amc technical brief

Analytical Methods Committee AMCTB No.74 Revised May 2016

z-Scores and other scores in chemical proficiency testing-their meanings, and some common misconceptions

z-score و سایر score ها در آزمون های مهارت شیمی – معانی آنها و برخی از ابهامات رایج

ترجمه : شرکت شیمی دانان ویرا

z-score *ها برای فراهم آوردن یک سیستم امتیازدهی شفاف و گسترده برای آزمایشگاه های تجزیه شرکت کننده در آزمون های مهارت ایجاد شدند. ایده اصلی این شاخص عملکرد، ارائه یک مقیاس مناسب برای بیان تفاوت بین نتیجه شرکت کننده و "مقدار تخصیص یافته" برای غلظت آنالیت است. تفسیر* ***z-score*** *به سادگی امکان پذیر است، اما برخی جنبه های آن نیاز به توجه دقیق دارند تا از بروز ابهام جلوگیری شود. در طول زمان، شاخص های عملکرد متعددی برای مواجهه با گستره متنوعی از کاربردها ایجاد شده است. شاخصه های اصلی عملکرد به تازگی در استاندارد (2015) ISO 13528 کدگذاری شده اند.*

آزمون های مهارت، مطالعات بین آزمایشگاهی منظمی هستند که برای شناسایی هرگونه عملکرد فاقد صحت معنی دار در نتیجه گزارش شده توسط شرکت کنندگان طراحی شده اند. در زمانی که امکان پذیر باشد، نتایج از طریق شاخص های عملکرد ارزیابی شده که هدف از آن ارائه یک مبنا برای آغاز اقدام اصلاحی در صورت لزوم است. در ابتدا، روش های امتیازدهی متنوعی بر اساس تفسیرهای دلخواه از نتیجه وجود داشت. با این حال، به سرعت مشخص شد که یک روش امتیازدهی ساده و یکپارچه، به تحلیلگران امکان می دهد که یک امتیاز را به صورت یکسان در مواد آزمایشی مختلف، آنالیت ها، بازه های غلظت و اصول اندازه گیری مختلف، حتی در طرح های آزمون تخصصی مختلف تفسیر کنند.[1] این درک باعث تشکیل پروتکل هماهنگ شده / ISO / IUPAC AOAC شد.[32] بنابراین، در آغاز (یا کمی بعد از آن)، **z-score** و **q-score** (یا شاخص های معادل با آنها) وجود داشت که توسط همه قابل درک بود.

افزایش گسترده آزمون های مهارت در پی خیزش اعتبارسنجی، باعث بروز نیاز به ایجاد تغییرات کوچکی در زمینه نحوه امتیازدهی به کمک z-score برای مواجهه با کاربردهای مختلف شد. بنابراین، در حال حاضر **z-score**، **$\acute{z}$-score**، **zeta(ζ) scores**، **z_L-score** ، **D-score** و **E_n-score** وجود دارد. نویسندگان و اسناد پی در پی، از نام های قدیمی برای معانی جدید و نام های جدید و نمادها برای معانی قدیمی استفاده کرده اند که این موضوع گیج کننده است. بنابراین، در این مقاله می خواهیم یک نگاه سریع به وضعیت کنونی z-score در استاندارد ISO 13528 (2015) بیندازیم[4].

amc technical brief

z-score

z-score، شاید پراستفاده ترین شاخص عملکرد باشد که به شرکت کننده ها اختصاص داشته و با فرمول $z = (x - x_{pt}) / s_{pt}$ محاسبه می شود. در این فرمول، x نتیجه شرکت کننده، x_{pt} شاخص عملکرد و s_{pt} ، "انحراف معیار برای آزمون تخصصی (SDPT) " است. مقادیر عددی برای x_{pt} و s_{pt} توسط ارائه دهنده طرح آزمون مهارت تعیین می شود. زمانی که z-score خارج از بازه ۲ باشد نتیجه مورد سؤال قرار می گیرد و اگر خارج از بازه ۳ باشد نیاز به اقدام اصلاحی هست (توجیه این تفسیر در بخش "z-score به ما چه می گوید؟" شرح داده شده است.)

تفسیر بر اساس Z-Score مبتنی بر این فرض نیست که نتایج شرکت‌کنندگان

در یک دوره آزمون مهارت به صورت نرمال توزیع شده اند.

شاخص عملکرد D و D% و مقدار δ_E

این شاخص عملکرد یعنی $D = ((x - x_{pt}) / x_{pt})$ و امتیاز مشتق شده از آن یعنی $(\%D = 100D)$ ، که اخیراً q-score نامیده می شد، نشان دهنده تفاوت نسبی نتیجه با x_{pt} است. %D در واقع اختلاف نسبی نتیجه شرکت کننده با مقدار تخصیص یافته را به صورت درصدی نشان می دهد. این مفهوم برای شیمیدانان تجزیه آشناست، اما بدون تفسیر "مناسب بودن برای هدف" قابلیت فهم برای خواننده را ندارد، مگر اینکه خواننده در مورد عملکرد معمول در قالب انحراف معیار نسبی مطلوب آگاه باشد. این امتیاز به سادگی قابل ارائه است، اما دو نقص دارد: (الف) امتیازهای مربوط به محدوده غلظت های مختلف برای یک اندازه گیری ممکن است قابل مقایسه نباشند، و (ب) در رابطه ی این شاخص عملکرد هیچ اشاره ای به "تناسب برای هدف" نشده است. نقص دوم با معرفی یک عبارت اضافی $\boldsymbol{\delta_E}$ قابل برطرف شدن است، اگرچه اعمال کردن عبارت $\boldsymbol{\delta_E}$ به رابطه در واقع تنها باعث تبدیل شدن D-score به z-score می شود.

amc technical brief

z´-score

انجام یک اصلاحیه برای z-score منجر به ارائه یک شاخص عملکرد به نام $\mathbf{z' = (x - xpt) / \sqrt{\sigma_{pt}^2 + u^2(x_{pt})}}$ شده است که در صورتی که عدم قطعیت $u(x_{pt})$ مربوط به یک مقدار تخصیص یافته مستقل به قدر کافی بزرگ باشد تا به طور قابل توجهی بر روی z-score تأثیر بگذارد، مورد استفاده قرار می گیرد. این اتفاق معمولاً زمانی اتفاق می افتد که $u(x_{pt}) > 0.3\sigma_{pt}$ باشد و در این صورت، z-score نسبتاً بیش از ۴٪ نسبت به حالت نرمال کاهش می یابد. z-score به درستی برای استاندارد سازی انحراف از مقدار تخصیص یافته عمل می کند، اما از نشان دادن تفاوت بین یک نتیجه نامطلوب و مقدار تخصیص یافته نامطلوب ناتوان است.

zeta-score and E_n

$\zeta - score$، $\zeta = (x - x_{pt}) / \sqrt{u^2(x) + u^2(x_{pt})}$ در شرایطی مورد استفاده قرار می گیرد که شرکت کننده نتیجه ای را با تخمین عدم قطعیت $u(x)$ ارائه می دهد و مقدار تخصیص یافته (x_{pt}) یک مقدار مرجع تأیید شده با عدم قطعیت $u(x_{pt})$ است. مانند z، امتیازهای زتا بدست آمده در خارج از دامنه ±۲ به عنوان نتیجه مشکوک و مقادیر خارج از دامنه ±۳ دلیلی برای آغاز اقدام اصلاحی یا حداقل بروز نگرانی هستند.

$\zeta - score$ با افزایش انحراف از مقدار تعیین شده یا با کاهش عدم قطعیت گزارش شده افزایش می یابد. بنابراین، یک $\zeta - score$ بزرگ می تواند خطای بزرگ، عدم قطعیت تخمین زده شده کمتر از حد، یا هر دو را نشان دهد. این ابهام باعث می شود تا $\zeta - score$ در صورتیکه شرکت کنندگان برای کاهش امتیاز خود، عدم قطعیت $u(x)$ را بیش از حد تخمین زده شده بیان کنند، معیار نادرستی برای ارزیابی انحراف باشد.

شاخص عملکرد E_n در واقع شبیه به $\zeta - score$ است، با این تفاوت که در رابطه E_n ، عدم قطعیت استاندارد با عدم قطعیت گسترده جایگزین شده است. در نتیجه، محدوده نتایج $E_n -$ $scores$، نصف محدوده نتایج $\zeta - scores$ هستند، بنابراین مقدار E_n خارج از ۱± معمولاً به عنوان نتیجه مشکوک در نظر گرفته می شود. این شاخص عملکرد در آزمایشگاه های کالیبراسیون بیشتر از آزمایشگاه های تجزیه ای مورد استفاده قرار می گیرد.

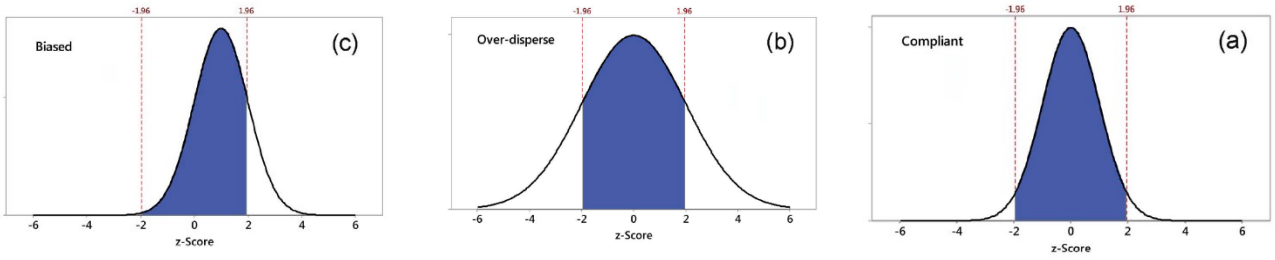
$z_L - scores$

این شاخص عملکرد $z_L = (x - x_{pt}) / u_f$ برای آزمایشگاهی طراحی شده است که با هدف خاصی در آزمون مهارت شرکت کرده و در نتیجه انحراف استاندارد ارائه شده توسط برگزار کننده PT در طرح آزمون مهارت (SDPT)، برای هدف آنها مناسب نیست (جهت اطلاعات بیشتر به AMCTB No. 2 مراجعه کنید). ایده اصلی z_L-score این است که شرکت کننده و مشتری (برای یک هدف ویژه) به صورت مشترک یک SDPT متفاوت (یعنی یک عدم قطعیت استاندارد مناسب برای برنامه کاربردی منظور هدف خودشان) تعیین کرده و از آن به عنوان انحراف استاندارد مناسب برای هدف برای محاسبه امتیاز اصلاح شده استفاده کنند. استفاده از $z_L - scores$ تا زمانی یک شیوه علمی صحیح است که (a) منبع آن معتبر و قابل ارائه به طرف سوم باشد و (b) مقدار تعیین شده در طرح PT به عنوان مقدار تخصیص یافته در محاسبه استفاده شود (توجه: این امتیاز بخشی از استاندارد ISO 13528 نیست. در ابتدا به آن امتیاز $\zeta - score$,zeta گفته می شد، اما این عبارت بعداً برای اهداف دیگری استفاده شد).

z−$score$ به ما چه می‌گوید؟

$z - score$ برای یک آزمایشگاه با این فرضیه تفسیر می شود که نتایج آزمایشگاه به صورت تصادفی از توزیع نرمال با میانگین صفر و انحراف معیار واحد خوردار است. در آزمایشگاهی مطابق با این شرایط، امتیازهای خارج از محدوده ۲± با احتمال ۰.۰۵ رخ می‌دهند که به عنوان یک رخداد محدود و در حد هشدار باید تفسیر شوند. امتیازهای خارج از محدوده ۳± با فراوانی کمتری (با احتمال

کمتر از ۰.۰۰۳) در فرض توزیع نرمال استاندارد رخ می‌دهند و می‌توانند به عنوان اقدامات محدودکننده برای آغاز بررسی علت مشکل در نظر گرفته شوند. آزمایشگاه‌هایی که با انحراف استاندارد σ_{pt} همپوشانی ندارند و یا میانگین نتایج آنها به صورت انحرافی تغییر کرده باشد، ممکن است درصد بیشتری از امتیازهای خارج از این محدوده‌ها داشته باشد. شکل ۱ این توضیحات را بخوبی نشان می‌دهد.



شکل ۱ توزیع امتیازهای زتا که از آزمایشگاه‌های شرکت‌کننده با شرایط زیر انتظار می‌رود را نشان می‌دهد (a) دقیقاً متناسب با مقدار تخصیص یافته و انحراف استاندارد SDPT ، (b) سازگاربا مقدار تخصیص یافته اما با پراکندگی نتایج بیشتر، و (c) متناسب با انحراف استاندارد SDPT اما دارای انحراف نسبت به مقدار تخصیص یافته. مناطقی که با رنگ آبی نشان داده شده اند، میزان همپوشانی با نتایج "مطلوب" یا آزمایش‌های با کیفیت را نشان می‌دهند، در واقع به عبارتی نتایجی که در محدوده $-1.96 < z < 1.96$ قرار دارند نتایج مطلوبی هستند. نسبت‌های مطلوب برای شرایط مذکور به ترتیب عبارتند از : (a) 0.95، (b) 0.67 و (c) 0.83.

ضروری است که تأکید شود که **تفسیرنتایج به کمک z-score با این فرضیه نیست که نتایج شرکت کنندگان در یک دوره به شکل نرمال توزیع شده اند**. این یک مشکل شایع در بین آماردانان و ناظرینی است که با آزمون مهارت آشنایی ندارند. تفسیر نتایج با z-score بر این ایده

است که اگر تمامی آزمایشگاه ها به طور مشابه و دقیقاً منطبق با موارد مورد نیاز برای گزارش یک مقدار تخصیص یافته و SDPT عمل کنند، نتایج آن ها تقریباً به شکل توزیع نرمال با میانگینx_{pt} و انحراف معیار σ_{pt} توزیع خواهد شد. در این صورت، z-score یک توزیع نرمال با میانگین صفر و

انحراف معیار یک واحدی را نشان می‌دهند. لازم به ذکر است که این فرضیه به معنای این نیست که نتایج واقعی شرکت کنندگان به شکل نرمال توزیع شده باشند، بلکه تنها به این معناست که عملکرد ایده‌آل تمامی شرکت کنندگان به شکل توزیع نرمال با میانگین x_{pt} و انحراف معیار σ_{pt} منجر به توزیع نرمال و استاندارد شاخص عملکرد خواهد شد. بنابراین در طول زمان، z-score عملکرد یک شرکت کننده را با معیار هدف در آزمون مهارت به عنوان عملکرد خوب مقایسه می‌کنند.

در طولانی مدت

آزمون مهارت می‌تواند به راحتی عملکرد بد را به نمایش بگذارد، اما توانایی آن در نشان دادن صلاحیت آزمایشگاه با قدرت کمتری همراه است. هر چند z-score خارج از محدوده ±۳ نشان دهنده نیاز به اقدام است، اما شاخص عملکرد داخل محدوده ±۲ به تنهایی به این معنا نیست که همه چیز خوب است. این موضوع به خوبی در شکل ۱ نشان داده شده است که حتی با انحراف استاندارد دو برابر σ_{pt} ، یک شرکت کننده هنوز با احتمالی حدود ۰.۶۷ به یک z-score در محدوده ±۲ می‌رسد. حتی با وجود یک انحراف معیار مناسب اما با وجود انحرافی (bias) برابر با σ_{pt} ، احتمال این که یک امتیاز در محدوده ±۲ باشد، حدود ۰.۸۳ است. به همین دلیل، اشتباه است که بگوییم می‌توانیم صلاحیت یک آزمایشگاه را تنها از طریق یک z-score قابل قبول یا حتی چندین z-score پشت سر هم نشان دهیم. برای نشان دادن صلاحیت، باید به دید بلندمدتی از z-score به همراه سایر عوامل بپردازیم.

این نادرست است که بگوییم ما می توانیم شایستگی خود را از طریق

یک z-score تنهای قابل قبول نشان دهیم.

---

[1] Analytical Methods Committee, Analyst, 1992, 117, 97–104.
[2] M. Thompson and R. Wood, Pure Appl. Chem., 1993, 65, 2123–2144.
[3] M. Thompson, S. L. R. Ellison and R. Wood, Pure Appl. Chem., 2006, 78, 145–196.
[4] ISO 13528, 2015.