

کاربرد مدل‌های چند سطحی خطی و معادلات برآورد تعمیم یافته در ارزشیابی روند کیفیت تدریس استادان در ترم‌های مختلف

محمد رضا بانثی^{۱*}، ملیحه رضایی^۲، سمیه شاهرودی^۲، فرزانه ذوالعلی^۳، مریم اخوتی^۴، علی اکبر حق دوست^۵

۱. دکترای تخصصی آمار و اپیدمیولوژی، دانشیار، مرکز تحقیقات مدل‌سازی در سلامت، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
۲. دانشجوی کارشناسی ارشد اپیدمیولوژی، کمیته تحقیقات دانشجویی، دانشگاه علوم پزشکی کرمان، کرمان، ایران
۳. دکترای تخصصی اپیدمیولوژی، استادیار، گروه اپیدمیولوژی و آمار حیاتی، مرکز تحقیقات مدل‌سازی در سلامت، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
۴. ۱. دکترای تخصصی کتابداری و اطلاع‌رسانی پزشکی، استادیار، گروه کتابداری، مرکز تحقیقات انفورماتیک پزشکی، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
۵. دکترای تخصصی اپیدمیولوژی، استاد گروه آمار و اپیدمیولوژی، مرکز تحقیقات مدل‌سازی در سلامت، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، ایران

• دریافت مقاله: ۹۲/۴/۱۹ آخرین اصلاح مقاله: ۹۲/۹/۲۳ • پذیرش مقاله: ۹۲/۱۰/۱

زمینه و هدف: بهبود کیفیت آموزش از محوری‌ترین ارزش‌های نهادهای دانشگاهی و آموزش عالی می‌باشد. با توجه به این‌که استادان نقش تعیین‌کننده‌ای در این راستا دارند، جهت بررسی روند کیفیت آموزش، به ارزشیابی کیفیت تدریس استادان نیاز است. یکی از روش‌های متداول ارزشیابی استادان، ارزشیابی توسط دانشجو می‌باشد که این امر به طور معمول در پایان هر ترم تحصیلی صورت می‌گیرد. با توجه به این‌که داده‌های ارزشیابی استادان از نوع داده‌های طولی با اندازه‌گیری‌های مکرر در طول زمان می‌باشد، برای بررسی روند کیفیت تدریس آنان باید از آنالیزهای مناسب داده‌های طولی استفاده شود، اما متدولوژی چنین روش‌هایی به طور معمول به صورت پیچیده توضیح داده شده است. در مطالعه حاضر به معرفی مدل چند سطحی خطی، روش معادلات برآورد تعمیم یافته و مدل‌های ارزش افزوده (که مدل‌های مناسبی برای داده‌های طولی ارزشیابی می‌باشند) پرداخته شد و سعی گردید که مفهوم آن با زبان ساده و کاربردی توضیح داده شود.

کلید واژه‌ها: ارزشیابی استادان، مدل ارزش افزوده، مدل چند سطحی خطی، روش معادلات برآورد تعمیم یافته

* نویسنده مسؤول: گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی، بلوار هفت باغ، کرمان، ایران

مقدمه

بهبود کیفیت یاددهی و یادگیری از محوری‌ترین ارزش‌های نهادهای دانشگاهی و آموزش عالی است (۱). استادان به عنوان یکی از مهم‌ترین اجزا در سیستم آموزشی، نقش تعیین‌کننده‌ای در دستیابی به اهداف آموزشی دارند (۲). آنچه که در دانشگاه‌ها و مؤسسات آموزش عالی به عنوان یک سیاست مهم در جهت ارتقای کیفیت آموزشی مورد توجه قرار گرفته است، ارزشیابی استادان می‌باشد (۳).

ارزشیابی از استاد را به صورت تعیین میزان موفقیت استاد در رسیدن به اهداف آموزشی تعریف کرده‌اند (۴). برای ارزشیابی استاد به روش‌هایی مانند مصاحبه، مشاهده از کلاس درس، خودسنجی، ارزشیابی به وسیله همکار، بررسی آثار علمی، عضو هیأت علمی و استفاده از نظر دانشجویان می‌توان اشاره کرد (۵) که یکی از رایج‌ترین روش‌ها، ارزشیابی استاد توسط دانشجویان در دانشگاه‌ها و مراکز آموزشی است (۶).

ارزشیابی استاد توسط دانشجو ضمن بحث برانگیز بودن، یکی از پرکاربردترین روش‌ها است؛ به نحوی که هیچ شیوه ارزشیابی دیگری تا به این حد کاربرد ندارد (۷). ارزشیابی توسط دانشجو در دانشگاه‌ها و مراکز آموزشی عالی به طور رسمی از سال ۱۹۲۷ میلادی به عنوان بخش مهمی از ارزشیابی‌های آموزشی به کار گرفته شد (۸). طرفداران این نوع ارزشیابی معتقد هستند که دانشجویان دارای نوعی فراشناخت می‌باشند که باعث می‌شود ارزشیابی آنان از استادان صحیح باشد (۴). این نوع ارزشیابی به صورت مدون طی هر نیم‌سال تحصیلی در دانشگاه‌ها و مراکز آموزش عالی انجام می‌شود. با استفاده از داده‌های ارزشیابی استادان می‌توان به سؤالات مهمی از جمله تأثیر جنسیت، پایه علمی و سابقه کار بر کیفیت تدریس پاسخ داد. کیفیت تدریس استاد به وسیله متغیر کمی نشان داده می‌شود و میانگین نمراتی است که هر استاد برای هر درس از فرم‌های ارزشیابی تکمیل شده توسط دانشجویان در هر نیم‌سال تحصیلی کسب می‌کند. با توجه به این‌که ارزشیابی در هر نیم‌سال تحصیلی و در ترم‌های متوالی تکرار می‌گردد،

برای هر استاد بیش از یک مشاهده (نمره) در دسترس می‌باشد. در واقع داده‌های طولی وجود دارد که در آن مشاهدات ارزشیابی مربوط به یک استاد در ترم‌های متوالی نمی‌تواند مستقل از هم باشد.

چنانچه فقط داده‌های یک ترم موجود باشد، چنین داده‌ای «مقطعی» نامیده می‌شود. برای بررسی عوامل مؤثر بر نمره استادان در داده‌های مقطعی از روش‌های رگرسیونی خطی معمولی استفاده می‌شود. پیش‌فرض اصلی آنالیزهای مقطعی، مستقل بودن داده‌ها از هم می‌باشد. اگر نمرات استادان در ترم‌های مختلف موجود باشد (در واقع داده‌های طولی در دسترس باشد)، نیاز است آنالیزهای پیچیده‌تری صورت گیرد که در آن وابسته بودن نمرات هر فرد در طول ترم‌های مختلف لحاظ گردد. این آنالیزهای پیچیده در واقع تعمیم رگرسیون خطی ساده هستند تا نیازی به فرض مستقل بودن داده‌ها نباشد. به طور معمول مباحث مربوط به تکنیک‌های طولی به شکل پیچیده در کتب آماری یافت می‌شود و خوانندگان غیر متخصص در درک مفهوم آن مشکل دارند. از این‌رو در مطالعه حاضر روش‌های مناسبی مانند مدل‌های آنالیز چند سطحی (آنالیز ضرایب تصادفی یا آنالیز ضرایب آمیخته) و روش معادلات برآورد تعمیم یافته (که در آموزش پزشکی تحت عنوان مدل‌های ارزش افزوده شناخته می‌شود) جهت تحلیل داده‌های ارزشیابی معرفی شد.

آنالیز داده‌های طولی در مجلات داخلی آموزش پزشکی بسیار اندک مورد توجه قرار گرفته‌اند. در بررسی مجله فارسی آموزش پزشکی اصفهان مشخص شد که در ده سال اخیر، هیچ مقاله‌ای تحت عناوین مدل‌های چند سطحی، مدل‌های داده‌های طولی، روش *Generalized estimation equation* (GEE) و روش *Mixed model* به چاپ نرسیده است، تنها یک مورد که در آن داده‌های طولی مورد بررسی قرار گرفته بود، اما از روش ANOVA برای تجزیه و تحلیل داده‌های طولی استفاده شده بود (۹)؛ در حالی که در بررسی مجله *Medical education* طی پنج سال اخیر در حدود ۲۴۴ مقاله یافت شد که در آن بر روی

با تعداد داده‌های کم امکان‌پذیر است (۱۰). از دیگر اشکالات این روش‌ها، از بین رفتن جزئیات و عدم بررسی و روند یا چگونگی تغییرات در طول زمان می‌باشد.

روش‌های پیشرفته

در مطالعات طولی متغیر پیامد موردنظر برای یک فرد در چندین موقعیت متفاوت اندازه‌گیری می‌شود؛ بنابراین باید از تکنیک‌های خاص آماری استفاده گردد و در آن‌ها این اصل که «مشاهدات تکرار شده برای یک فرد یا یک موضوع مشخص، با هم همبستگی دارند» در نظر گرفته شود (۱۱، ۱۰). از جمله مدل‌هایی که در سال‌های اخیر در زمینه ارزشیابی کیفیت تدریس استادان به آن توجه ویژه‌ای شده است، مدل آنالیز چند سطحی یا ارزش افزوده (Value added model) یا (VAM) و معادلات برآورد تعمیم یافته می‌باشد (۱۲). اصلی‌ترین فرض این مدل آن است که توزیع هم‌زمان متغیرهای وابسته (به طور مثال نمرات ارزشیابی استادان در طول ترم‌های مختلف) از یک توزیع نرمال چند متغیره پیروی می‌کند (۱۳). به علاوه، عوامل مختلفی از جمله تعداد دفعات اندازه‌گیری متغیر وابسته و شدت وابستگی نمرات در ترم‌های مختلف بر روی حجم نمونه تأثیر دارد. در این زمینه جداگانه نیز طراحی شده است که بر اساس این ویژگی‌ها حجم نمونه مطلوب را مشخص می‌کند (۱۴).

مدل چند سطحی

تجزیه و تحلیل چند سطحی در دهه ۸۰ میلادی (توسط Goldstion در سال ۱۹۸۷، Cuttance در سال ۱۹۸۸، Nuttall در سال ۱۹۸۹ و Woodhouse در سال ۱۹۹۱) برای انجام تحقیقات آموزشی توسعه داده شد (به نقل از Twisk (۱۱)). تجزیه و تحلیل چند سطحی جهت تجزیه و تحلیل داده‌ها با الگوهای پیچیده و با تمرکز بر ساختار تودرتوی متغیرها می‌باشد. در آنالیز چنین داده‌هایی ارتباطی که متغیرها در هر سطح دارند، در نظر گرفته می‌شود (۱۵).

داده‌های طولی کار شده و از روش‌های آماری مختلفی برای آنالیز استفاده شده بود.

در مطالعه حاضر سعی شد که روش‌های آنالیز تا حد امکان ساده و کاربردی توضیح داده شود. قبل از آن چند روش کلاسیک که زیاد مورد استفاده قرار می‌گیرند، ولی چندان مناسب نیستند، معرفی گردید.

روش‌های کلاسیک

قبل از پیشرفت روش‌های پیچیده آماری از روش‌های کلاسیک برای آنالیز داده‌های طولی استفاده می‌شد. دید کلی روش‌های قدیمی این بود که ساختار طولی داده‌ها را به صورت یک ساختار مقطعی کاهش دهد. حتی امروزه نیز از این روش‌ها برای آنالیز داده‌های طولی استفاده می‌شود.

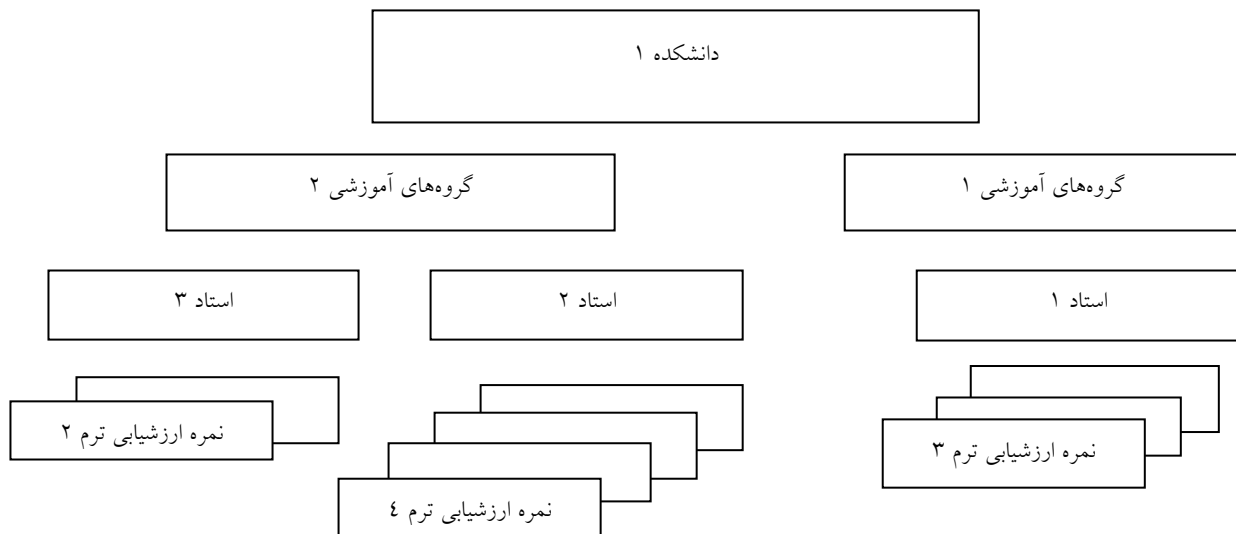
از جمله ساده‌ترین روش‌های کلاسیک، آنالیز ارتباط بین تغییرات دو نقطه زمانی در شاخص‌های مختلف است؛ بدین معنی که نمرات هر ترم به صورت جداگانه آنالیز می‌شود. با این راهکار به طور طبیعی نمی‌توان به بررسی روند تغییرات نمرات استادان در طول ترم‌های مختلف پرداخت. روش دیگر برای آنالیز ارتباط طولی بین چندین متغیر، استفاده از اندازه متغیر در پایان دوره زمانی به عنوان متغیر پاسخ می‌باشد. محدودیت دو روش ذکر شده این است که از همه داده‌های طولی در دسترس استفاده نمی‌کنند.

روش دیگری که از همه داده‌ها برای آنالیز ارتباط طولی بین متغیر پاسخ و متغیرهای پیشگویی کننده استفاده می‌کند، رگرسیون خطی فردی با زمان است. در این روش داده‌های هر استاد به صورت مجزا آنالیز می‌گردد. پس اگر ۳۰۰ عضو هیأت علمی وجود داشته باشند، به ۳۰۰ آنالیز جداگانه نیاز است. پس از محاسبه رگرسیون خطی متغیر پاسخ و زمان برای هر فرد، ضریب رگرسیونی با زمان به عنوان متغیر پاسخ استفاده می‌شود و بدین ترتیب ارتباط طولی با سایر متغیرها بررسی می‌گردد. بزرگ‌ترین اشکال این روش، فرض ارتباط خطی بین متغیر پاسخ و زمان است. علاوه بر این، استفاده از چنین روشی فقط

می‌توان یافت. داده‌های طولی یا داده‌های اندازه‌گیری مکرر می‌توانند به عنوان داده‌های چند سطحی در نظر گرفته شوند که در آن اندازه‌گیری‌های مکرر درون افراد آشیانه شده‌اند. در ساده‌ترین شکل، این داده‌ها به صورت یک مدل دو سطحی می‌باشد که در آن داده‌های طولی در پایین‌ترین سطح و افراد در بالاترین سطح قرار می‌گیرد (۱۸).

انواع بسیار زیادی از داده‌ها از جمله داده‌های مشاهده‌ای جمع‌آوری شده در علوم انسانی دارای ساختار سلسله مراتبی یا خوشه‌ای هستند (۱۶). به عنوان مثال در بررسی سطح علمی دانش‌آموزان در شهر، دانش‌آموزان به عنوان واحدهای سطح یک در داخل مدارس یعنی واحدهای سطح دو قرار گرفته‌اند (۱۷).

همان‌طور که بیان شد، در مطالعات طولی نیز همین اصل را



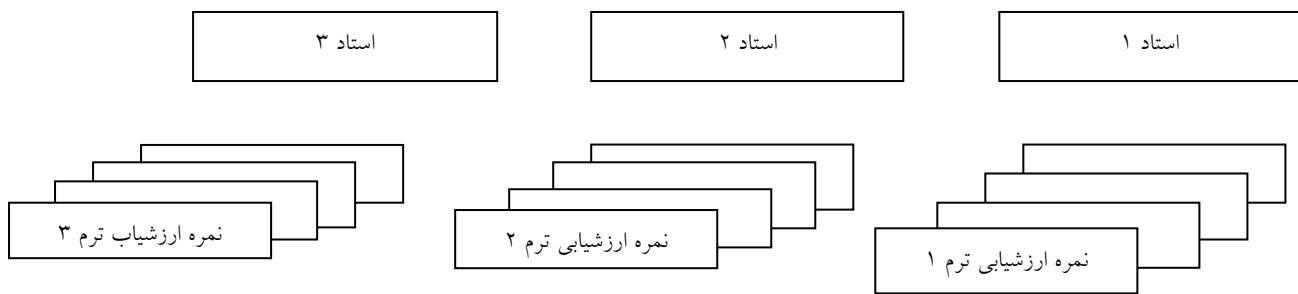
شکل ۱: ساختار چهار سطحی داده‌های مربوط به ارزشیابی استادان

فرد تحت تأثیر گروه‌های اجتماعی که با آن‌ها در تعامل است، می‌باشد و از این گروه‌ها تأثیر می‌پذیرد. برای این‌که کمترین تورش در برآورد پارامترهای مدل وجود داشته باشد، لازم است که این خوشه‌بندی در مشاهدات ارزشیابی در نظر گرفته شود (۱۹).

همان‌طور که پیش‌تر گفته شد، این مدل‌ها تعمیم یافته مدل‌های رگرسیون خطی هستند (۱۱). برای آشنایی با منطق مدل‌های چند سطحی خطی و کاربرد آن در مدل ارزش افزوده، در ذیل مثالی از مدل‌های دو سطحی ارائه شده است (شکل ۲).

بر اساس شکل ۱، مشاهدات مربوط به ارزشیابی هر استاد در ترم‌های متوالی در پایین‌ترین سطح قرار دارد. در سطح دوم استادان و سطح سوم و چهارم به ترتیب گروه‌های آموزشی و دانشکده‌ها می‌باشد. در واقع داده‌های ارزشیابی درون سطح استادان آشیانه شده و سطح استادان درون گروه‌های آموزشی و گروه‌های آموزشی نیز درون دانشکده‌ها خوشه‌بندی شده‌اند.

در مدل چند سطحی خطی نه تنها وابستگی مربوط به ارزشیابی طولی استادان در نظر گرفته می‌شود، بلکه وابستگی که ممکن است درون هر گروه آموزشی و دانشکده وجود داشته باشد نیز در مدل وارد می‌شود تا تأثیری که گروه‌های آموزشی و دانشکده‌ها بر استادان دارند، محسوب گردد؛ چرا که برداشت کلی در بررسی‌ها و تحقیقات اجتماعی این است که



شکل ۲: ساختار دو سطحی داده‌های مربوط به ارزشیابی استادان

$$\text{Year} + \beta_2 \times \text{Gender} + \beta_3 \times \text{Gender} \times \text{Year} + \varepsilon$$

$$\text{Score} = \beta_0 + \beta_1 \times$$

اگر مقادیر صفر و یک برای متغیر جنسیت در معادله جایگذاری شود، هم عرض از مبدأ خط‌ها و هم شیب آن‌ها متفاوت است. همین ایده را می‌توان به شکل کلی‌تر زیر بسط داد:

چنانچه برای هر استاد یک خط رگرسیونی در نظر گرفته شود که رابطه بین سابقه کار و نمره ارزشیابی را نشان دهد، حدس زده می‌شود که روند صعودی یا نزولی نمرات در طول زمان از فردی به فرد دیگر متفاوت باشد. در واقع اگر عرض از مبدأ و شیب خطوط از فردی به فرد دیگر تغییر کند (به دلیل این‌که متغیر استادان یک متغیر اسمی است)، باید متغیرهای دو حالتی ساخته شود؛ یعنی اگر ۳۰۰ استاد باشد، لازم است ۲۹۹ متغیر دو حالتی ساخته شود و در مدل قرار گیرد تا ضرایب رگرسیونی مربوطه محاسبه گردد. این مسأله باعث کاهش توان و کارایی مدل می‌شود؛ چرا که تعداد شاخص‌هایی که باید برآورد کنیم، بسیار زیاد است.

یک راه‌حل کارا و پرتوان برای رفع این مشکل، استفاده از مدل‌های چند سطحی خطی می‌باشد. در مدل‌های چند سطحی لازم نیست چندین عرض از مبدأ و شیب خط برآورد گردد، بلکه واریانس آن‌ها محاسبه و در مدل وارد می‌شود. در برآورد این واریانس‌ها اجازه داده می‌شود که عرض از مبدأها تصادفی باشند؛ به همین دلیل به مدل‌های چند سطحی خطی «آنالیز ضرایب تصادفی» نیز می‌گویند. در این حالت کافی است

چنانچه مطالعه‌ای برای بررسی رابطه بین سال‌های کاری استادان (Year) و نمره ارزشیابی (Score) آن‌ها در طی چند سال انجام گیرد؛ مدل رگرسیون خطی که برای داده‌های مقطعی کاربرد دارد، به صورت زیر توصیف می‌گردد:

$$\text{Score} = \beta_0 + \beta_1 \times \text{Year} + \varepsilon$$

در این مدل Score متغیر پیامد و میانگین نمرات ارزشیابی توسط دانشجویان طی چند ترم می‌باشد. β_0 عرض از مبدأ مدل، β_1 شیب خط مدل، Year متغیر مستقل وابسته به زمان و ε خطای مدل می‌باشد. اگر محقق بخواهد اثر متغیر جنسیت استاد را نیز به مدل اضافه کند، در این حالت مدل به صورت زیر ارائه می‌شود:

$$\text{Score} = \beta_0 + \beta_1 \times \text{Year} + \beta_2 \times \text{Gender} + \varepsilon$$

در مدل فوق β_2 ضریب رگرسیونی مربوط به جنسیت می‌باشد. چنانچه جنس مرد با کد صفر و جنس زن با کد یک مشخص گردد، اگر این مقادیر برای متغیر جنسیت در معادله قرار گیرد، مشاهده می‌گردد که β_0 عرض از مبدأ خط رگرسیونی مربوط به مردان و $\beta_0 + \beta_2$ عرض از مبدأ خط رگرسیونی مربوط به زنان خواهد بود؛ بنابراین وقتی که جنسیت در مدل اضافه و تصحیح می‌شود، عرض از مبدأ خطوط رگرسیونی برای دو جنس مرد و زن فرق خواهد کرد و در عمل دو خط رگرسیونی با شیب‌های برابر و عرض از مبدأهای متفاوت وجود دارد (۱۱).

چنانچه مقدار شیب خط برای مردان و زنان متفاوت باشد، باید اثر متقابل جنسیت در سابقه کار وارد شود؛ یعنی مدل به صورت زیر خواهد بود:

داده‌های طولی با استفاده از برآورد Quasi-likelihood است (۲۱).

معادلات برآورد تعمیم یافته چون بر فرضیات آماری کلاسیک (استقلال داده‌ها و نرمال بودن آن‌ها) غلبه کرد، به طور افزایشی در کاربردهای علوم زیستی، اقتصادی و روان‌شناسی محبوبیت یافت (۲۰). این روش مشخص می‌کند که چگونه میانگین متغیر پاسخ یک فرد به وسیله متغیرهای همراهی (Covariates) (زمانی که همبستگی بین اندازه‌های تکرار شده یک فرد در طول زمان وجود دارد)، تغییر می‌کند (۲۲).

در روش معادلات برآورد تعمیم یافته می‌توان با در نظر گرفتن یک ساختار همبستگی، همبستگی بین مشاهدات را در آنالیز لحاظ کرد (۱۰). پنج نوع ساختار همبستگی اصلی وجود دارد که می‌توان با توجه به ساختار داده‌ها انتخاب کرد. فرم ماتریسی این ساختارهای همبستگی در پیوست ارائه شده است.

آزمون کنیم که آیا این واریانس اختلاف معنی‌داری با عدد صفر دارد یا خیر؟

مفهوم مهم دیگری که در مدل‌های چند سطحی وجود دارد، ضریب همبستگی درون خوشه‌ای (ICC یا Intraclass correlation coefficient) است که میزان همبستگی مشاهداتی که متعلق به یک خوشه هستند را بیان می‌کند (۱۶، ۱۱). بالا بودن این شاخص نشان دهنده همبستگی شدید داده‌ها درون هر خوشه می‌باشد (۱۱).

در حال حاضر با توجه به کاربرد روزافزون روش آنالیز چند سطحی، این روش پیشرفت قابل ملاحظه‌ای کرده است و بیشتر نرم‌افزارهای آماری مانند SPSS، Stata، SAS و R توانایی انجام این مدل آماری را دارند.

روش معادلات برآورد تعمیم یافته Zeger و Liang معرفی شد (به نقل از Nakai و Ke) (۲۰). این روش راهکار متفاوتی را برای حل مشکل وابستگی بین داده‌ها ارائه می‌کند. روش معادلات برآورد تعمیم یافته گسترده مدل‌های خطی تعمیم یافته (Generalized linear models) برای آنالیز

پیوست ۱. شکل ماتریسی ساختارهای همبستگی

	t_1	t_2	t_3	t_4	t_5	t_6
t_1	—	0	0	0	0	0
t_2	0	—	0	0	0	0
t_3	0	0	—	0	0	0
t_4	0	0	0	—	0	0
t_5	0	0	0	0	—	0
t_6	0	0	0	0	0	—

	t_1	t_2	t_3	t_4	t_5	t_6
t_1	—	ρ	ρ	ρ	ρ	ρ
t_2	ρ	—	ρ	ρ	ρ	ρ
t_3	ρ	ρ	—	ρ	ρ	ρ
t_4	ρ	ρ	ρ	—	ρ	ρ
t_5	ρ	ρ	ρ	ρ	—	ρ
t_6	ρ	ρ	ρ	ρ	ρ	—

شکل ۲. ساختار همبستگی مستقل

	t_1	t_2	t_3	t_4	t_5	t_6
t_1	—	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5
t_2	ρ^1	—	ρ^1	ρ^2	ρ^3	ρ^4
t_3	ρ^2	ρ^1	—	ρ^1	ρ^2	ρ^3
t_4	ρ^3	ρ^2	ρ^1	—	ρ^1	ρ^2
t_5	ρ^4	ρ^3	ρ^2	ρ^1	—	ρ^1
t_6	ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	—

شکل ۴. ساختار همبستگی اتورگرسیو

شکل ۱. ساختار همبستگی قابل مبادله

	t_1	t_2	t_3	t_4	t_5	t_6
t_1	—	ρ_1	ρ_2	0	0	0
t_2	ρ_1	—	ρ_1	ρ_2	0	0
t_3	ρ_2	ρ_1	—	ρ_1	ρ_2	0
t_4	0	ρ_2	ρ_1	—	ρ_1	ρ_2
t_5	0	0	ρ_2	ρ_1	—	ρ_1
t_6	0	0	0	ρ_2	ρ_1	—

شکل ۳. ساختار همبستگی ساکن ($m=2$)

رابطه توانی متناسب با فاصله ترم‌ها از یکدیگر باشد، ساختار مناسب برای همبستگی مشاهدات، ساختار اتورگرسیو در نظر گرفته می‌شود.

آخرین ساختار همبستگی ممکن «ساختار همبستگی Unstructured» نامیده می‌شود که با این ساختار همه اندازه‌های همبستگی بین اندازه‌ها با هم تفاوت دارند (۱۰). برای مثال اگر در نمره‌های ارزشیابی استاد در ترم‌های مختلف روند خاصی از همبستگی دیده نشود و اندازه‌های همبستگی نمرات با هم متفاوت باشد، ساختار همبستگی بدون ساختار برای داده‌ها در نظر گرفته می‌شود.

اگرچه روش معادلات برآورد تعمیم یافته نسبت به ساختار همبستگی نادرست قوی است، اما انتخاب ساختار اشتباه موجب نتیجه‌گیری نادرست در مورد ارتباط طولی بین متغیرها می‌شود. در انتخاب ساختار همبستگی باید سادگی ساختار، تعداد شاخص کم (ضرایب همبستگی) برای برآورد و این که ساختار انتخابی به خوبی به داده‌ها برازش شود، مدنظر قرار گیرد (۱۰). شاخص QIC (Quasilikelihood under the independence model criterion) نیز شاخصی برای انتخاب ساختار همبستگی مناسب برای داده‌ها در روش معادلات برآورد تعمیم یافته می‌باشد (۲۳، ۲۲). بعد از مشخص کردن ساختار همبستگی، معادله رگرسیونی در روش GEE به صورت زیر می‌باشد:

$$Y_{it} = \beta_0 + \sum_{j=1}^J \beta_{1j} X_{itj} + \beta_2 t + \dots + \text{CORR}_{it} + \varepsilon_{it}$$

در این معادله، Y_{it} مشاهدات برای فرد i ام در زمان t ، β_0 مقدار ثابت، X_{itj} متغیر مستقل i ام برای فرد i ام در زمان t ، β_{1j} ضریب رگرسیونی برای متغیر مستقل j ام، J تعداد متغیرهای مستقل، t زمان، β_2 ضریب رگرسیونی برای زمان، CORR_{it} ساختار همبستگی و ε_{it} خطا برای فرد i ام در زمان t می‌باشد. آنالیز با روش GEE را می‌توان از طریق نرم‌افزارهای R، SAS و Stata انجام داد.

اولین ساختار ممکن، «ساختار همبستگی Independent» می‌باشد که فرض می‌کند همبستگی بین مشاهدات صفر است (۱۰). به عنوان مثال اگر فرض بر آن باشد که نمره ارزشیابی کسب شده توسط استاد در یک ترم از نمره ارزشیابی که در ترم‌های قبل کسب کرده تأثیر نپذیرفته است، ساختار همبستگی مستقل برآورد می‌شود.

انتخاب دیگر برای ساختار همبستگی «ساختار همبستگی Exchangeable» می‌باشد. در این ساختار، همبستگی بین اندازه‌ها مقدار مشابهی فرض می‌شود (۱۰). به عنوان مثال ارتباط بین نمره ارزشیابی استاد در یک ترم با ترم قبل برابر ارتباط نمره استاد در همان ترم و دو ترم قبلی می‌باشد. به بیان دیگر، تأثیر نمره ارزشیابی استاد در هر ترم از ترم‌های قبلی یکسان و ثابت بوده است.

سومین ساختار همبستگی ممکن «ساختار همبستگی M-Dependent (or stationary)» می‌باشد. فرض این است که همبستگی اندازه‌های با فاصله زمانی t با هم و همبستگی اندازه‌های با فاصله زمانی $t + 1$ نیز با هم برابر هستند و به همین صورت از $t = 1$ تا $t = m$. همبستگی اندازه‌های با فاصله زمانی بیشتر از m نیز صفر فرض می‌شوند (۱۰). به عنوان مثال در حالت $t = 1$ ، اگر میزان تأثیری که نمره ارزشیابی استاد در یک ترم از ترم قبل می‌گیرد برابر با میزان تأثیری باشد که نمره ارزشیابی استاد در ترم آینده از ترم جاری خواهد گرفت و این روند فقط در ترم‌های متوالی دیده شود و در ترم‌های با فاصله زمانی بیشتر از یک، ارتباطی وجود نداشته باشد، ساختار همبستگی به عنوان ساختار همبستگی ساکن با $t = 1$ در نظر گرفته می‌شود.

چهارمین ساختار ممکن «ساختار همبستگی Autoregressive» می‌باشد که در آن همبستگی اندازه‌های با فاصله زمانی یک برابر با ρ ، همبستگی اندازه‌های با فاصله زمانی دو برابر با ρ^2 و همبستگی اندازه‌های با فاصله زمانی t برابر با ρ^t در نظر گرفته می‌شود (۱۰). به عنوان مثال، اگر ارتباط بین نمره ارزشیابی استاد در هر ترم با ترم‌های دیگر

بحث و نتیجه‌گیری

لحاظ می‌شود (۱۰). همان گونه که بیان شد، در نظر نگرفتن یک ساختار خاص موجب می‌شود که نتایج از نظر کیفی نادرست باشد. در پایان با توجه به اهمیت ارزشیابی کیفیت تدریس استادان و با استفاده از مدل‌های چند سطحی خطی و معادلات برآورد تعمیم یافته (که به عنوان روش‌هایی برای آنالیز داده‌های طولی ارزشیابی استادان) خیلی مناسب می‌باشد (۱۰). چنین طرح‌هایی هم از نظر آماری و هم از نظر علمی بسیار قوی هستند؛ چرا که این توانایی را به محقق می‌دهد تا تغییرات درون فردی را در طول زمان یا تحت تغییر شرایط مختلف مطالعه کند و علت توان بالای این طرح‌ها در آن است که اندازه‌های تکرار شده به همبستگی با یکدیگر گرایش دارند و این همبستگی باید در زمان تجزیه و تحلیل داده‌ها تعیین شود (۲۴).

در روش معادلات برآورد تعمیم یافته، این همبستگی به وسیله فرض یک ساختار همبستگی لحاظ می‌شود و در مدل چند سطحی به وسیله ضرایب رگرسیونی برای تغییر بین افراد

ویژگی اصلی بسیاری از تحقیقات اپیدمیولوژی و علم کارآزمایی، مطالعات با اندازه‌های تکرار شده می‌باشد (۲۴). مدل‌های چند سطحی و معادلات برآورد تعمیم یافته برای آنالیز داده‌های طولی (مانند داده‌های ارزشیابی استادان) خیلی مناسب می‌باشد (۱۰). چنین طرح‌هایی هم از نظر آماری و هم از نظر علمی بسیار قوی هستند؛ چرا که این توانایی را به محقق می‌دهد تا تغییرات درون فردی را در طول زمان یا تحت تغییر شرایط مختلف مطالعه کند و علت توان بالای این طرح‌ها در آن است که اندازه‌های تکرار شده به همبستگی با یکدیگر گرایش دارند و این همبستگی باید در زمان تجزیه و تحلیل داده‌ها تعیین شود (۲۴).

- در روش معادلات برآورد تعمیم یافته، این همبستگی به وسیله فرض یک ساختار همبستگی لحاظ می‌شود و در مدل چند سطحی به وسیله ضرایب رگرسیونی برای تغییر بین افراد
- Dehghani M, Nakhaee N. Faculty evaluation by students: a review of criticisms. *Strides Dev Med Educ* 2012; 9(2):102-9. [In Persian]
 - Shakurnia A, Fakour M, Elhampoor H, Taherzadeh M, Chab F. Consistent with the faculty Self ' evaluation scores by Evaluation by students: students in Ahvaz Jundishapur University of Medical Sciences in 2009. *Iran J Educ Med Sci* 2010; 10(3): 229-37. [In Persian]
 - Shakurnia A. Teachers' attitudes towards students' evaluations: Viewpoints of students will it really matter? *Iran J Educ Med Sci* 2011; 11(2): 84-93. [In Persian]
 - Adib-Hajbaghery M, Azizi-Fini E. Longitudinal Study of CardioPulmonary Resuscitation Knowledge and Skills among Nurse Interns of Kashan University of Medical Sciences. *Iran J Med Educ* 2013; 13(2):134-45. [In Persian]
 - Twisk JWR. Applied longitudinal data analysis for epidemiology. 2nd ed. New York: Cambridge University Press; 2013.
 - Twisk JWR. Applied multilevel analysis: a practical guide for medical researchers. New York: Cambridge University Press; 2006.

References:

- Yamani N, Yousefy A, Changiz T. Proposing a Participatory Model of Teacher Evaluation. *Iran J Med Sci Educ* 2006; 6(2): 120-7. [In Persian]
- Ranjbar M, Vahid Shahi K, Mahmoudi M. Viewpoints of the attending and medical students about the Students' evaluation of the attending, Mazandaran University of Medical Sciences, Sari, 2005. *J Mazandaran Univ Med Sci* 2007; 16(56): 116-26. [In Persian]
- Tootoonchi M, Changiz T, Alipour L, Yamani N. Faculty Members' Viewpoints towards Teacher Evaluation Process in Isfahan University of Medical Science. *Iran J Med Sci Educ* 2006; 6(1):23-31. [In Persian]
- Amini M, Honardar M. The view of faculties and medical students about evaluation of faculty teaching experiences. *Koomesh* 2008; 9(3): 171-8. [In Persian]
- Kaikhaei A, Navidian A, Tabassi MA, Sargaz GH. The view of Teachers' Zahedan University of Medical Sciences about evaluations of faculty members. *Zahedan J Res Med Sci* 2003; 4(3): 135-40 [In Persian]

12. Goe L, Bell C, Little O. Approaches to evaluating teacher effectiveness: A research synthesis. Washington, DC: National Comprehensive Center for Teacher Quality; 2008.
13. Rencher AC, Christensen WF. Methods of multivariate analysis. 3rd ed. New York: John Wiley & Sons, INC; 2012
14. Twisk JW. Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide. 2nd ed. New York: Published by Cambridge University Press; 2013
15. Snijders T, Bosker R. Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: Sage; 2011.
16. Goldstion H. Multilevel statistical models. London: Arnold publishers; 1999.
17. Goldstion H. Multilevel models in education and social research. London: Oxford University Press; 1987.
18. Hox JJ. Applied multilevel analysis. Amsterdam: TT-publikaties; 1995.
19. Amir Kafi M. The Importance and logic of multilevel models in social research. Iran J Sociol 2007; 7(4): 38-71. [In Persian].
20. Ziegler A, Vens M. Generalized Estimating Equations. Notes on the choice of the working correlation matrix. Methods Inf Med 2010; 49(5):421-5
21. Nakai M, Ke W. Statistical Models for Longitudinal Data Analysis. Applied Mathematical Sciences 2009; 3(40): 1979-89.
22. Cui J. QIC program and model selection in GEE analyses. Stata Journal 2007;7(2): 209-20.
23. Jang MJ. (dissertation). Working correlation selection in generalized estimating equations. Iowa: The University of Iowa; 2011.
24. Burton P, Gurrin L, Sly P. Tutorial in biostatistics. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modeling. Stat Med 1998; 17(11): 1261-91.