۷ تا ۱۱۳ و فقط ۰.۳٪ احتمال داره که مقدار واقعی کمتر از ۱۰۷ یا بیشتر از ۱۱۳ باشه.

عدم قطعیت سنجش‌های آزمایشگاهی به سه عامل بستگی داره: انحراف معیار سنجش (SD)، عدم قطعیت کالیبراتور (u_{cal}) و عدم قطعیت Bias (u_{Bias})؛

$$u = \sqrt{(SD^2 + u_{cal}^2 + u_{Bias}^2)}$$

به این u می‌گیم عدم قطعیت معیار (Standard Uncertainty) (یه نکته کوچیک: به u می‌گیم عدم قطعیت معیار چون در واقع «انحراف معیار» کلی حاصل از این سه تا عامله). اگه نتیجه رو با R نشون

می‌دیم، با داشتن U می‌تونیم بگیم که مقدار واقعی با اطمینان ۶۸٪ به جایی‌یه تو بازه $R \pm u$ (یه یادآوری کوچیک: توی توزیع گوسی، ۶۸٪ از سطح وسط منحنی می‌افته تو فاصله $\bar{X} \pm 1SD$). چون اطمینان ۶۸٪ نسبتاً کمه، U رو دو برابر می‌کنیم، به اون می‌گیم «عدم قطعیت بسط یافته» (Expanded Uncertainty) و با نشانه U نشون می‌دیم. با داشتن U ، می‌تونیم بگیم مقدار واقعی با اطمینان ۹۵٪ به جاییه توی فاصله $R \pm U$. اگه اجزای طرف راست فرمول U رو به صورت درصدی از نتیجه بنویسیم، عدم قطعیت به صورت درصد، $U\%$ ، حساب می‌شه:

$$\%u = \sqrt{(CV^2 + \%u_{cal}^2 + \%u_{Bias}^2)}$$

$$\%U = 2 \times \%u$$

هر کدوم از سه عامل مؤثر در عدم قطعیت که بزرگ باشه، عدم قطعیت نتایج ما بزرگ می‌شه و جواب ما رو غیر قابل اطمینان می‌کنه. اگه به خاطر CV بالای سنجش، عدم قطعیت ما بزرگ باشه و کار تصمیم‌گیری بالینی رو خراب بکنه، یه راه چاره اینه که با سنجش تکراری نمونه‌ها، CV رو کم بکنیم و در نتیجه عدم قطعیت کوچکتری داشته باشیم. به عنوان مثالی از تأثیر تکرار سنجش روی کاهش عدم قطعیت، از همون مثال ذکر شده تو صفحه ۵۷ کتاب استفاده می‌کنم: یه جواب کراتینین داریم برابر 2.6 mg/dL که حاصل از سه وضعیت مختلفه: ۱ بار سنجش، ۲ بار سنجش و ۱۶ بار سنجش. فرض می‌کنیم CV این روش برابره با ۱۰٪، $\%u_{cal}$ برابره با ۱٪ و $\%u_{Bias}$ هم برابره با ۱٪. عدم قطعیت بسط یافته (U) برای این سه وضعیت به ترتیب می‌شه: ۲۰.۲٪، ۱۴.۵٪ و ۵.۷٪. همونطور که می‌بینیم وقتی برای به دست آوردن جواب، نمونه رو ۱۶ بار اندازه تکرار می‌کنیم، عدم قطعیت ما به مقدار زیادی کم می‌شه (که البته انتظار هم همینه) و بنا بر این اطمینان‌پذیری جواب خیلی بالا می‌ره. با این عدم قطعیت‌ها، در مورد این که غلظت واقعی کراتینین چقدره، می‌شه با اطمینان ۹۵٪ گفت:

- اگه نتیجه 2.6 حاصل ۱ بار سنجش باشه، غلظت واقعی یه مقداری‌یه بین 2.1 تا 3.1 ؛
- اگه نتیجه 2.6 حاصل ۲ بار سنجش باشه، غلظت واقعی یه مقداری‌یه بین 2.2 تا 3.0 ؛ و
- اگه نتیجه 2.6 حاصل ۱۶ بار سنجش باشه، غلظت واقعی یه مقداری‌یه بین 2.5 تا 2.7 .

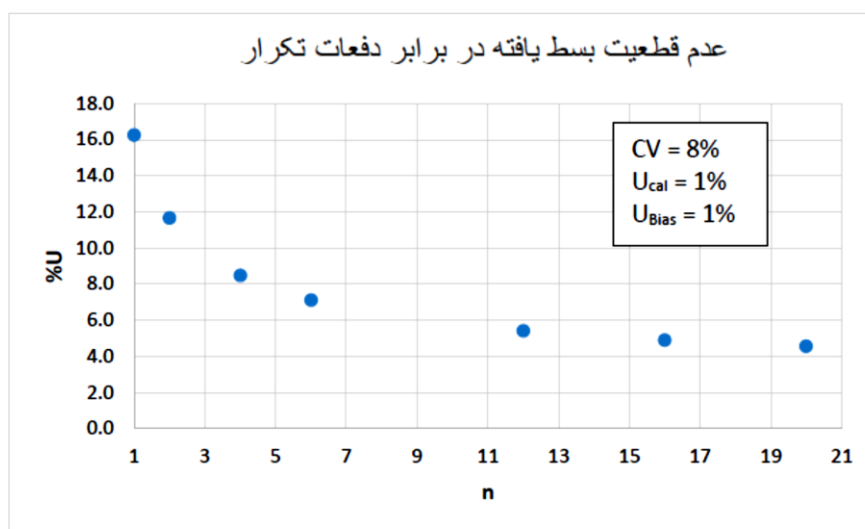
همونطور که می‌بینیم با ۱ بار سنجش، جواب واقعی با اطمینان ۹۵٪ ممکنه تا 0.5 mg/dL دورتر از جواب ما باشه. با ۲ بار سنجش هم وضعیت خیلی بهتر نمی‌شه و جواب واقعی با اطمینان ۹۵٪ ممکنه تا 0.4 mg/dL دورتر از جواب ما باشه؛ اما با ۱۶ بار تکرار، شرایط خیلی فرق می‌کنه و جواب واقعی با اطمینان ۹۵٪ ممکنه فقط 0.1 mg/dL دورتر از جواب ما باشه. اگه ما این اطلاعات رو به یه پزشک بگیم و از اون بپرسیم در کدوم حالت به جواب ما بیشتر اطمینان می‌کنه و با خیال راحت‌تری برای بیمارش تصمیم می‌گیره، مسلماً وضعیت سوم رو ترجیح می‌ده (حتا ممکنه از اون به بعد تو درخواست‌هاش بنویسه «کراتینین با ۱۶ بار تکرار»!).

حالا پرسش اینه که چطور می‌شه گفت جواب حاصل از ۱ بار سنجش خوبه (چون TAE اون به اندازه کافی از TEa کوچیک‌تره)، جواب حاصل از ۲ بار سنجش هم هنوز خوبه اما نه به اندازه جواب حاصل از ۱ بار سنجش (چون TAE اون خیلی از $mTEa$ کوچیک‌تر نیست)، اما جواب حاصل از ۱۶ بار سنجش اصلاً خوب نیست (چون TAE اون از $mTEa$ بزرگ‌تره)؟! هم همچین ادعایی نادرسته، هم هر

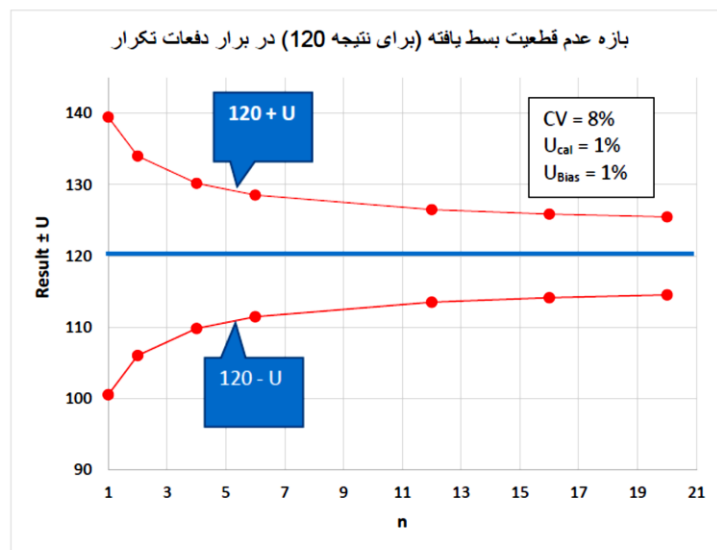
جور فرمولی که درست بشه تا همچین چیزی رو نشون بدیم. بدون نیاز به هیچ فرمول و محاسبه‌ای، می‌شه فهمید که هر چی تکرار اندازه‌گیری بیشتر بشه کیفیت جواب بهتر می‌شه، نه بدتر.

برای برآورد کردن میانگین وزن یه جمعیت ۱۰۰۰۰ نفری، کافی نیست که فقط ۱ نفر رو وزن کنیم و وزن اون رو به عنوان برآوردی از میانگین وزن جمعیت در نظر بگیریم. میانگین یه نمونه ۲ نفری برآورد بهتریه و البته میانگین یه نمونه ۱۶ نفری برآورد خیلی بهتریه. هر چی CV وزن جمعیت بزرگتر باشه، باید نمونه بزرگتری داشته باشیم برای این که برآورد مطمئنی از میانگین وزن جمعیت داشته باشیم. در مورد سنجش‌های آزمایشگاهی هم همینطوره؛ هر چی CV سنجش بزرگتر باشه، تکرار سنجش باید بیشتر بشه تا برآورد مطمئنی از مقدار واقعی داشته باشیم. اگه یه فرمول درست کنیم که با اون نشون بدیم وزن ۱ نفر برآورد بهتریه برای میانگین جمعیت تا میانگین وزن ۱۰ نفر و حتا این که میانگین وزن ۱۶ نفر (یا بیشتر) برآورد غیرقابل قبولی‌یه، باید در اون فرمول که مخالف بدیهیاتیه تجدیدنظر بکنیم.

شکل زیر، کاهش عدم قطعیت برای یک روش با CV برابر ۸٪، U_{cal} برابر ۱٪، و U_{Bias} برابر ۱٪ رو در برابر افزایش تکرار نشون می‌ده.



تو شکل بعدی، عدم قطعیت‌های شکل بالا برای محاسبه بازه عدم قطعیت ۹۵٪ برای یه جواب گلوکز برابر با ۱۲۰ mg/dL استفاده شده. همونطور که دیده می‌شه با افزایش n، بازه عدم قطعیت باریکتر می‌شه.



این که با تکرار اندازه‌گیری جواب مطمئن‌تری به دست می‌آوریم، موضوعی نیست که به بلد بودن ریاضی و آمار و عدم قطعیت نیاز داشته باشد. افرادی هم که سر و کارشون با این جور مباحث نیست از تکرار اندازه‌گیری برای اطمینان بیشتر استفاده می‌کنن. خیلی وقت‌ها شاهد صحنه‌هایی مثل این بودیم که مثلاً یه نجار از شاگردش خواسته طول یه تخته رو اندازه بگیره. وقتی که شاگرد اندازه تخته رو گفته، نجار پرسیده «مطمئنی اندازه درسته؟» و شاگرد گفته «آره، ۲ بار اندازه گرفتم». وقتی با ۲ بار اندازه گرفتن خیال یه نجار راحت می‌شه، با بیشتر اندازه گرفتن خیالش خیلی راحت‌تر می‌شه؛ نه این که بگه «اندازه قابل اطمینان نیست چون ۱۰ بار اندازه گرفتی». اگه ۱ بار اندازه زده بودی، جوابت مطمئن‌تر بود.»

تکرار سنجش و نوسان بیولوژیک درون-فردی (CVI)

وقتی قراره دو تا جواب پشت سر هم رو با هم مقایسه بکنیم، فقط بزرگ بودن CV سنجش نیست که مشکل‌ساز می‌شه، بلکه اگه CV_I (بیولوژیک درون-فردی) هم بزرگ باشه، مشکل ایجاد می‌کنه. با اقدامات اصلاحی مختلف می‌شه CV سنجش رو کم کرد، مثل نصب پایدارساز واسه کاهش نوسان برق، نو کردن قطعات کهنه مثل سرنگ‌های مکنده، و غیره، و دست آخر سنجش تکراری نمونه‌ها. شاید هم کلاً بی‌خیال اون دستگاه/ کیت بشیم و اون رو با دستگاه/ کیت بهتری جایگزین بکنیم. اما در مورد CV_I راه چاره‌چندانی نداریم؛ مثلاً امکانش نیست که واسه بیمار پایدارساز بذاریم یا روی بیمار کار اصلاحی/ تعمیراتی انجام بدیم؛ همینطور نمی‌شه (مثل عوض کردن دستگاه/ کیت) به بیمار بگیم اصلاً بی‌خیال اون آزمایش بشه و به جای اون یه آزمایش با CV_I کوچیکتر براش انجام بدیم. تنها راه چاره‌ای که برای مشکل بزرگ بودن CV_I داریم «تکرار» کردنه. باید بیمار بیشتر از ۱ بار و با فاصله زمانی آزمایش بده و پزشک بر اساس میانگین نتایج تصمیم بگیره. به همین دلیل که مثلاً در مورد آزمایش چربی‌ها، برای این که جواب آزمایش برآورد خوبی از نقطه- تنظیم باشه توصیه می‌شه بیمار طی ۱ ماه، ۲ بار یا ۳ بار آزمایش بده و میانگین نتایج، ملاک تصمیم‌گیری پزشک باشه. مثلاً با در نظر گرفتن CV_I برابر ۶٪ و CV سنجشی برابر ۲٪ برای کلسترول، اگه از بیمار فقط ۱ بار نمونه بگیریم و نتیجه بشه ۱۲۰ mg/dL

۲۲۰، نقطه- تنظیم کلاسترول بیمار با احتمال ۹۵٪ به جایی‌به بین ۱۹۴ و ۲۴۶ (فاصله ۱۲٪ اطراف میانگین) و آگه ۲ بار نمونه بگیریم و میانگین نتایج بشه ۲۲۰، نقطه- تنظیم کلاسترول به احتمال ۹۵٪ به جایی‌به بین ۲۰۲ و ۲۳۸ (فاصله ۸٪ اطراف میانگین). آگه دفعات نمونه‌گیری از ۲ بار هم بیشتر بشه، بازه‌ای که نقطه- تنظیم داخلش قرار می‌گیره باریک‌تر می‌شه؛ مثلاً آگه ۱۶ بار از بیمار نمونه بگیریم و از نتایج میانگین بگیریم، نقطه- تنظیم این بیمار با احتمال ۹۵٪ به جایی‌به بین ۲۱۳ و ۲۲۷ (فاصله ۳٪ پیرامون میانگین). همونطور که می‌بینیم، هر چی دفعات تکرار نمونه‌گیری رو بیشتر می‌کنیم (مثل تکرار سنجش به نمونه)، بیشتر می‌تونیم به نتایج اطمینان بکنیم و این یعنی کیفیت جواب بهتر شده، نه بدتر. مطمئناً در این مورد هم نمی‌شه بگیم هر چی دفعات نمونه‌گیری بیشتر بشه، کیفیت جواب پائین‌تر می‌آد چون خطای مجاز تعدیل شده‌ش کوچکتر می‌شه.

تأثیر تکرار سنجش بر عیار سیگما و پایش کیفیت

همونطور که می‌دونیم عیار سیگما (Sigma Metric; SM) کیفیت تولید رو نشون می‌ده و از فرمول زیر حساب می‌شه:

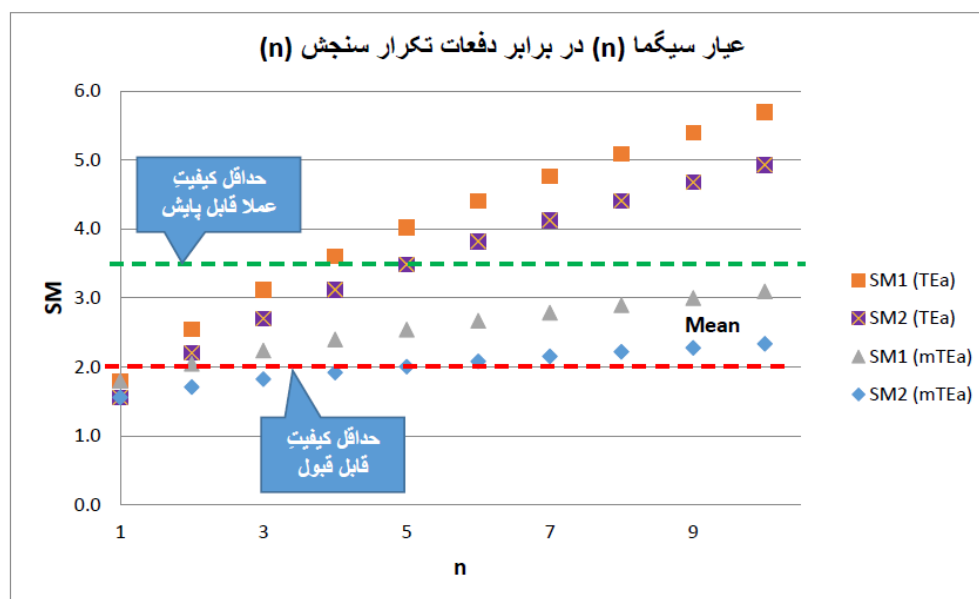
$$SM = (TEa - |B|)/SD$$

یا

$$SM = (\%TEa - |\%B|)/CV$$

هر چی عیار سیگما بالاتر باشه، درصد محصولات ناقص کمتره؛ در مورد آزمایشگاه، هر چی عیار سیگما بزرگتر باشه، جواب‌های نادرست ما کمتره. مثلاً با عیار سیگمای ۲، حدود ۵٪ از جواب‌های ما می‌افته بیرون از محدوده مجاز، اما با عیار سیگمای ۶ فقط ۲ تا از ۱ میلیارد جواب می‌افته بیرون از محدوده مجاز. لازمه توجه داشته باشیم که عیار سیگما هم در مورد «عملکرد» اطلاعات می‌ده نه در باره «تک-جواب‌ها». هر چی عیار سیگما بزرگتر باشه، عملکرد ما بهتره به این معنا که درصد جواب‌های نادرست ما کمتره؛ پس می‌شه به درست بودن هر تک-جواب اطمینان بیشتری داشت. همونطور که توی فرمول عیار سیگما می‌بینیم، CV تو مخرج فرموله و بنا بر این، برای به TEa و Bias معین، هر چی CV کوچکتر بشه به همون نسبت عیار سیگما بزرگتر می‌شه. بنا بر این، آگه عیار سیگمامون کوچک باشه، به عنوان به گزینه می‌شه نمونه‌ها رو به صورت تکراری آزمایش کرد تا CV کوچکتر بشه و به تبع اون SM بزرگتر بشه. شکل زیر عیار سیگمای دو تا روش سنجش مختلف رو برای آنالیتی با TEa برابر ۹٪ نشون می‌ده. CV هر دو روش برابر با ۵ درصده، Bias روش شماره ۱ صفره اما Bias روش شماره ۲ برابر با ۳ درصده. با این مشخصات، عیار سیگمای این دو روش به ترتیب برابره با ۱.۸ سیگما و ۱.۶ سیگما. همونطور که دو تا نمودار بالایی توی شکل زیر نشون می‌ده، آگه نمونه‌ها به صورت تکراری آزمایش بشه، با زیاد شدن n، عیار سیگما با ثابت گرفتن TEa به طور قابل ملاحظه‌ای بزرگتر می‌شه. مثلاً عیار سیگمای روش اول [SM1 (TEa)] با ۵ تکرار می‌رسه به ۴ سیگما و با ۱۰ تکرار می‌رسه به نزدیک ۶ سیگما. عیار سیگمای روش دوم [SM2 (TEa)] با ۵ تکرار می‌رسه به ۳.۵ سیگما و با ۱۰ تکرار می‌رسه تقریباً به ۵ سیگما. در

مقابل، آگه بنا باشه به دلیل تکرار، خطای کل مجاز رو تعدیل بکنیم و mTEa رو بذاریم توی فرمول عیار سیگما، زیاد کردن تکرار نمونه‌ها تأثیر چندانی روی بالا رفتن عیار سیگما نمی‌ذاره. مثلاً عیار سیگمای روش اول [SM1 (mTEa)] با ۵ تکرار می‌رسه به ۲.۵ سیگما و با ۱۰ تکرار می‌رسه به ۳.۵ سیگما؛ عیار سیگمای روش دوم [SM2 (mTEa)] با همین تعداد تکرار به ترتیب می‌رسه به ۲ سیگما و ۲.۳ سیگما.



یادمون بیاد که کمترین عیار سیگمای قابل قبول برابره با ۲ سیگما (چون با ۲ سیگما، ۵٪ از جواب‌های ما نادرسته؛ آگه عیار سیگما از ۲ کمتر بشه، درصد جواب‌های نادرست ما از ۵٪ بیشتر می‌شه). همینطور یادمون بیاد که هر چی عیار سیگما بزرگتر باشه، پایش کیفیت راحت‌تر می‌شه. مثلاً برای پک یه روش ۶ سیگما کافی‌یه فقط یه کنترل تو هر دور بذاریم، از معیار ساده 1:3s استفاده بکنیم و (بر اساس مدل (MaxE (Nuf) برای اطمینان از این که از وقتی که مشکلی پیش می‌آد تا وقتی که برنامه پک اون رو شناسایی می‌کنه حداکثر ۱ جواب اشتباه اضافی گزارش شده باشه، می‌تونیم به ازای هر ۳۸۰ تا نمونه بیمار ۱ نمونه کنترل بذاریم (یعنی بعد از خوندن هر ۳۸۰ تا نمونه بیمار، یه بار هم نمونه کنترل رو بخونیم). برای پک یه روش ۵ سیگما، می‌شه از همون ۱ کنترل و 1:3s استفاده کرد اما باید بعد از هر ۵۰ تا نمونه بیمار یه بار کنترل رو خوند، یا این که ۲ تا کنترل تو هر دور بذاریم، معیارمون همون 1:3s باشه و به ازای هر ۲۸۰ تا نمونه بیمار یه بار کنترل‌ها رو بخونیم. در مقابل، برای پک یه روش ۳.۵ سیگما، باید ۴ کنترل تو هر دور بذاریم (خوندن ۲ سطح کنترل ۲ بار)، معیار چند-قانونی و پرزحمت 1:3s/2of3:2s/R:4s/4:1s رو به کار ببریم و به ازای هر ۴۰ تا نمونه بیمار ۲ تا کنترل رو ۲ بار بخونیم (یعنی حدود ۱۰٪ از کار دستگاه و معرف‌های ما واسه خوندن کنترل‌ها مصرف بشه). پایش کیفیت یه روش ۳ سیگما از این هم پرهزینه‌تره: با همون ۴ تا کنترل و همون معیار چند-قانونی، باید به ازای هر ۱۱ نمونه بیمار هر ۲ کنترل رو ۲ بار بخونیم (یعنی تقریباً ۳۰٪ از هزینه و زحمت ما صرف کنترل خوندن می‌شه). به خاطر همین که اگرچه روش‌های با عیار سیگمای ۲ تا ۳.۵ سیگما

«قابل قبول» به حساب میان، اما عملاً «غیر قابل کنترل» به حساب میان چون پک اون‌ها عملی و مقرون به صرفه نیست و آگه از اون‌ها استفاده کنیم باید امیدوار باشیم که اشکالی برایشون پیش نیاد (ایشالا گربه‌س!) چون با برنامه‌های معمول پک، اون اشکالات به سرعت شناسایی نمی‌شن. با این اوصاف، در مورد روش ۱ و ۲ در مثال بالا، در صورت به کار بستن تعدیل خطای کل مجاز، حتا ۱۰ بار تکرار هم کمکی نمی‌کنه چون کیفیت روش شماره ۱ رو از ۱.۸ سیگما می‌رسونه به ۳.۵ سیگما که برای پک اون باید کلی هزینه و زحمت صرف بکنیم؛ و روش شماره ۲ رو از ۱.۶ سیگما می‌رسونه به ۲.۳ سیگما که عملاً قابل پایش نیست.

چند تا نکته فرعی

گذشته از این که اساساً ایده تعدیل خطای مجاز اشتباهه و اصل موضوع اینه، اما روش به دست آوردن فرمول mTEa هم خالی از اشکال نیست و بد نیست نگاهی هم به اشکالات جنبی بندازیم.

اشکال اول اینه که نویسنده با بیان این که "اگر میزان مجاز هر یک از پارامترهای Bias و CV را معادل یک سوم TEa در نظر بگیریم" مقادیر مجاز جداگانه برای Bias و CV قائل می‌شه؛ در حالی که، همونطور که به طور کامل توضیح داده شد، همچین چیزی توی مدل خطای کل نداریم. اساساً مدل خطای کل ارائه شد که به جای صحبت از خطاهای مجاز جداگانه برای Bias و CV، خطای حاصل از این دو تا روی هم بذاریم بشه خطای سنجشی کل و مقدار مجاز برای این خطای کلی تعیین کنیم (TAE و TEa). واژه «کل» توی اسم این مدل به همین دلیله و به این موضوع اشاره داره. مثال صد هزار تومن اعتبار واسه خرید سیب و پرتقال رو یادمون بیاریم. نمی‌شه بر اساس مدل خطای کل، به یه نفر گفت برو با این صد هزار تومن «خرید کل مجاز»، با هر ترکیبی که دوست داری سیب و پرتقال بخر و فقط حواست باشه «خرید کل» تو از صد هزار تومن بیشتر نشه و در عین حال به اون گفت «خرید مجاز سیب» اینقدره و «خرید مجاز پرتقال» اونقدر. نمی‌شه توی مدل خطای کل از خطاهای مجاز جداگانه برای Bias و CV صحبت کرد. یا رومی روم، یا زنگی زنگ؛ یا باید مثل بعضی کشورهای اروپایی‌ها و همینطور مثل امریکای قبل از ۱۹۷۶، رویکرد خطاهای جداگانه برای Bias و CV رو انتخاب کرد، یا باید مثل امریکایی‌ها و بعضی کشورهای دیگه، رویکرد خطای کل رو انتخاب کرد.

اشکال بعدی اینه که نویسنده، فرمول mTEa رو از فرمول $\text{TAE} = \text{Bias} + 2\% \text{CV}$ به دست می‌آره. نکته اینجاس که توی فرمول TAE، ضریب ۲ یا دقیقتر بگیریم ۱.۹۶ برای وقتی‌یه که Bias برابر صفر باشه. در صورتی که Bias صفر باشه، ۹۵٪ از سطح میانی زیر منحنی می‌افته تو فاصله $\pm X$ 2% CV. اما در صورتی که Bias داشته باشیم، باید بسته به مقدار Bias، این ضریب رو کاهش بدیم. آگه فرمول کلی رو به صورت $\text{TAE} = \text{Bias} + Z\% \text{CV}$ بنویسیم، هر چی Bias از صفر برگتر بشه، باید Z از ۱.۹۶ کوچیکتر بشه. مثلاً وقتی که Bias برابر CV می‌شه باید Z برابر با ۱.۶۵ استفاده بکنیم. بعد از این که Bias از CV بزرگتر می‌شه، سرعت کاهش Z به ازای افزایش Bias خیلی کم می‌شه تا جایی که عملاً با ۱.۶۵ خیلی فرق نمی‌کنه. به خاطر همینه که تو منابع جدید، با فرض وجود مقداری Bias (حداقل به اندازه CV)، از Z برابر با ۱.۶۵ توی فرمول خطای سنجشی کل استفاده

می‌شه. و باز به همین دلیل که برای محاسبه «خطای کل مجاز بیولوژیک» از Z برابر با ۱.۶۵ اتوی فرمول $TE(\%) = B(\%) + 1.65 \cdot I(\%)$ استفاده می‌شه تا وجود Bias در نظر گرفته شده باشه. از قضا، نویسنده فرض رو بر این می‌ذاره که "Bias و CV% هر دو معادل یک سوم TEa%" هستن. با این فرض، Bias با CV مساوی‌یه و بنا بر این، دیگه ضریب ۲ مناسب نیست و باید از ضریب ۱.۶۵ استفاده بشه. با این ضریب، فرمول متفاوتی واسه mTEa به دست می‌آد:

$$\%TAE = \%Bias + 1.65 \cdot \%CV$$

$$\%mTEa = \%Bias + 1.65 \cdot (\%CV / \sqrt{n})$$

$$\%mTEa = 1/3 \cdot \%TEa + 1.65 \cdot [(1/3 \cdot \%TEa) / \sqrt{n}] = 1/3 \cdot \%TEa + (1.65 \cdot \%TEa) / (3 \cdot \sqrt{n})$$

$$\%mTEa = 2.65 \cdot \%TEa / 3 \sqrt{n}$$

با این فرمول، اگه تکرار در کار نباشه و نمونه‌ها فقط ۱ بار سنجش بشن ($n = 1$):

$$\%mTEa = 2.65 \cdot \%TEa / 3 \sqrt{1}$$

$$\%mTEa = \sim 0.88 \cdot \%TEa!!!$$

با همچین فرمولی برای mTEa (که برای به دست آوردنش از مقدار درست برای Z استفاده شده) باید مدعی بشیم حتا بدون تکرار هم، باید خطای کل مجاز رو تعدیل کنیم (TEa رو در ۰.۸۸ ضرب بکنیم)! نکته بعدی اینکه که با شیوه‌ای که نویسنده برای رسیدن به فرمول کلی mTEa استفاده می‌کنه، نمی‌شه برای تعدیل خطای کل بیولوژیک، $TE(\%)$ ، فرمول کلی به دست آورد. این نکته از این نظر جالبه که $TE(\%)$ تنها خطای کلی‌یه که از Bias مجاز و CV مجاز حساب می‌شه:

$$TE(\%) = B(\%) + 165I(\%)$$

دلیل اینکه که $B(\%)$ و $I(\%)$ هر کدام معادل یک سوم $TE(\%)$ نیستن و بنا بر این $B(\%)$ و $I(\%)$ با هم برابر نیستن. هر آنالیت نسبت‌های خاص خودش رو داره و $B(\%)$ و $I(\%)$ برای هر آنالیت، کسر خاصی از $TE(\%)$ رو تشکیل می‌ده و بنا بر این، نمی‌شه به یه فرمول کلی mTEa برای مدل بیولوژیک رسید.

البته باز هم تأکید می‌کنم که اصل موضوع اینکه که اساساً ایده تعدیل خطای مجاز اشتباهه؛ یعنی اساساً «چرایی» تعدیل خطای کل زیر سؤاله؛ و این سه نکته که گفتم اشکالاتی فرعی‌یه توی «چگونگی» به دست آوردن فرمول برای این تعدیل.

کلام آخر

با توضیحاتی که ارائه شده، بر اساس بدیهیات نباید خطای مجاز رو بسته به تلاشی که برای برآورده کردن اون می‌شه، تعدیل کرد. اول باید «چرایی» یه اقدام درست باشه، بعد بشینیم واسه «چگونگی» اون فرمول درست کنیم. اگه فقط ارائه فرمول کافی باشه، مثلاً اون‌هایی که ریاضی بلدن می‌تونن واسه سرگرمی با یه سری فرمول نشون بدن که ۱ با ۱ مساوی نیست. البته هم اون ریاضی‌بلدها و هم آدم‌های ریاضی‌نابلد مثل من، مساوی بودن ۱ با ۱ رو نمی‌پذیریم و می‌دونیم یه جایی توی اون فرمول‌ها یه نکته‌ای هست که به چشم نمیاد.

علاوه بر این، اگه بنا باشه خطای کل مجاز رو تعدیل بکنیم، اون وقت دستاوردمون از تکرار سنجش نمونه‌ها چی می‌شه؟ غیر از این که خطای مجاز، بیشتر و بیشتر آب بره تا جایی که از کوتاه شدن قد خطای ما جلو بزنه و عملکرد ما غیر قابل قبول بشه؟ غیر از این که احتمال پذیرفته نشدن نتایجمون بیشتر بشه؟ و غیر از این که عیار سیگمای روش‌مون تغییر چندانی نکنه، طوری که یا همچنان کیفیت روش ما قابل قبول نباشه یا این که اگه به سختی قابل قبول می‌شه، اینقدر جزئی باشه که قابل پایش نباشه (و عملاً قابل استفاده نباشه)؟ با این اوصاف، منطقی‌یه بپرسم چرا اصلاً باید اندازه‌گیری‌ها رو به شکل تکراری انجام بدیم؟ چرا بیشتر زحمت بکشیم و بیشتر هزینه کنیم و دست آخر بدهکارتر بشیم؟ خوب از اول از خیر، ببخشید از شر، تکرار می‌گذریم.

نتیجه این که، $mTEa$ ایده نادرستی‌یه که از پیش‌فرضی نادرست مشتق شده و استفاده از اون برای ارزیابی تک-جواب‌ها یا عملکرد، و همینطور هر کاربرد دیگه‌ای شامل استفاده از $mTEa$ توی بررسی خطی بودن، به نتیجه‌گیری‌های اشتباه منجر می‌شه.

تشکر

از آقای محمد طالبی همکار علاقمند به مباحث پایش کیفیت، شاغل در آزمایشگاه‌های بیمارستان رازی و دکتر طیبی (قائم‌شهر) به خاطر ویرایش این نوشته سپاسگزارم. علاوه بر این، نامبرده با مطالعه جستجوگرانه و پرسشگرانه کتاب یاد شده و طرح نکاتی از کتاب با من، این انگیزه را در من به وجود آورد که این نقد را بنویسم.