فصلنامه اندازه‌گیری تربیتی
دانشگاه علامه طباطبائی
سال نهم، شماره ۳۳، پاییز ۹۷، ص ۵۳ تا ۷۸

Quarterly of Educational Measurement
Allameh Tabataba'i University
Vol. 9, No. 33, Fall 2018

مقایسه خطای استاندارد اندازه‌گیری شرطی نمرات مقیاس تحت توزیع دوجمله‌ای و بتا-دوجمله‌ای

سمیه کاوه[۱]، ابراهیم خدایی[۲]، سید امین موسوی[۳]، علی مقدم‌زاده[۴]، جلیل یونسی[۵]

تاریخ دریافت: ۹۷/۰۳/۰۸
تاریخ پذیرش: ۹۷/۰۸/۲۵

چکیده

به منظور تسهیل تفسیر نمرات خام، معمولا آنها را به نمرات مقیاس تبدیل می‌کنند که در بعضی مواقع این تبدیلات یک سری تبدیلات غیرخطی هستند و می‌توانند بر مقدار خطای استاندارد اندازه‌گیری شرطی در طول مقیاس نمره تأثیرگذار باشند. لذا اهداف این پژوهش، معرفی و مرور روش‌های متداول برای برآورد خطای استاندارد اندازه‌گیری شرطی براساس چارچوب نظریه نمره واقعی و نیز مقایسه خطای استاندارد اندازه‌گیری شرطی تحت تبدیل غیرخطی نرمال شده و همصدک‌سازی بر روی نمرات خام آزمون‌های پیشرفت تحصیلی دانش‌آموختگان رشته ریاضی فیزیک سال ۱۳۹۳ بودند. در این پژوهش، از نمونه تصادفی۳۹۴۳ تایی از دانش‌آموختگان رشته ریاضی فیزیک سال ۱۳۹۳ که در کنکور سراسری سال ۱۳۹۴ شرکت کرده بودند، استفاده نمودیم. خطای استاندارد اندازه‌گیری شرطی این تبدیلات براساس روش دوجمله‌ای «برنان» و «لی» و همچنین روش «چنگ» و بر اساس توزیع بتا-دوجمله‌ای برآورد شد. نتایج این پژوهش نشان از آن دارد که خطای استاندارد اندازه‌گیری شرطی تحت روش چنگ، هموارتر از روش دوجمله‌ای است، اما در هر دو نوع تبدیلات برآورد خطا در نقاط میانی، دارای مقدار بالاتر از نقاط ابتدایی و انتهایی حدود نمرات بود و خطای استاندارد اندازه‌گیری شرطی همصدک‌سازی همواره کمتر از نرمال شده بود؛ بنابراین براساس این معیار، همصدک‌سازی بهتر از نرمال شده است.

واژگان کلیدی: خطای استاندارد اندازه‌گیری شرطی، دوجمله‌ای، بتا-دوجمله‌ای، همصدک‌سازی، نرمال شده.

---

۱. دانشجوی دکتری تخصصی سنجش آموزش، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران، نویسنده مسئول، (kaveh1@ut.ac.ir).

۲. دانشیار گروه روش‌ها و برنامه‌های آموزشی و درسی دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران.

۳. استادیار گروه روان‌شناسی تربیتی و آموزش استثنایی، دانشگاه ساسکاچوان کانادا.

۴. استادیار گروه روش‌ها و برنامه‌های آموزشی و درسی دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران.

۵. دانشیار گروه سنجش و اندازه‌گیری دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبایی.

مقدمه

تجدیدنظر اخیر و انتشار استانداردهای آزمون‌های آموزشی و روان‌شناختی[1] (انجمن تحقیقات آموزشی آمریکا[2]، انجمن روان‌شناسی آمریکا، شورای ملی اندازه‌گیری در آموزش و پرورش[3]) مسئولیت‌های جدیدی را برای ناشر و کاربران آزمون مشخص کرده‌اند. توصیه خاص استانداردهای جدید این است که ناشران آزمون، برآوردهای خطای استاندارد اندازه‌گیری شرطی[4] را نیز گزارش دهند تا کاربران این آزمون‌ها بتوانند این اطلاعات را در تفسیر نمرات اعمال کنند (پرایس و راجو[5]، ۲۰۰۶).

خطای استاندارد اندازه‌گیری[6] عبارت است از: انحراف استاندارد فرضی از اندازه‌گیری‌های مکرر در تعداد دفعات زیاد بر روی آزمودنی (فلت، استفن و گاپتا[7]، ۱۹۸۵، ص. ۳۵۱). برآورد خطای استاندارد اندازه‌گیری برای تمام آزمودنی‌ها به‌طور یکسان به‌کار برده می‌شود که به صورت زیر حاصل می‌شود:

$$S_E = S_x\sqrt{1 - r_{xx'}} \qquad (1)$$

S_E، S_x و $r_{xx'}$ به ترتیب برابر با خطای استاندارد اندازه‌گیری، انحر اف استاندارد نمرات مشاهده‌شده و برآورد ضریب پایایی برای جامعه مورد نظر هستند. برآورد خطای استاندارد اندازه‌گیری، متوسط مقدار دقت اندازه‌گیری را برای کل جامعه و در سطح آزمون نشان می‌دهد؛ از این رو، تغییرپذیری خطاهای اندازه‌گیری برای آزمودنی‌هایی با نمرات کم، میانی و خیلی بالا را منعکس نمی‌کند. به همین منظور، اکثر نسخه‌های جدید آزمون‌های آموزشی و روان‌شناسی از برآورد خطاهای استاندارد اندازه‌گیری شرطی؛ یعنی به دست آوردن خطای اندازه‌گیری برای آزمودنی‌ها در سطوح خاصی از نمره آزمون استفاده می نمایند (فلت، استفن و گاپتا، ۱۹۸۵). در واقع، بر اساس خطای استاندارد اندازه گیری شرطی می توان تغییرات دقت اندازه گیری را در سراسر طیف نمرات بررسی کرد.

تبدیلات غیرخطی بر روی نمرات خام امری متداول در فعالیتهای آزمون سازی است که باعث تسهیل تفسیر نمرات می شود. اما این نوع تبدیلات بر مقدار پایایی و خطای استاندارد اندازه‌گیری شرطی نیز تأثیرگذار است؛ بنابراین برآورد خطاهای استاندارد اندازه‌گیری شرطیِ نمرات مقیاس (تبدیل شده) نیز بسیار مهم و ضروری است (کولن، هنسون و برنان[8]، ۱۹۹۲). بنا بر اهمیت این نوع خطا، بسیاری از

1. Standards for Educational and Psychological Testing
2. American Educational Research Association
3. National Council on Measurement in Education
4. Conditional Standard Error of Measurement (CSEM)
5. Price and Raju
6. Standard Error of Measurement (SEM)
7. Feldt, Steffen ,and Gupta
8. Kolen, Hanson ,and Brennan

پژوهشگران روش‌های گوناگونی را برای برآورد خطای استاندارد اندازه‌گیری شرطی نمرات مقیاس ارائه نمودند که معمولاً به دو دسته کلی تقسیم می شوند:

- برآورد خطای استاندارد اندازه گیری شرطی نمرات مقیاس بر اساس «مدل های نمره ضعیف [1]».

در این روش‌ها محاسبه خطا بدون نیاز به فرضیات نظری و آماری است و معمولاً نیازمند در اختیار داشتن نمره تک تک سوالات آزمون‌ها است؛ مانند تقریبی از «روش دلتا[2]». فلت و کوالز[3] (۱۹۹۸)، X، T، E و S را به ترتیب نمره مشاهده‌شده، واقعی ، خطا و نمره مقیاس (نمره حاصل از تبدیل مورد نظر بر روی نمره خام) در نظرگرفتند و چنین فرض کردند که f یک تابع خطی برای تبدیل نمرات خام به نمرات مقیاس است. آنگاه:

$$\begin{aligned} S = f(X) = B + A(X) = B + A(T_x + E_x) \\ = [B + A(T_x)] + A(E_x) = T_s + E_s \end{aligned} \quad (۲)$$

در معادله (۲)، A و B اعداد ثابت هستند.

در نتیجه از معادله (۲)، معادله (۳) حاصل می‌شود:

$$E_s = A(E_x) \rightarrow \sigma_{E_s} = A\sigma_{E_x} \quad (۳)$$

به عبارت دیگر؛ برای هرمقدار از نمره آزمودنی[4] (x)، خطای استاندارد اندازه‌گیری شرطی نمره مقیاس(S) برابر است با:

$$\sigma_{E(S|x)} = A\sigma_{E(X|x)} \quad (۴)$$

معادلات (۲) و (۳) در صورتی قابل‌قبول هستند که $f(x)$ یک تابع خطی باشد، درصورتی‌که $f(x)$ یک تبدیل غیرخطی باشد، A برابر است با مشتق اول تابع؛ یعنی $f'(x)$ است. آنگاه خطای استاندارد اندازه‌گیری شرطی در هر مقدار X به صورت زیر در نظر گرفته می‌شود:

$$\sigma_{E(S|x)} = f'(x)\, \sigma_{E(X|x)} \quad (۵)$$

1. weak true score models
2. delta method
3. Feldt and Qualls
4. Individual score points

برنان و لی[1] (۱۹۹۹) به نقل از کولن و برنان[2] (۱۹۹۵) چنین بیان کردند که معادله (۵) تقریبی از «روش دلتا» است[3]. برنان و لی (۱۹۹۹)، برآورد $\sigma_{E(S|x)}$ را در معادله (۵) یک فرآیند دو مرحله‌ای در نظر گرفتند:

الف) برآورد $f'(x)$:

برای برآورد $f'(x)$ می توان از روش‌هایی مانند چندجمله‌ای[4]، اسپیلین مکعبی[5] و روش تقریبی فلت و کوالز (۱۹۹۸) استفاده نمود (این روشها درقسمت پیوست به تشریح معرفی شده اند).

ب) برآورد $\sigma_{E(X|x)}$

برای برآورد $\sigma_{E(X|x)}$ می توان از روش‌هایی مانند روش ثرندایک (۱۹۵۱) و روش چندجمله‌ای مولنکف (۱۹۴۹) استفاده نمود (این روشها درقسمت پیوست به تشریح معرفی شده اند).

- برآورد خطای استاندارد اندازه گیری شرطی نمرات مقیاس بر اساس «مدل‌های نمره واقعی[6]».

در این روشها محاسبه برآورد خطا بر اساس فرضیات نظری و آماری قوی است که با مدل بندی کردن توزیع نمرات واقعی و توزیع خطای اندازه‌گیری شرطی مشروط به نمره واقعی می توان مقدار آن را محاسبه نمود؛ در این روش می توان مدل‌های خطای دوجمله ای[7]، دوجمله ای مرکب[8] و مدل‌های نظریه سوال-پاسخ[9] را نام برد.

کولن، هنسون و برنان[10] (۱۹۹۲)، روشی را برای برآورد پایایی و خطاهای استاندارد اندازه‌گیری شرطی نمرات مقیاس براساس مدل‌بندی‌کردنِ توزیع نمرات واقعی و توزیع خطای اندازه‌گیری شرطی مشروط به نمره واقعی مطرح نمودند. این مدل‌سازی بر اساس مدل‌های نمره واقعی یا مدل‌های نظریه سؤال- پاسخ انجام می‌گیرد.

- روش کولن، هنسون و برنان

---

1. Brennan and Lee
2. Kolen and Brennan

۳. از آنجایی که تبدیلات مورد استفاده در این پژوهش بر روی نمرات خام، تبدیلات غیرخطی هستند، پس در این پژوهش بر روی تبدیلات غیر خطی تمرکز نمودیم.

۴. polynomial

۵. cubic Spline

6. strong true score models
7. binomial
8. compound binomial
9. Item respone theory (IRT)
10. Kolen, Hanson ,and Brennan

کولن، هنسون و برنان (۱۹۹۲) معتقدند از طریق مدل‌بندی‌کردنِ توزیع نمرات واقعی و توزیع خطای اندازه‌گیری شرطی مشروط به نمره واقعی می‌توان خطاهای استانداردِ اندازه‌گیری شرطی نمرات خام را بدست آورد؛ لذا این روش، می‌تواند بعد از مشخص‌نمودن تبدیلات نمره خام به نمره مقیاس، برای برآورد خطاهای استاندارد اندازه‌گیری شرطی و پایایی نمرات مقیاس استفاده شود. آنها چنین فرض نمودند که نمرات خام از یک آزمون k سؤالی است که به صورت دووضعیتی نمره‌گذاری شده‌اند، پس متغیر تصادفی نمره خام X نشان‌دهنده تعداد پاسخ‌های صحیح i بر روی آزمون است، که می تواند مقادیر از ۰ تا k را با احتمال زیر اختیار کند:

$$P(X = i) = \int_{0}^{1} P(X = i|\tau)g(\tau)d\tau \qquad (۶)$$

τ برابر با نسبت نمرات صحیح در یک آزمون با تابع چگالی $g(\tau)$ در جامعه آزمودنی‌ها است. $P(X = i|\tau)$ برابر با توزیع شرطی نمرات خام برای آزمودنی‌های تحت τ است. با برآورد نمودن $P(X = i|\tau)$ و $g(\tau)$، احتمال این‌که آزمودنی نمره i در معادله (۶) اختیار کند، مشخص و برآورد می‌شود.

میانگین نمره مقیاس برای آزمودنی‌هایی با نسبت نمره واقعی صحیح τ برابر است با:

$$\xi(\tau) = E(S(X)|\tau) = \sum_{i=0}^{k} S(i)P(X = i|\tau) \qquad (۷)$$

که $\xi(\tau)$ نمره مقیاس واقعی برای چنین آزمودنی‌هایی است. واریانس خطای اندازه‌گیری برای نمره مقیاس و آزمودنی‌هایی با نسبت نمره واقعی τ برابر است با:

$$\sigma^2[S(X)|\tau] = E([S(X) - \xi(\tau)]^2|\tau) = \sum_{i=0}^{k} [S(i) - \xi(\tau)]^2 P(X = i|\tau) \qquad (۸)$$

جذر معادله (۸) برابر با خطای استاندارد اندازه‌گیری شرطی تحت نسبت نمره واقعی τ مورد نظر است.

- روش دوجمله‌ای برنان و لی[1]

یکی از روش‌هایی که برنان و لی (۱۹۹۹) برای محاسبه این معیار ارائه داده‌اند، براساس مفروضات توزیع دوجمله‌ای است. بدین‌گونه که فرض می‌شود آزمون‌ها دارای k سؤال هستند که به صورت دووضعیتی (یعنی ۰ و ۱) نمره‌گذاری شده‌اند. احتمال این‌که یک آزمودنی، به Y سؤال از این آزمون پاسخ صحیح دهد برابر است با:

1. Brennan and Lee

$$P(Y = i|k, \tau) = \binom{k}{i} \tau^{i} (۱ - \tau)^{k-i} \qquad i = ۰,۱, \dots, k \tag{۹}$$

τ به عنوان نسبت واقعی پاسخ‌های درست برای یک آزمودنی لحاظ می‌گردد که برآورد آن برابر است با تعداد پاسخ‌های درست به تعداد کل آن‌ها؛ یعنی $\bar{x} = \frac{x}{k}$. قابل توجه است که x برابر با نمره‌ای است که قصد داریم خطای استاندارد اندازه‌گیری شرطی را برای آن محاسبه کنیم؛ اما Y متغیری با توزیع دوجمله‌ای است که مقادیر ۰ تا k را با احتمال دوجمله‌ای اختیار می‌کند. برنان و لی (۱۹۹۹) برآورد خطای استاندارد اندازه‌گیری شرطی به إزای تبدیل هر نمره y به $f(y)$ را برای هر مقدار نمره خاص x به صورت زیر محاسبه نمودند:

$$\hat{\sigma}_{E(S|x)} = c_k \sqrt{\sum_{y=۰}^{k} [f(y)]^{۲} P(y|\bar{x}, k) - \left[\sum_{y=۰}^{k} [f(y)] P(y|\bar{x}, k)\right]^{۲}} \tag{۱۰}$$

$$c_k = \sqrt{k/(k - ۱)} \tag{۱۱}$$

c_k میزانی است که به إزای آن $\hat{\sigma}_{E(S|x)}$ به یک برآورد نااریب تبدیل می‌شود.

- روش دوجمله‌ای مرکب برنان و لی

برنان و لی (۱۹۹۹)، روشی را برای برآورد خطای استاندارد اندازه‌گیری شرطی نمرات مقیاس بر اساس چارجوب کولن، هنسون و برنان (۱۹۹۲) و مشابه خطای استاندارد اندازه‌گیری فلت (۱۹۸۴) ارائه نمودند. آنها فرض نمودند سؤالات را می‌توان بر حسب محتوایشان به H طبقه دسته‌بندی نمود بگونه‌ای که در هر طبقه، خطا به صورت دوجمله‌ای توزیع شده است و این خطاها با یکدیگر ناهمبسته هستند و همچنین؛

$k = \{k_1, \dots, k_h, \dots, k_H\}$ = تعداد سوالات در طبقات

$x = \{x_1, \dots, x_h, \dots, x_H\}$ = تعداد پاسخ‌های صحیح در طبقات برای هر آزمودنی

$\bar{x} = \{\bar{x}_1, \dots, \bar{x}_h, \dots, \bar{x}_H\}$ = نسبت پاسخ‌های صحیح مشاهده شده در طبقات برای هر آزمودنی

$\pi = \{\pi_1, \dots, \pi_h, \dots, \pi_H\}$= نسبت واقعی پاسخ‌های صحیح در طبقات برای هر آزمودنی

$w = \{w_1, \dots, w_h, \dots, w_H\}$ = وزن نسبت داده شده به طبقات

$x_t = \sum_{h=1}^{H} w_h x_h$= نمره موزون کل[1] هر آزمودنی بر روی تمام طبقات

y_h = تعداد پاسخ‌های صحیح در طبقه ام h

y_t= نمره موزون کل بر روی تمام طبقات

1. Weighted total score

آنگاه خطای استاندارد اندازه‌گیری شرطی نمره موزون کل خام x_t به نمره مقیاس $f(x_t)$ به صورت زیر محاسبه می‌گردد:

$$\hat{\sigma}_{E(S|x)} = c_k \sqrt{\sum_{y_t=0}^{\max(y_t)} [f(y_t)]^2 P(y_t|\bar{x}, k) - \left[\sum_{y_t=0}^{\max(y_t)} [f(y_t)] P(y_t|\bar{x}, k)\right]^2} \qquad (۱۲)$$

c_k میزانی است که به إزای آن $\hat{\sigma}_{E(S|x)}$ به یک برآورد نااریب تبدیل می‌شود که در معادله (۱۱) مشخص شده است.

- روش بتا-دوجمله‌ای[1] چنگ[2]

چنگ (۲۰۰۶) برای ارزیابی اثرات نمرات مقیاس حاصل از تبدیلات گوناگون بر روی نمره خام از خطای استاندارد اندازه‌گیری شرطی ارائه شده توسط کولن، هنسون و برنان (۱۹۹۲) استفاده نمود؛ یعنی توزیع نمرات واقعی و خطای اندازه‌گیری را تحت نمره واقعی مدل‌بندی نمود. در این حالت، پارامترهای مدل نمره واقعی براساس توزیع بتا-دوجمله‌ای ارائه شده کارلین و روبین(۱۹۹۱) برآورد شدند. کارلین و روبین(۱۹۹۱) معتقدند؛ آزمودنی‌هایی که آزمون بر آنها اجرا می‌شود، دارای توانایی متفاوت هستند؛ به عبارت دیگر، احتمال پاسخ صحیح در بین آنها یکسان نخواهد بود؛ بنابراین آنها از توزیع بتا-دوجمله‌ای -که بسط توزیع دوجمله‌ای می باشد- برای منعکس‌نمودن تغییرپذیری توانایی آزمودنی‌ها استفاده نمودند.

کارلین و روبین (۱۹۹۱) چنین فرض نمودند که تعداد پاسخ‌های آزمودنی بر روی آزمونی با k سؤال دارای توزیع دوجمله‌ای با نسبت پاسخ درست به هر سؤال مانند τیی است که دارای توزیع پیشین[3] بتا با پارامترهای$\phi = (\alpha, \beta)$ با میانگین $\mu = \mu(\phi)$ و واریانس $\sigma^2 = \sigma^2(\phi)$ و χ ضریب همبستگی بین پارامترهای α, β به صورت زیر است:

$$f(\tau|\phi) = \left(B(\alpha, \beta)\right)^{-1} \tau^{\alpha-1} \left(1 - \tau\right)^{\beta-1} \quad 0 \le \tau \le 1,\ \alpha > 0,\ \beta > 0,\ \phi = (\alpha, \beta) \qquad (۱۳)$$

$$\mu = \frac{\alpha}{\alpha+\beta} \text{ و } \sigma^2 = \chi\mu(1-\mu) \text{ و } \chi = (\alpha + \beta + 1)^{-1} \qquad (۱۴)$$

که در معادله (۱۳)، $B(\alpha, \beta) = \int_0^1 \tau^{\alpha-1} (1-\tau)^{\beta-1} d\tau$ تعریف می‌شود.

در نتیجه، توزیع Y (تعداد پاسخ‌های درست) دارای توزیع بتا- دوجمله‌ای به صورت زیر است:

1. Beta-binomial
2. Chang
3. Prior distribution

$$P(Y|\phi) = \frac{1}{k+1}\frac{B(\alpha+y,\beta+k-y)}{B\left(y+1,k-y+1\right)B(\alpha,\beta)} \quad (۱۵)$$

سپس، خطای استاندارد اندازه گیری شرطی به صورت زیر برآورد می شود:

$$\widehat{\sigma}_{E(S|x)} = \sqrt{\sum_{y=0}^{k}[f(y)]^2 P(y|\widehat{\phi}) - \left[\sum_{y=0}^{k}[f(y)]P(y|\widehat{\phi})\right]^2} \quad (۱۶)$$

- روش نظریه سؤال- پاسخ

کولن، هنسون و برنان (۱۹۹۲) برای بدست آوردن خطای استاندارد اندازه‌گیری شرطی نمره مقیاس تحت نظریه سوال-پاسخ بدین صورت بیان نمودند که نخست، توزیع شرطی تعداد پاسخ‌های صحیح تحت سطح توانایی مشخص θ مشخص می‌شود. لذا فرض می‌شود، احتمال اینکه متغیرتصادفی X برابر با نمره خام کسب شده i $(i = 0,1,\dots,k)$ در یک آزمونی با k سوال تحت توانایی θ باشد برابر با $P(X=i|\theta)$ است. سپس با استفاده از یک فرمول بازگشتی، احتمال اینکه X_r برابر با نمره i شود $(P(X_r=i|\theta))$، محاسبه می گردد که X_r متغیرتصادفی بر روی r سوال نخست آزمون است. همچنین فرض می‌شود، احتمال پاسخ درست به سوال j ام در سطح توانایی θ برابر با $p_j(\theta)$ است، پس برای $r=1$(یعنی سوال اول) می‌توان نتیجه گرفت که؛ $P(X_1=1|\theta)=p_1(\theta)$ و $P(X_1=0|\theta)=1-p_1(\theta)$ و برای $r>1$ فرمول بازگشتی به صورت زیر حاصل می شود:

$$\begin{aligned} P(X_r=i|\theta) &= P(X_{r-1}=i|\theta)(1-p_r(\theta)) && i=0 \\ &= P(X_{r-1}=i|\theta)(1-p_r(\theta)) + P(X_{r-1}=i-1|\theta)p_r(\theta) && 0<i<r \\ &= P(X_{r-1}=i-1|\theta)p_r(\theta) && i=r \end{aligned} \quad (۱۷)$$

فرمول بازگشتی (۱۶) از $r=1$ شروع و در $r=k$ که $P(X=i|\theta)=P(X_k=i|\theta)$ خاتمه می‌یابد.

سپس، با فرض اینکه S تبدیل نمره خام به نمره مقیاس باشد آنگاه توسط معادلات (۷) و (۸) خطای استاندارد اندازه‌گیری شرطی نمرات مقیاس برآورد می‌گردد.

از میان روشهای ارائه شده برای برآورد خطای استاندارد اندازه گیری شرطی نمرات مقیاس بر اساس مدل‌های نمره واقعی تنها روش دوجمله‌ای برنان ولی و بتا-دوجمله‌ای چنگ نیاز به نمره تک تک سوالات ندارد و در سایر روشها باید نمره تک تک سوالات را در اختیار داشته باشیم.

مولنکف[۱] (۱۹۴۹) در میان اولین کسانی است که خطای استاندارد اندازه‌گیری شرطی را پیشنهاد داد (برنان، ۱۹۹۸). او بررسی کرد که آیا خطای استاندارد اندازه‌گیری در بین نمرات آزمون ثابت است یا

۱. Mollenkopf

خیر. به همین منظور، چنین فرض کرد که خطای استاندارد اندازه‌گیری به عنوان تابعی، حداکثر درجه دوم، از نمرات آزمون است و این‌که نمره آزمون از نمرات دو آزمون موازی تشکیل شده است؛ یک معادله برای پیش‌بینی خطای استاندارد اندازه‌گیری از نمره آزمون به دست آورد. نتایج حاصل از تحلیل آزمون‌ها در ۱۰۰۰ نمونه، نشان داد که خطای استاندارد اندازه‌گیری تنها برای توزیع نمرات متقارن ثابت است.

بعد از زمان کوتاهی، لرد[1] (۱۹۵۵) مدل نظری خطای دوجمله‌ای را ارائه نمود که یکی از معروف‌ترین فرمول‌های خطای اندازه‌گیری شرطی است. لرد (۱۹۵۵) چنین فرض نمود: اگر سؤالات دو یا چند آزمون از یک مجموعه بزرگ سوالات به‌طور تصادفی انتخاب شده باشند، این آزمون‌ها را می توان موازی نامیده می‌شوند. نمره حاصل برای هر آزمودنی بر روی هر یک از چنین آزمون‌هایی دارای توزیع دوجمله‌ای است و جذر واریانس توزیع نمرات برابر با خطای استاندارد اندازه‌گیری برای آزمودنی مورد نظر می‌باشد.

فلت (۱۹۸۴) تعمیمی از خطای استاندارد اندازه‌گیری شرطی لرد بر اساس مدل دوجمله‌ای مرکب ارائه نمود و چنین اظهار کرد که سؤالات را می‌توان بر حسب محتوایشان به طبقاتی دسته‌بندی نمود؛ در هر طبقه، خطا به صورت دوجمله‌ای توزیع شده است و این خطاها با یکدیگر ناهمبسته هستند. همچنین بیان نمود که داده‌های تجربی نشان می‌دهد که روند خطای استاندارد اندازه گیری در طول پیوستار نمره مشاهده شده از مدل خطای دوجمله‌ای مرکب تبعیت می‌کند.

فلت، استفن و گاپتا (۱۹۸۵) بر اساس پنج روش، خطاهای استاندارد اندازه‌گیری شرطی را برآورد و با یکدیگر مقایسه نمودند. قابل توجه است که تمامی این پنج روش، برای محاسبه خطای استاندارد اندازه گیری شرطی نیازمند نمره تک تک سوالات آزمون می‌باشند، که آنها عبارتندا ز:

- روش ثرندایک: در این روش، انحراف استاندارد خطای اندازه‌گیری کل نمره آزمون برابر است با تفاضل انحراف استانداردهای خطای اندازه‌گیری نیمه‌های آزمون.
- روش چند جمله‌ای: این روش مبتنی بر مدل بندی مربع تفاضل نمرات نیمه‌های آزمون $(X_1 - X_2)^2$ بر روی نمره کل X برای هر آزمودنی است که می‌تواند با استفاده از مدل‌های چندجمله‌ای درجه دو، مکعبی و یا درجه چهار پیش‌بینی شود. در واقع، این مدل با استفاده از روش‌های رگرسیون خطی حداقل مربعات پیش‌بینی می‌شود. در نهایت، متوسط مربع تفاضل نمرات نیمه‌های آزمون برابر با واریانس تفاضل دو نیمه آزمون است که همان‌گونه که

1. Lord

ثرندایک (۱۹۵۱) نشان داد؛ این کمیت نشان‌دهنده واریانس خطای اندازه‌گیری برای کل آزمون است.

- روش دوجمله‌ای: بر اساس روش دوجمله‌ای لرد (۱۹۵۵)، خطای استاندارد اندازه‌گیری آزمودنی با نمره کل X در یک آزمون k سؤالی برابر است با:

$$S_E = \sqrt{\frac{X(k-X)}{k-1}} \qquad (۱۸)$$

- روش دوجمله‌ای مرکب: سؤالات بر حسب محتوایشان به طبقاتی دسته‌بندی می‌شوند که در هر طبقه خطا به صورت دوجمله‌ای توزیع شده است و $S_{E(i)}$ خطای استاندارد آزمودنی i برابراست با:

$$S_{E(i)} = \sqrt{\sum_{h=1}^{c} \frac{X_{ih}(k_h - X_{ih})}{k_h - 1}} \qquad (۱۹)$$

که X_{ih} و k_h به ترتیب برابرند با نمره آزمودنی i و تعداد سؤالات در طبقه h می‌باشند.

- برآورد مؤلفه‌های واریانس[1]: بر اساس روش آنالیز واریانس می‌باشد که توسط ماتریسی از نمرات سوالات آزمودنیها در یک آزمون، میانگین مربعات آزمودنی‌ها (MS_s)، سؤالات (MS_t) و تعامل بین سؤال و آزمودنی ($MS_{s\times t}$) محاسبه می گردد. سپس، انحراف استاندارد خطای اندازه‌گیری برای یک آزمون k سوالی برابر با $\sqrt{k(MS_{s\times t})}$ است.

- روش نظریه منحنی سؤال- پاسخ[2]: احتمال اینکه آزمودنی i ام در سطح توانایی θ_i به سوال j ام پاسخ درست دهد برابر با $P_j(\theta)$ است که معمولا به شکل اجایو نرمال[3] به صورت زیر در نظر گرفته می شود:

$$P_j(\theta_i) = c_j + (1 - c_j) \int_{-\infty}^{a_j(\theta_i - b_j)} \frac{1}{\sqrt{2\pi}} e^{-z^2} dz \qquad (۲۰)$$

a_j ، b_j و c_j به ترتیب پارامترهای دشواری، تمیز و حدس سوال هستند. در صورتی که این پارمترها نسبتاً دقیق برآورد شوند، آنگاه واریانس خطای استاندارد اندازه گیری شرطی برای آزمودنی i ام توسط دو مرحله برآورد می شود:

۱. نخست، پارامتر توانایی آزمودنی برآورد می شود.

1. Variance component estimated
2. Item response curve theory
3. Normal ogive

۲. سپس، برآورد خطای استاندارد اندازه گیری در سطح توانایی θ_i با استفاده از فرمول زیر برآورد می‌شود:

$$S_{E(i)} = \sqrt{\sum_{j=1}^{k} P_j(\theta_i)(1 - P_j(\theta_i))} \qquad (۲۱)$$

هر پنج روش، دارای حداکثر مقدار در نقاط میانی مقیاس نمره بودند، سپس یک کاهش خیلی زیاد در نقاط انتهایی مقیاس وجود داشت. آنها معتقدند که این نوع از روندهای خطای استاندارد اندازه گیری شرطی، احتمالا در میان اکثر آزمون‌های استانداردشده پیشرفت تحصیلی و توانایی رخ می‌دهد و برای انواع دیگر آزمون‌ها، روند خطای استاندارد اندازه‌گیری شرطی به صورت دیگری است که معمولا برای برآورد خطا، دو روش چند جمله‌ای و مدل خطای دو جمله‌ای پیشنهاد داده می‌شوند.

فلت و کوالز (۱۹۹۸)، برای تعیین دقت در یک سطح نمره معین، برآورد خطاهای استاندارد اندازه‌گیری شرطی را الزامی دانستند. آنها دو روش نسبتا ساده را برای برآورد خطاهای استاندارد اندازه‌گیری شرطی برای نمره مقیاس تحت تبدیلات غیرخطی پیشنهاد دادند که عبارتند از روش چندجمله ای و تقریبی[1]. در روش اول، از طریق تابع چند جمله‌ای، رابطه بین نمرات خام و نمرات حاصل تحت تبدیلات غیرخطی برازش داده می‌شوند، سپس از حاصل‌ضرب شیب تابع به دست‌آمده در نمره خام مورد نظر در مقدار خطای اندازه‌گیری در همان نمره، خطای استاندارد اندازه‌گیری شرطی نمره مقیاس به إزای نمره خام مورد نظر محاسبه می‌شود. روش دوم که یک روش تقریبی است، به این صورت تعیین می‌شود: به إزای هر نمره خام $X_.$، کاربر می‌تواند با استفاده از جداول در دسترس -که شامل جفت‌های بین نمرات خام و نمرات تحت تبدیل غیرخطی است- نمرات متناظر با $X_. + C$ و $X_. - C$ ؛ یعنی به ترتیب S_U و S_L را به دست آورد که S_U و S_L به ترتیب نشان دهنده نمره تحت تبدیل غیرخطی متناظر با حد بالای نمره خام $X_.$ (یعنی $X_. + C$) و حد پایین نمره خام $X_.$ (یعنی $(X_. - C)$ هستند. آنها بیان نمودند C یک عدد ثابت است که برای آزمون‌های پیشرفت تحصیلی ۳ یا ۴، یک انتخاب راحت و منطقی است. تقسیم خطای استاندارد اندازه‌گیری نمره خام مورد نظر بر تفاوت بین این نمرات حاصل از تبدیل‌های غیرخطی نشان‌دهنده این است که خطای استاندارد اندازه‌گیری شرطی در حدود 2C در مجاورت و نزدیکی $X_.$ چقدر می‌باشد. همچنین آن‌ها معتقدند، کاربرد تجربی این روش‌ها نشان می‌دهد که خطاهای استاندارد اندازه‌گیری نمرات مقیاس (یا به عبارت دیگر، نمرات حاصل از تبدیلات غیرخطی) همانند نمره خام در کل حدود نمره مقیاس تغییر می‌کند. با این حال، برخلاف

1. approximation method

خطاهای استاندارد اندازه‌گیری نمرات خام، خطاهای استاندارد اندازه‌گیری نمرات مقیاس تغییرات نامنظم‌تری را نمایش می‌دهد که می‌تواند در نتیجه گردکردن و یا تبدیل نمرات باشد. با استفاده از این دو روشِ پیشنهادی فلت و کوالز (۱۹۹۸)، برآوردهای نسبتا یکسانی ایجاد می‌شود که در انتهای مقیاس نمرات دارای بیشترین مقدار و در سطح میانی مقیاس، دارای حداقل مقدار هستند. برنان و لی (۱۹۹۹)، دو روش دوجمله‌ای و دوجمله‌ای مرکب را برای برآورد خطای استاندارد اندازه‌گیری شرطی نمرات مقیاس ایجاد نمودند، البته تحت این فرض که سوالات آزمونها به صورت دووضعیتی (یعنی ۰ و ۱) نمره‌گذاری شده‌اند. ملاحظات نظری و نتایج تحلیل‌ها نشان داد که روش‌های دوجمله‌ای و دوجمله‌ای مرکب دارای ویژگی‌هایی هستند که می‌تواند آنها را به روش‌های چندجمله ای، و فلت و کوالز ترجیح داد. تجزیه و تحلیل حاصل از روش‌های دوجمله ای و دوجمله‌ای مرکب به‌طور کلی نتیجه هموارتری را نسبت به چند جمله‌ای، و فلت و کوالز ارائه می‌دهد. علاوه بر آن، این روش‌ها همانند روش چندجمله‌ای نیاز به قضاوت‌های شخصی برای برازش منحنی ندارند. اگرچه روش‌های دوجمله‌ای و دوجمله‌ای مرکب نسبت به دو روش دیگر دارای مزایایی می‌باشند؛ اما روش فلت و کوالز از لحاظ محاسباتی، بسیار ساده تر است و حتی در بسیاری از حالت‌ها دارای نتایج خیلی متفاوتی نسبت به روش‌های دوجمله ای و دوجمله ای مرکب نمی‌باشند.

لی، برنان و کولن[1] (۲۰۰۰)، چهار روشی را که قبلا برای برآورد خطاهای استاندارد اندازه‌گیری شرطی و نمرات مقیاس ارائه شده بودند، شرح دادند. این روش‌ها عبارتند از: روش نظریه سؤال- پاسخ، روش دوجمله‌ای (برنان و لی ۱۹۹۹)، روش دوجمله‌ای مرکب (برنان و لی ۱۹۹۹) و روش فلت و کوالز (۱۹۹۸) که هر کدام از این روش‌ها دارای فرضیات مختلفی می‌باشند. با استفاده از مطالعه شبیه‌سازی، آنها برای مقایسه دقت برآورد خطاهای استاندارد اندازه‌گیری از اریبی متوسط [2] و متوسط واریانس خطا[3] استفاده نمودند. نتایج نشان داد که روش نظریه سؤال- پاسخ نسبت به سایر روش‌ها، دارای اریبی و واریانس خطای کمتراست. در بین روش‌های غیر نظریه سؤال- پاسخ، روش دوجمله‌ای مرکب نسبت به روشهای دوجمله‌ای، و فلت و کوالز برای همه انواع مقیاس نمرات بهتر است.

شایان ذکر است که در روشهای نظریه سؤال- پاسخ و فلت و کوالز برای محاسبه خطای استاندارد اندازه‌گیری شرطی نیازمند در اختیار داشتن نمره تک تک سوالات آزمون هستیم اما برخلاف آنها، در روشهای دوجمله‌ای با در اختیار داشتن نمره کل آزمون می توان این معیار را محاسبه نمود. روش جمله ی مرکب نیازمند نمره تک تک سوالات برای محاسبه خطای استاندارد اندازه گیری شرطی

1. Lee ,Brennan ,and Kolen
2. average biased
3. average error variance

نیست اما از آنجایی که سؤالات بر حسب محتوایشان به طبقاتی دسته‌بندی می‌شوند باید میزان نمره کل هر طبقه مشخص باشد.

چنگ (۲۰۰۶)، اثرات استفاده از روش‌های تبدیل خطی، نرمال شده[1] و آرک سینوس را برای ساخت نمرات مقیاس بررسی کرد. برای ارزیابی اثرات نمرات مقیاس حاصل از تبدیلات گوناگون بر روی نمره خام از خطای استاندارد اندازه گیری شرطی ارائه شده توسط کولن، هنسون و برنان (۱۹۹۲) و برای برآورد پارامترهای مدل نمره واقعی براساس توزیع بتا-دوجمله‌ای ارائه شده کارلین و روبین(۱۹۹۱) استفاده نمود. یافته‌ها نشان دادند که با استفاده از تبدیلات آرک سینوس تغییرات خطا در کل مقیاس تقریبا ثابت است.

در چند سال اخیر، برای ورود به دانشگاه‌های دولتی و نیمه دولتی، برای محاسبه نمره کل شرکت‌کنندگان در کنکور سراسری ایران، فقط به نمره کل آزمون ورودی کنکور، توجه نشده است، بلکه نمره کل سوابق تحصیلی[2] دانش‌آموزانی که دیپلم‌های خود را در سال ۱۳۸۴ تا ۱۳۹۳ أخذ کرده‌اند نیز تأثیرگذار بوده است. در کنکور سال ۱۳۹۴ تأثیر آن به صورت قطعی و به میزان ۲۵ درصد بوده است. این نمرات سوابق تحصیلی، نمرات حاصل از آزمون‌های پیشرفت تحصیلی است که به صورت تشریحی و با دامنه تغییرات بین ۰ تا ۲۰ بر روی دانش‌آموزان سال سوم دبیرستان در خرداد ماه به صورت نهایی و در سرتاسر کشور برگزار می‌گردد. با توجه به این‌که دامنه و پراکندگی نمرات حاصل از آزمون‌های پیشرفت تحصیلی و نمره‌های خام آزمون سراسری با یکدیگر متفاوت هستند، این نمرات خام با استفاده از تبدیلات نرمال شده به نمرات تراز تبدیل می‌شوند. این امر با استفاده از فرمول‌های آماری انجام می‌گیرد؛ بدین صورت که برای تعیین نمره تراز، باید تمامی نمرات خام داوطلبان در هر گروه آزمایشی و هر درس، در دسترس باشد. از روی نمرات هر درس به ترتیب نزولی فراوانی نمرات، فراوانی تجمعی و مقادیر فراوانی تجمعی نسبی زیر سطح منحنی نمره تراز محاسبه گردد. سپس نمره کل سوابق تحصیلی در هر زیر گروه، با استفاده از میانگین وزنی نمره‌های تراز هر یک از درس‌های مؤثر در سوابق تحصیلی محاسبه می‌شود (پیک سنجش ۸۸۲، ۱۳۹۳). در سازمان سنجش آموزش کشور، در ترکیب و همسان‌سازی نمره‌های سوابق تحصیلی و کنکور، ویژگی‌های روان‌سنجی آزمون‌ها؛ از قبیل دشواری و محتوا آزمون‌ها در نظر گرفته نمی‌شود و همچنین تطابقی بین این نمره‌ها وجود ندارد، به عبارت دیگر می‌توان بیان نمود؛ مقیاس‌سازی[3] نمرات با استفاده از روشهای نرمال شده تنها باعث قرار دادن نمرات

1. normalized

۲. طبق تعریف سازمان سنجش، سوابق تحصیلی عبارت است از نمرات دروس سال سوم متوسطه (دیپلم) که زمان امتحان آن سال ۱۳۸۴ و بعد از آن بوده و به صورت نهایی و کشوری برگزار شده است.

3. scaling

روی یک مقیاس مشترک می شود که این امر تنها باعث تسهیل تفسیرپذیری نمرات می‌گردد (کولن و برنان ۲۰۰۴). اما در این روش مقیاس‌سازی دشواری آزمون‌های پیشرفت تحصیلی و کنکور لحاظ نمی-گردد و این نادیده گرفتن دشواری آزمون‌ها می‌تواند در نمره کل آزمون کنکور- که ترکیبی از نمرات آزمون‌های پیشرفت تحصیلی با خرده آزمون‌های کنکور است- تأثیرگذار باشد. به طور کلی نمی‌توان تصمیم عادلانه‌ای را برای شرکت‌کنندگان در کنکور سراسری أخذ نمود. به ویژه این‌که در کنکور سراسری، دانش‌آموختگان سال‌های مختلف شرکت می‌نمایند، حال آن‌که دشواری آزمون‌های پیشرفت تحصیلی آن‌ها با یکدیگر برابر نیست.

لذا اهدافی که در این پژوهش مد نظر است، عبارتند از:

- معرفی روش‌هایی برای برآورد خطای استاندارد اندازه‌گیری شرطی براساس چارچوب نظریه نمره واقعی
- مقایسه تاثیر تبدیلات غیرخطی بر روی نمره خام آزمون‌های پیشرفت تحصیلی دانش‌آموختگان رشته ریاضی فیزیک سال ۱۳۹۳.

در این پژوهش، به دلیل اینکه تنها نمره کل آزمون‌های پیشرفت تحصیلی دانش‌آموختگان را در اختیار داشتیم نه نمره تک تک سوالات، لذا تنها از روش‌هایی برای برآورد خطای استاندارد اندازه‌گیری شرطی استفاده شد که براساس نظریه مدل نمره واقعی بودند.

روش

از آنجایی که هدف این پژوهش، ارائه روش‌هایی برای محاسبه خطای استاندارد اندازه‌گیری شرطی بر اساس چارچوب نظریه نمره واقعی و نیز مقایسه تاثیر روش‌های تبدیلات غیرخطی بر روی نمره خام آزمون‌های پیشرفت تحصیلی دانش‌آموختگان رشته ریاضی فیزیک سال ۱۳۹۳ است؛ لذا این پژوهش کمی و از نوع توصیفی - غیرآزمایشی محسوب می‌گردد. پژوهش‌های توصیفی شامل مجموعه روش‌هایی است که هدف آن‌ها توصیف شرایط یا پدیده‌های مورد بررسی است. اجرای پژوهش‌های توصیفی می‌تواند صرفاً برای شناخت بیشتر شرایط موجود یا کمک به فرایند تصمیم‌گیری باشد (سرمد، بازرگان و حجازی، ۱۳۸۴).

جامعه آماری و گروه نمونه: به منظور دستیابی به اهداف این پژوهش، از نمونه ۵۰۰۰ تایی از شرکت‌کنندگان گروه ریاضی و علوم فنی کنکور سال ۱۳۹۴ که توسط سازمان سنجش به صورت تصادفی ساده انتخاب شدند، استفاده نمودیم[1]. از این تعداد، ۳۹۴۳ نفر از آن‌ها دانش‌آموزانی بودند که

۱. به علت مسائل امنیتی سازمان سنجش، امکان دسترسی به نمونه‌هایی با حجم بیشتر نبود.

آزمون‌های پیشرفت تحصیلی‌شان در خرداد ماه سال ۱۳۹۳ به صورت جداگانه و در سرتاسر کشور برگزار شدند.

داده‌های پژوهش: داده‌های این پژوهش از بانک اطلاعات سازمان سنجش آموزش کشور دریافت شد که شامل؛ نمرات خام و نرمال شده (ترازشده) سوابق تحصیلی دانش‌آموزان رشته ریاضی فیزیک؛ یعنی درس‌های «فیزیک»، «شیمی»، «جبر و احتمال»، «هندسه ۲»، «حسابان»، «فرهنگ و معارف اسلامی»، «زبان فارسی»، «ادبیات فارسی»، «عربی» و «زبان خارجی» سال سوم دبیرستان بودند که در سال ۱۳۹۳ دانش‌آموخته شدند. هر کدام از این آزمون‌ها، در خرداد ماه سال ۱۳۹۳ به صورت جداگانه و به شکل سراسری در کل کشور برگزار شدند. با توجه به ضعف شیوه کنونی تبدیل نمرات خام به تراز نسبت به عدم درنظر گرفتن ویژگیهای روانسنجی نمرات سوابق تحصیلی و کنکور، انواع روشهای مرتبط‌سازی را بر اساس دوران (۲۰۰۴)؛ یعنی همترازسازی، تطابق و پیش‌بینی در نظر گرفتیم و مرتبط‌سازی مناسب را بر اساس سه معیار معرفی شده دوران (۲۰۰۴)؛ یعنی شباهت محتوای آزمون‌ها، قدرت رابطه بین نمرات آزمون‌ها و نامتغیر بودن انتخاب کردیم. از بین سه روش تطابق (میانگین، خطی و همصدک سازی)، همصدک‌سازی انتخاب گردید به علت اینکه این روش از لحاظ ویژگی نامتغیر بودن بهتر بود.

لذا، برای قابل مقایسه‌پذیرنمودن نمرات سوابق تحصیلی حاصل از آزمون‌های پیشرفت تحصیلی، آنها را با استفاده از همصدک‌سازی به نمرات خرده آزمون متناظر کنکور مرتبط نمودیم. درواقع، در تبدیل همصدک‌سازی، توزیع نمرات تبدیل‌شده آزمون X (پیشرفت تحصیلی) به مقیاس آزمون Y (خرده آزمون کنکور متناظر با آزمون پیشرفت تحصیلی) با توزیع نمرات آزمون Y یکسان می‌باشند؛ به‌عبارت‌دیگر، تابع همصدک‌سازی نمراتی را روی آزمون X مشخص می‌کند که رتبه‌های صدکی آنها (درصد آزمودنی‌های که نمره آنها کمتر یا مساوی با نمره مورد نظر باشند) برابر با نمراتی روی آزمون Y است که همان رتبه صدکی را داشته باشند (کولن و برنان، ۲۰۰۴).

در نهایت، با استفاده از نرم افزار R خطای استاندارد اندازه‌گیری شرطی برای دو نوع تبدیل غیرخطی؛ نرمال شده و همصدک‌سازی بر روی نمرات حاصل از آزمون‌های پیشرفت تحصیلی محاسبه گردید.

روش تجزیه و تحلیل داده‌ها: نحوه محاسبه تبدیلات غیرخطی نرمال شده و همصدک‌سازی

- نرمال شده

طبق هفته‌نامه پیک سنجش[1] ۸۸۲ (۱۳۹۳)، در سازمان سنجش، با استفاده از تبدیل غیرخطی مقیاس‌سازی، نمره‌های خام به مقیاس نرمال شده تبدیل شدند. برای تعیین این نمرات، لازم است تمامی

۱. نشریه رسمی خبری و اطلاع رسانی سازمان سنجش آموزش کشور

نمره‌های خام داوطلبان در هر گروه آزمایشی و هر درس در دسترس باشند. سپس از روی نمره‌های هر درس به ترتیب نزولی مراحل زیر انجام می شود:

- محاسبه توزیع فراوانی نسبی نمره‌ها $(\hat{g}(y))$:
- محاسبه رتبه درصدی نمره‌ها با استفاده از فراوانی نسبی و توزیع تراکمی نمره‌ها $(\hat{Q}(y))$:
- یافتن نمره z مربوط به نسبت $\frac{\hat{Q}(y)}{۱۰۰}$ از روی معکوس تابع زیر:

$$\phi(z) = \frac{\hat{Q}(y)}{۱۰۰} = \frac{۱}{\sqrt{۲\pi}} \int_{-\infty}^{z} e^{-w^{۲}/۲} dw \qquad (۲۲)$$

تابع فوق، تابع توزیع تراکمی نرمال استاندارد است و W متغیر انتگرال‌گیری با دامنه از $-\infty$ تا z است.

- تبدیل خطی z نمره‌های به دست آمده در مرحله قبل:

$$sc(y) = \sigma(sc)z + \mu(sc) \qquad (۲۳)$$

Sc نمره مقیاس‌بندی شده است. $\mu(sc)$ و $\sigma(sc)$ به ترتیب میانگین و انحراف استاندارد نمره مقیاس بندی شده هستند که در سازمان سنجش برای دانش آموختگان سال ۱۳۹۳به ترتیب برابر با ۵۰۰۰ و ۲۲۵۰ در نظر گرفته شدند (جهانی فر و همکاران، ۱۳۹۶).

- نمره‌های به دست آمده در مرحله قبل به نزدیک‌ترین نمره صحیح گرد می‌شود.

از این رو، داوطلبان کنکور نمی‌توانند نمره تراز درس‌ها –اعم از درس‌های عمومی و اختصاصی– را به دست بیاورند و این کار را سازمان سنجش آموزش کشور برای داوطلبان انجام می‌دهد.

- همصدک‌سازی

تابع همصدک‌سازی، نمراتی را روی آزمون X (آزمون پیشرفت تحصیلی) مشخص می‌کند که رتبه‌های صدکی[۱] آن‌ها برابر با نمرات روی آزمون Y (خرده آزمون کنکور متناظر با آزمون پیشرفت تحصیلی) است که همان رتبه صدکی را داشته باشند. تابع تبدیل نمرات آزمون X به نمرات آزمون Y از طریق همصدک‌سازی $(eq_Y(x))$ به صورت زیر است:

$$eq_Y(x) = G^{-۱}[F(x)] \qquad (۲۴)$$

F و $G^{-۱}$ به ترتیب تابع توزیع تجمعی نمرات X و معکوس تابع توزیع تجمعی Y هستند.

۱. درصد آزمودنی‌های که نمره آن‌ها کمتر یا مساوی با نمره مورد نظر باشند.

روش‌های برآورد خطای استاندارد اندازه‌گیری شرطی: در این پژوهش، برای محاسبه خطای استاندارد اندازه‌گیری شرطی نمره مقیاس تحت تبدیلات نرمال شده و همصدک‌سازی، مبنا را براساس روش کولن، هنسون و برنان (۱۹۹۲) قرار دادیم و احتمال تعداد پاسخ‌های صحیح مشروط به نسبت نمره واقعی خاص را برای هر آزمودنی براساس دوجمله‌ای برنان و لی (۱۹۹۹) و بتا-دوجمله‌ای چنگ[1] (۲۰۰۶) به صورت زیر به دست آوردیم:

روش دوجمله‌ای برنان و لی: یکی از روش‌هایی که توسط آن خطای استاندارد اندازه‌گیری شرطی تحت تبدیلات غیرخطی همصدک سازی و نرمال شده (تراز) محاسبه گردید، روش دوجمله ای برنان ولی (۱۹۹۹) است. بدون کاسته‌شدن از کلیت فرض شرط توزیع دوجمله‌ای، سؤالات آزمون‌های پیشرفت تحصیلی به صورت ساختگی و به شکل دووضعیتی در نظر گرفته شد؛ یعنی هر آزمودنی به سؤال، درست یا غلط پاسخ می‌دهد. احتمال این‌که یک آزمودنی، به Y سؤال از یک آزمون پیشرفت تحصیلی پاسخ صحیح دهد برابر با معادله (۹) درنظر گرفته شد.

از آنجایی که در این پژوهش، آزمون های پیشرفت تحصیلی به صورت تشریحی می باشند و حدود تغییرات نمرات آنها از ۰ تا ۲۰ است، پس بدون کاسته شدن از کلیت فرض، نمرات آزمونهای پیشرفت تحصیلی هر آزمودنی گرد شدند و x و k به ترتیب برابر با نمره کسب شده آزمودنی از آزمون و نمره کل آزمون (یعنی ۲۰) در نظر گرفته شدند و τ هم به عنوان نسبت واقعی پاسخ‌های درست برای یک آزمودنی لحاظ شد که برآورد آن برابر است با تعداد پاسخ‌های درست به نمره کل ؛ یعنی $\bar{x} = \frac{x}{۲۰}$.

برآورد خطای استاندارد اندازه‌گیری شرطی برای هر نمره آزمون پیشرفت تحصیلی (y) که به معادل‌های حاصل از مرتبط‌سازی و نرمال شده (تراز) $(f(y))$ تبدیل شده است، بر اساس معادله (۱۰) محاسبه شدند.

روش بتا-دوجمله‌ای چنگ: روش دیگری که در این پژوهش برای محاسبه خطای استاندارد اندازه-گیری شرطی مورد استفاده قرار گرفت، بر اساس روش چنگ (۲۰۰۶) بود. در این حالت، بدون کاسته‌شدن از کلیت فرض شرط توزیع دوجمله‌ای-بتا، سؤالات آزمون‌های پیشرفت تحصیلی به صورت ساختگی و به شکل دووضعیتی در نظر گرفته شد؛ یعنی هر آزمودنی به سؤال، درست یا غلط پاسخ می‌دهد. تعداد پاسخ‌های آزمودنی (Y) بر روی هر آزمون پیشرفت تحصیلی با k سؤال دارای توزیع بتا-دوجمله‌ای با پارامترهای $\phi = (\alpha, \beta)$ برابربا معادله (۱۵) در نظر گرفته شدند.

سپس با استفاده از نرم افزار R، خطای استاندارد اندازه گیری شرطی به صورت زیر برآورد گردید:

1. Chang

۱. ابتدا نمرات خام هر آزمون پیشرفت تحصیلی به توزیع بتا- دوجمله ای برزاش یافت (همان گونه که یی[1] (۲۰۱۸) توضیح داده است)، سپس تحت این برازش دو مقدار μ و χ در معادله (۱۴) برآورد گردیدند.

۲. از آنجایی که هدف در این پژوهش، برآورد خطای استاندار اندازه گیری شرطی به ازای نسبت واقعی پاسخ‌های صحیح هر آزمودنی است، پس برآورد متوسط نسبت واقعی پاسخ های صحیح (μ) برای هر آزمودنی برابر با $\overline{X} = \frac{x}{k}$ در نظر گرفته شد. از آنجایی که در این پژوهش، آزمون های پیشرفت تحصیلی به صورت تشریحی می باشند و حدود تغییرات نمرات آنها از ۰ تا ۲۰ است، پس بدون کاسته شدن از کلیت فرض، نمرات آزمونهای پیشرفت تحصیلی هر آزمودنی گرد شدند و x و k به ترتیب برابر با نمره کسب شده آزمودنی از آزمون و نمره کل آزمون (یعنی ۲۰) در نظر گرفته شدند.

۳. بر اساس برآورد μ در مرحله ۲ و برآورد χ از مرحله ۱، تابع توزیع بتا-دو جمله ای در نرم افزار R به ازای برآورد نسبت واقعی پاسخ های صحیح هر آزمودنی برآورد شدند.

۴. سپس، خطای استاندارد اندازه گیری شرطی بر اساس معادله (۱۶) برآورد شد.

$f(y)$ تبدیل غیر خطی به ازای هر نمره خام y یی است که مقادیر ۰ تا k را با احتمال بتا- دوجمله‌ای اختیار می‌کند. بر اساس این معیار، خطای استاندارد اندازه‌گیری شرطی برای هر نمره آزمون پیشرفت تحصیلی که به معادل‌های حاصل ازهمصدک سازی و نرمال شده (تراز) تبدیل شده است، محاسبه گردید.

یافته‌ها

در این پژوهش، دو نوع تبدیل غیرخطی بر روی نمرات آزمون‌های پیشرفت تحصیلی؛ «فیزیک»، «شیمی»، «جبر و احتمال»، «هندسه ۲»، «حسابان»، «ریاضیات[2]»، «فرهنگ و معارف اسلامی»، «زبان

1. Yee

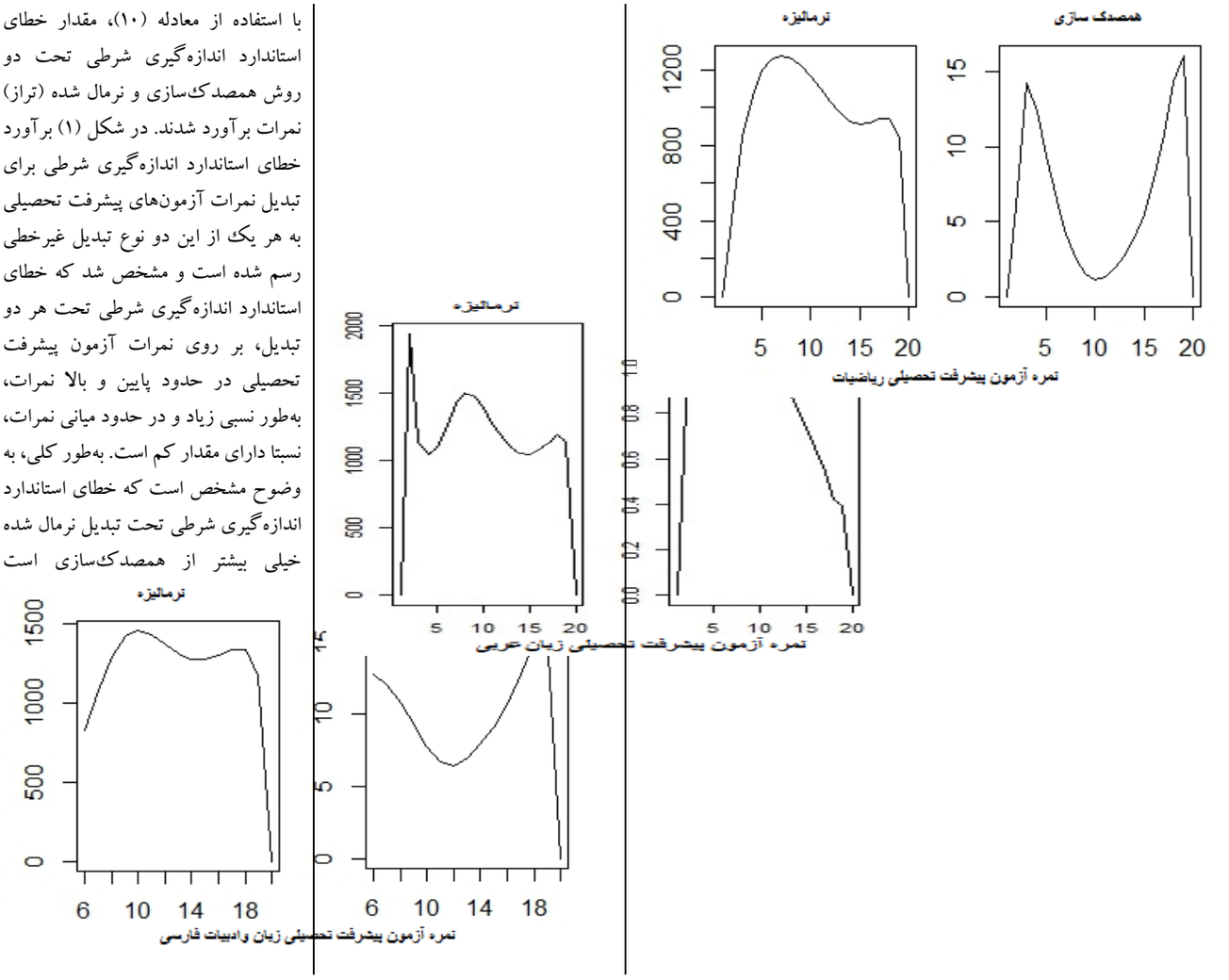
۲. برای مرتبط نمودن نمره هر درس آزمون پیشرفت تحصیلی به نمره خرده‌آزمون کنکور، از همصدک‌سازی استفاده شد. با توجه به این‌که در کنکور تنها یک خرده آزمون نمره ریاضیات است که دارای مباحث؛ «حسابان»، «هندسه»، «جبر و احتمال»، «آمار و مدل‌سازی»، «حساب و دیفرانسیل و انتگرال ۱و۲»، «جبر خطی» و «ریاضیات گسسته» می‌باشد؛ اما سه آزمون مجزا پیشرفت تحصیلی «جبر و احتمال»، «هندسه ۲»، «حسابان» داریم؛ بنابراین میانگین وزنی این نمرات به نمره خرده آزمون ریاضیات کنکور مرتبط می‌شود. از آن‌جایی‌که در سازمان سنجش برای محاسبه نمره کل سوابق تحصیلی به درس‌های جبر و احتمال، هندسه و حسابان به ترتیب وزن‌های ۰/۸۹، ۱/۳۴ و ۱/۷۸ نسبت داده شده است؛ از این وزن‌ها برای محاسبه میانگین وزنی این سه درس استفاده می‌شود. همچنین برای دروس ادبیات وزبان فارسی یه همین صورت است، با این تفاوت که وزن‌های به کاربرده شده در سازمان سنجش برای محاسبه نمره کل سوابق تحصیلی به درس‌های «ادبیات فارسی» و «زبان فارسی» وزن‌های یکسان نسبت داده شده است.

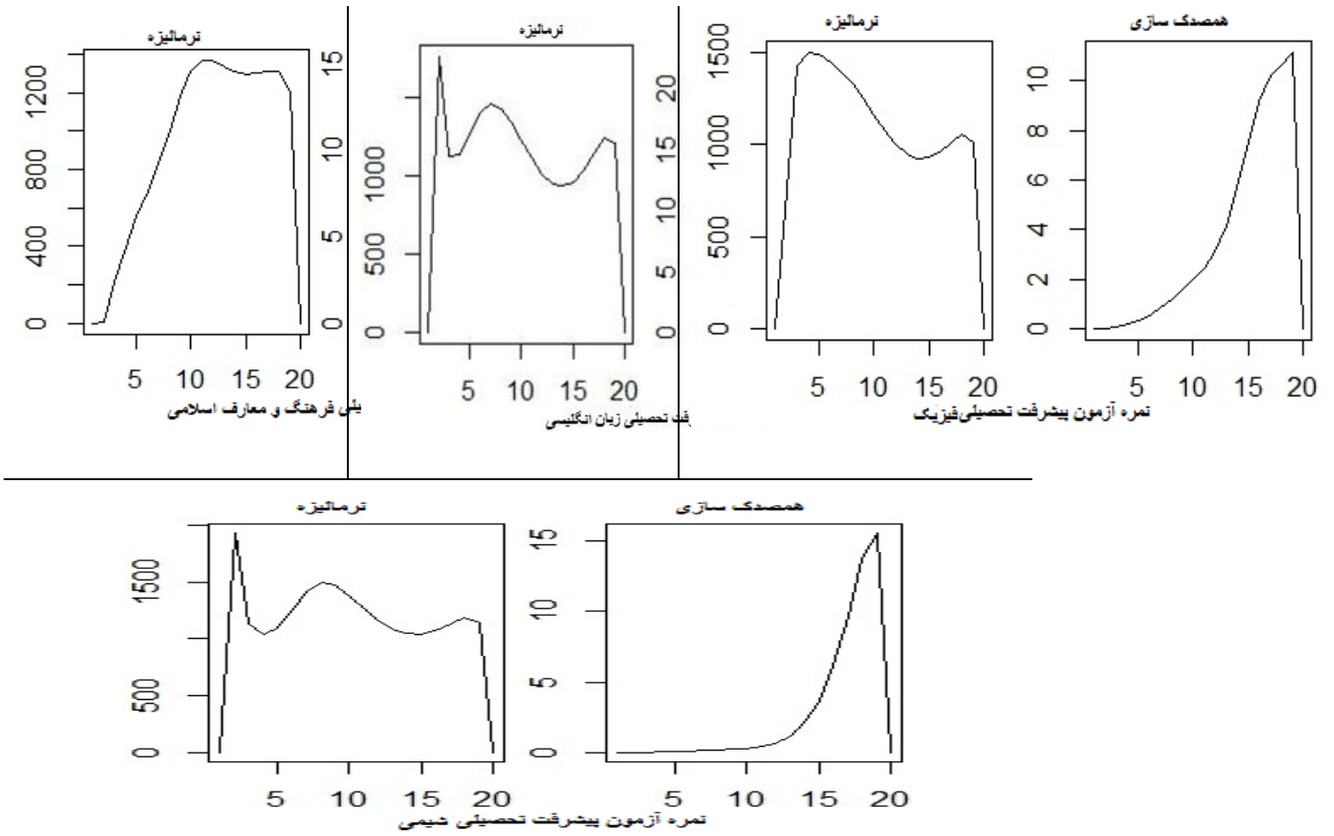
فارسی»، «ادبیات فارسی»، «عربی» و «زبان خارجی» سال سوم دبیرستان که در خرداد سال **۱۳۹۳** برگزار شدند، به کار برده شد.

در این پژوهش، برای محاسبه خطای استاندارد اندازه‌گیری شرطی چندین روش معرفی شد؛ اما چون فقط به نمرات کل آزمون‌های پیشرفت تحصیلی دسترسی داشتیم، نه نمرات سؤالات آزمون‌ها، پس از دو روش دوجمله‌ای برنان و لی (**۱۹۹۹**) و بتا- دوجمله‌ای چنگ (**۲۰۰۶**) استفاده نمودیم.

الف) محاسبه خطای استاندارد اندازه‌گیری شرطی بر طبق روش دوجمله‌ای برنان و لی (**۱۹۹۹**)

با استفاده از معادله (۱۰)، مقدار خطای استاندارد اندازه‌گیری شرطی تحت دو روش همصدک‌سازی و نرمال شده (تراز) نمرات برآورد شدند. در شکل (۱) برآورد خطای استاندارد اندازه‌گیری شرطی برای تبدیل نمرات آزمون‌های پیشرفت تحصیلی به هر یک از این دو نوع تبدیل غیرخطی رسم شده است و مشخص شد که خطای استاندارد اندازه‌گیری شرطی تحت هر دو تبدیل، بر روی نمرات آزمون پیشرفت تحصیلی در حدود پایین و بالا نمرات، به‌طور نسبی زیاد و در حدود میانی نمرات، نسبتا دارای مقدار کم است. به‌طور کلی، به وضوح مشخص است که خطای استاندارد اندازه‌گیری شرطی تحت تبدیل نرمال شده خیلی بیشتر از همصدک‌سازی است





شکل ۱. برآورد خطای استاندارد اندازه‌گیری شرطی برای نمرات حاصل از تبدیلات غیرخطی همصدک‌سازی و نرمال شده تحت توزیع دوجمله‌ای

جدول ۱. مقادیر آمارهای توصیفی نمرات خام آزمون‌های پیشرفت تحصیلی با نمرات برازش‌شده به توزیع دوجمله‌ای

آزمون پیشرفت تحصیلی	نمرات خام				نمرات برازش‌شده به توزیع دوجمله‌ای				
	میانگین	انحراف استاندارد	چولگی	کشیدگی	میانگین	انحراف استاندارد	چولگی	کشیدگی	τ[1]
زبان و ادبیات فارسی	۱۴/۹۲۸	۳/۰۹۵	-۰/۴۱۰	-۰/۷۳۳	۱۴/۹۲۰	۱/۹۴۷	-۰/۲۵۳	-۰/۰۳۶	۰/۷۴۶
زبان عربی	۱۵/۴۶۷	۳/۵۵۳	-۰/۶۵۶	-۰/۳۸۲	۱۵/۴۶۰	۱/۸۷۳	-۰/۲۹۱	-۰/۰۱۵	۰/۷۷۳
فرهنگ و معارف اسلامی	۱۶/۵۰۰	۲/۸۴۰	-۰/۷۸۰	-۰/۰۱۲	۱۶/۵۰۰	۱/۶۹۹	-۰/۳۸۳	۰/۰۴۶	۰/۸۲۵

۱. τ نسبت نمرات صحیح پاسخ داده شده است.

آزمون پیشرفت تحصیلی	نمرات خام				نمرات برازش‌شده به توزیع دوجمله‌ای				
	میانگین	انحراف استاندارد	چولگی	کشیدگی	میانگین	انحراف استاندارد	چولگی	کشیدگی	τ^1
زبان انگلیسی	۱۴/۵۴۸	۳/۹۶۰	-۰/۵۰۶	-۰/۷۶۲	۱۴/۵۴۰	۱/۹۹۲	-۰/۲۲۸	-۰/۰۴۸	۰/۷۲۷
ریاضیات	۱۲/۸۹۴	۴/۰۷۶	۰/۰۲۳	-۱/۰۹۸	۱۲/۹۰۰	۲/۱۴۰	-۰/۱۳۶	-۰/۰۸۲	۰/۶۴۵
فیزیک	۱۲/۰۱۹	۴/۳۲۲	۰/۰۲۰	-۰/۹۳۸	۱۲/۰۲۰	۲/۱۹۰	-۰/۰۹۲	-۰/۰۹۱	۰/۶۰۱
شیمی	۱۴/۷۶۸	۳/۴۷۷	-۰/۵۲۷	-۰/۳۶۷	۱۴/۷۶۰	۱/۹۶۶	-۰/۲۴۲	-۰/۰۴۱	۰/۷۳۸

همان‌گونه‌که جدول (۱) مشخص شد، وقتی نمرات آزمون‌های پیشرفت تحصیلی «فیزیک»، «شیمی»، «جبر و احتمال»، «هندسه ۲»، «حسابان»، «ریاضیات»، « فرهنگ و معارف اسلامی»، «زبان فارسی»، «ادبیات فارسی»، «عربی» و «زبان خارجی» سال سوم دبیرستان به توزیع دوجمله‌ای برازش داده شدند؛ برای تمام آزمون‌ها تقریبا میانگین نمرات خام با میانگین توزیع دوجمله‌ای برازش داده‌شده برابر بودند، اما از لحاظ چولگی و کشیدگی و انحراف استاندارد با یکدیگر تا حدی متفاوت بودند؛ پس از روش بتا-دوجمله‌ای چنگ برای برآورد خطای استاندارد اندازه‌گیری شرطی استفاده نمودیم.

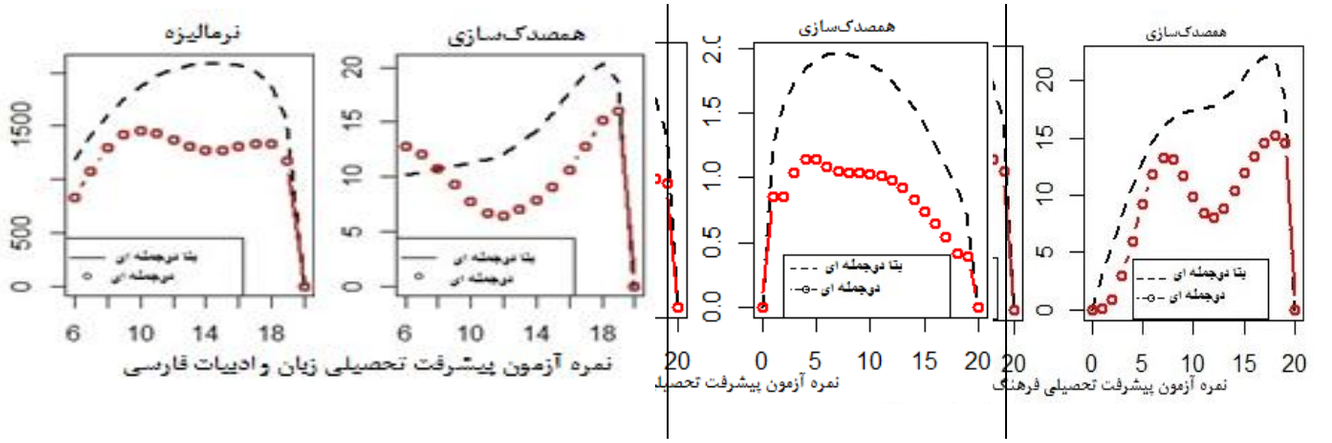
ب) محاسبه خطای استاندارد اندازه‌گیری شرطی طبق روش بتا-دوجمله‌ای چنگ (۲۰۰۶)

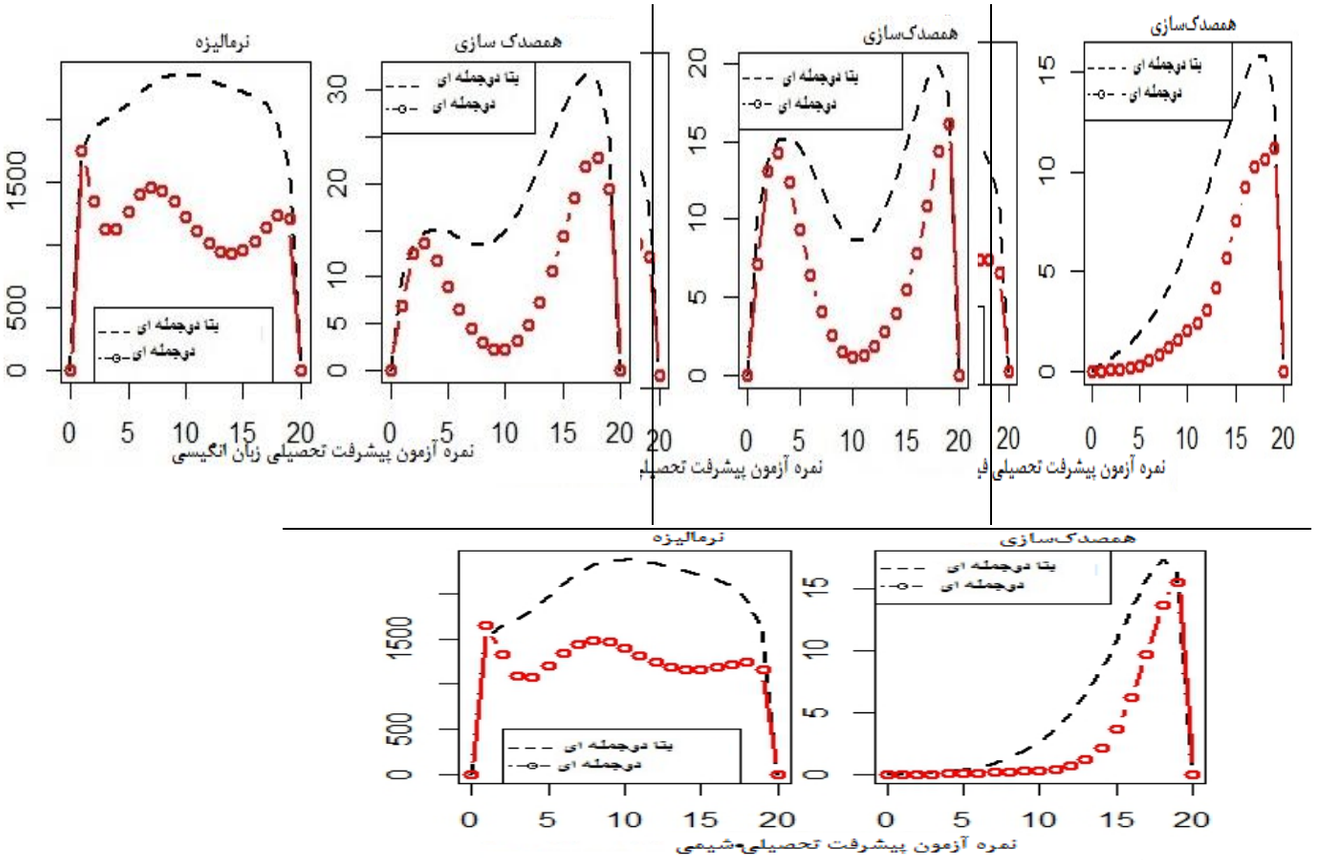
نمرات آزمون‌های پیشرفت تحصیلی را به توزیع بتا-دوجمله‌ای برازش نمودیم. همان‌گونه‌که در جدول (۲) مشخص شده است، تفاوت بین مقادیر آمارهای توصیفی نمرات خام با نمرات برازش‌شده به توزیع بتا-دوجمله‌ای، نسبت به برازش توزیع دو جمله‌ای خیلی کمتر است.

جدول ۲. مقادیر آمارهای توصیفی نمرات خام آزمون‌های پیشرفت تحصیلی با نمرات برازش‌شده به توزیع بتا دوجمله‌ای

آزمون پیشرفت تحصیلی	نمرات خام				نمرات برازش شده به توزیع بتا دوجمله ای				
	میانگین	انحراف استاندارد	چولگی	کشیدگی	میانگین	انحراف استاندارد	چولگی	کشیدگی	α	β
زبان و ادبیات فارسی	۱۴/۹۲۸	۳/۰۹۵	-۰/۴۱۰	-۰/۷۳۳	۱۴/۹۴۰	۳/۱۰۵	-۰/۶۱۵	۰/۰۶۶	۸/۴۰۷	۲/۸۴۷
زبان عربی	۱۵/۴۶۷	۳/۵۵۳	-۰/۶۵۶	-۰/۳۸۲	۱۵/۵۰۰	۳/۵۹۱	-۰/۸۵۲	۰/۲۸۲	۴/۶۸۳	۱/۳۶۰
فرهنگ و معارف اسلامی	۱۶/۵۰۰	۲/۸۴۰	-۰/۷۸۰	-۰/۰۱۲	۱۶/۵۲۰	۲/۸۸۷	-۱/۰۰۶	۰/۸۴۶	۷/۴۳۴	۱/۵۶۶
زبان انگلیسی	۱۴/۵۴۸	۳/۹۶۰	-۰/۵۰۶	-۰/۷۶۲	۱۴/۵۸۰	۳/۹۵۸	-۰/۷۰۹	-۰/۰۶۹	۳/۹۴۴	۱/۴۶۶
ریاضیات	۱۲/۸۹۴	۴/۰۷۶	۰/۰۲۳	-۱/۰۹۸	۱۲/۹۸۰	۴/۱۴۷	-۰/۴۲۰	-۰/۴۶۳	۳/۷۹۶	۲/۰۵۳
فیزیک	۱۲/۰۱۹	۴/۳۲۲	۰/۰۲۰	-۰/۹۳۸	۱۲/۱۰۰	۴/۳۶۳	-۰/۲۹۶	-۰/۶۱۸	۳/۲۴۹	۲/۱۲۱
شیمی	۱۴/۷۶۸	۳/۴۷۷	-۰/۵۲۷	-۰/۳۶۷	۱۴/۷۸۰	۳/۵۰۵	-۰/۶۷۱	۰/۰۲۱	۵/۶۳۱	۱/۹۸۹

برای برآورد خطای استاندارد اندازه‌گیری شرطی به إزای هر مقدار $\hat{\tau}$ (نسبت سؤالات صحیح پاسخ شده توسط هر آزمودنی)، به جای $P(X = i|\tau)$ در معادله (۱۶)، معادله (۱۵) گذاشته شد. در شکل (۲) برآورد خطای استاندارد اندازه‌گیری برای تمام نمرات آزمون‌های پیشرفت تحصیلی رسم شده است. همان‌گونه‌که از شکل مشخص است، همانند روش برآورد خطای از طریق دوجمله‌ای برنان و لی، برای حدود پایین و بالا، نمرات به‌طور نسبی زیاد و در حدود میانی، نمرات نسبتا دارای مقدار کمی است. به‌طور کلی، به وضوح مشخص است که خطای استاندارد اندازه‌گیری شرطی تحت تبدیل نرمال شده، خیلی بیشتر از همصدک‌سازی است. از شکل (۲) مشخص است که برآورد خطای روش بتا-دوجمله‌ای چنگ، با استفاده از برآورد توزیع بتا-دوجمله‌ای نسبت به برآورد خطای استاندارد اندازه‌گیری با استفاده از روش دوجمله‌ای خیلی هموارتر است. همچنین تحت دو روش دو جمله‌ای و بتا-دو جمله‌ای، برآورد خطای استاندارد اندازه‌گیری نسبت به روش تبدیلات غیرخطی همصدک‌سازی کمتر از نرمال شده است.





شکل ۲. مقایسه برآورد خطای استاندارد اندازه‌گیری شرطی برای نمرات حاصل از تبدیلات غیرخطی همصدک‌سازی و نرمال شده تحت توزیع بتا دوجمله‌ای و دوجمله‌ای[1]

بحث و نتیجه‌گیری

در این پژوهش، چندین روش را برای محاسبه خطای استاندارد اندازه‌گیری شرطی براساس نمره واقعی و برای مقایسه تبدیلات غیرخطی همصدک‌سازی و نرمال شده بر روی نمرات آزمون‌های پیشرفت تحصیلی دانش‌آموختگان رشته ریاضی فیزیک سال ۱۳۹۳ معرفی نمودیم. با توجه به این‌که نمرات سؤالات این آزمون‌ها را در اختیار نداشتیم؛ بنابراین در این پژوهش، تنها بعضی از روش‌های ذکرشده برای داده‌های مورد نظر استفاده شد. در این پژوهش، سه روشِ دلتا، دوجمله‌ای برنان و لی (۱۹۹۹) و روش بتا-دوجمله‌ای چنگ (۲۰۰۶) براساس برازش توزیع نمرات خام به توزیع بتا-دوجمله‌ای کارلین و

روبین معرفی شد. روش دلتا شامل برآورد تابع بین نمرات خام و نمرات تحت تبدیلات غیر خطی و خطای استاندارد اندازه گیری شرطی به ازای هر مقدار نمره خام است که برآورد تابع بر اساس روش چند جمله ای، فلت- کوالز و اسپلاین مکعبی معرفی شد. با توجه به این‌که برنان و لی (۱۹۹۹) معتقدند انتخاب درجه چندجمله‌ای می‌تواند براساس برخی از تجزیه و تحلیل داده‌های باشد که نیاز به قضاوت ذهنی و شخصی دارند و از طرفی دیگر، کولن (۱۹۸۴) چنین بیان کرد که اسپیلین‌های مکعبی نسبت به چندجمله‌ای کنترل بیشتری برای هموارکردن دارند اما به دلیل نداشتن نمرات سؤالات آزمون، امکان برآورد خطای استاندارد اندازه‌گیری شرطی برای هر مقدار از نمره از طریق روش‌های ذکر شده، وجود نداشت. لذا، از روشهای دیگری برای محاسبه برآورد خطای استاندارد اندازه‌گیری شرطی استفاده نمودیم که یکی آنها از طریق دوجمله‌ای برنان و لی بود. با استفاده از این روش، خطای استاندارد اندازه‌گیری شرطی هر آزمودنی به إزای نسبت پاسخ‌های صحیح در آزمون برآورد شد. با توجه به این‌که توزیع دوجمله‌ای، نمرات را به خوبی برازش نمی‌دهد، از روش بتا-دوجمله‌ای چنگ استفاده گردید. نمرات خام آزمون‌های پیشرفت تحصیلی به توزیع بتا–دوجمله‌ای برازش داده شد و خطای استاندارد اندازه‌گیری شرطی به إزای نسبت پاسخ سؤالات صحیح هر آزمودنی محاسبه شد. دو مزیت این روش نسبت به دوجمله‌ای برنان و لی این است که نخست آن که برای برازش نمرات خام، توزیع بتا-دوجمله‌ای نسبت به توزیع دوجمله‌ای بهتر بود. در ثانی، خطای استاندارد اندازه‌گیری شرطی براساس توزیع بتا–دوجمله‌ای نسبت به دوجمله‌ای هموارتر است. شایان ذکر است که در هر دو روش دوجمله‌ای برنان و لی و بتا-دوجمله‌ای، برآورد خطای استاندارد اندازه‌گیری شرطی تحت هر دو نوع تبدیلات در نقاط میانی دارای مقدار بالاتر از نقاط ابتدایی و انتهایی حدود نمرات بود و خطای استاندارد اندازه‌گیری شرطی تبدیلات غیرخطی همصدک‌سازی کمتر از نرمال شده بود؛ پس براساس این معیار تبدیلات غیرخطی همصدک‌سازی بهتر از نرمال شده است.

منابع

راهنمای داوطلبان برای انتخاب رشته و ضوابط پذیرش دانشجو در آزمون سراسری سال۱۳۹۳. (۱۳۹۳). *هفته‌نامه پیک سنجش ۸۸۲*، شماره ۱۴، سال ۱۹، ۱-۱۶.

سرمد، زهره؛ بازرگان، عباس؛ حجازی، الهه. (۱۳۸۴). *روش های تحقیق در علوم رفتاری*. تهران: نشر آگاه.

جهانی‌فر، مجتبی؛ خدایی، ابراهیم؛ یونسی، جلیل؛ موسوی، امین. (۱۳۹۶). نقش هموارسازی در تبدیل غیرخطی نمره‌های خام به نمره‌های مقیاس نرمال. فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی، ۷ (۲۰). ۱۳۲-۱۰۳.

Brennan, R. (1998). Raw- Score Conditional Standard Errors of Measurement in Generalizability Theory. Applied Psychological Measurement, 22, 307-331.

Brennan, R. L., & Lee, W. (1999). Conditional Scale-Score Standard Errors of Measurement un¬d¬¬¬er Binomial and Compopund Binomial Assumption. Educational and Psychological Measurement, 59, 5-24.

Carlin, J. B., & Rubin, D. B. (1991). Summarizing Multiple- Choise Tests Using Three Informative Statistics. Psychological Bulletin,11, 338-349.

Chang, S. W. (2006). Methods in Scaling the Basic Competence Test. Educational and Psychological Measurement, 66, 907-929.

Dorans, N. J. (2004).¬¬ Equating, Concordance and Expectation. Applied Psychological Measurement, 28 (4), 227–246.

Feldt, L. S., & Brennan, R. L. (1989). Reliabaility. In R.L. Linn (Ed.), Educational Measurement (3rd ed., pp. 105-146). New York: American Council on Educational and Macmillan.

Feldt, L. S., & Qualls, A. L. (1996). Estimation of measurement error variance at specific score levels. Journal of Educational Measurement, 33, 141-156.

Feldt, L. S., & Qualls, A. L. (1998). Approximating Scale Score Standard Error of Measurement from the Raw Score Standard Error. Applied Measurement in Education, 11, 159-177.

Feldt, L. S., Steffen, M & ,. Gupta, N. C .(1985) . A of Five Methods for Estimating of Measurement Levels. Applied Pshychological Measurement, 9, 351-361.

Kolen, M. J & ,. Brennan, R. L. (2004). Test Equating, Scaling, and Linking: Methods and Practice (3rd Ed .(New York: Springer.

Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional Standard Errors of Measurement for Scale Scores. Journal of Educational Measurement, 29, 285-307.

Lee, W., Brennan, R. L., & Kolen, M. J. (2000). Etimators of Conditional Scale-Score Standard Errors of Measurement: A Simulation Study. Journal of Educational Measurement, 37, 1-20.

Lord, F. M .((1995) . Estimating Test Reliability. Educational Testing Service, 1955, 1-17.

Mollenkopf, W. G. (1949) . Variation of the Standard Eerror of Measurement . Psychometrika,14, 189-229.

Price, L. R., & RAJU, N. (2006). Conditional Standard Errors of Measurement for Composite Scores on the WECHSLER Preschool and Primary Scale of Intelligance-Third Edition. Psychological Reports, 98, 237-252.

Thorndike, R. L. (1951). Reliability. In E. F. Lindquist (Ed.), Educational measurement (pp. 560-620). Washington, DC: American Council on Education.