

## مدل‌سازی چند سطحی داده‌های طولی ارزشیابی کیفیت تدریس اعضای هیأت علمی دانشگاه علوم

### پزشکی کرمان

سمیه شاهرودی<sup>۱</sup>، علی‌اکبر حق‌دوست<sup>۲</sup>، فرزانه ذوالعلی<sup>۳</sup>، مریم اخوتی<sup>۴</sup>، محمدرضا بانשי<sup>۵\*</sup>

۱. دانشجوی کارشناسی ارشد آمار زیستی، کمیته تحقیقات دانشجویی، دانشگاه علوم پزشکی کرمان، کرمان، ایران

۲. دکتری اپیدمیولوژی، استاد، مرکز منطقه‌ای آموزش نظام مراقبت HIV/ایدز، مرکز همکار سازمان بهداشت جهانی، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران

۳. دکتری اپیدمیولوژی، استادیار، مرکز منطقه‌ای آموزش نظام مراقبت HIV/ایدز، مرکز همکار سازمان بهداشت جهانی، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران

۴. دکتری کتابداری و اطلاع‌رسانی، استادیار، مرکز تحقیقات فیزیولوژی، پژوهشکده نوروفارماکولوژی، دانشگاه علوم پزشکی کرمان، کرمان، ایران

۵. دکتری آمار زیستی، دانشیار، مرکز تحقیقات مدل‌سازی در سلامت، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران

• دریافت مقاله: ۹۳/۱/۲۹ آخرین اصلاح مقاله: ۹۳/۶/۱۵ • پذیرش مقاله: ۹۳/۶/۱۶

**زمینه و هدف:** اعضای هیأت علمی دانشگاه علوم پزشکی کرمان به منظور ارتقای کیفیت آموزشی، هر نیم‌سال تحصیلی توسط دانشجویان مورد ارزشیابی قرار می‌گیرند. این مطالعه با هدف بررسی چگونگی تغییرات نمرات ارزشیابی اعضای هیأت علمی توسط دانشجو در دانشگاه علوم پزشکی کرمان در طول ترم‌های مختلف و عوامل مؤثر بر آن با استفاده از مدل‌سازی چند سطحی و مقایسه آن با برازش مدل رگرسیون خطی ساده انجام گرفت.

**روش کار:** در مطالعه تحلیلی حاضر اطلاعات مربوط به ارزشیابی کیفیت تدریس ۳۳۶ نفر از اعضای هیأت علمی دانشگاه علوم پزشکی کرمان که در فاصله زمانی سال‌های ۹۱-۱۳۸۷ از طریق وب‌سایت پایش توسط دانشجویان به وسیله پرسش‌نامه الکترونیکی ارزشیابی شده بودند، استخراج گردید. کد مدرس به عنوان عامل شناسایی امتیازات او در ترم‌های مختلف در نظر گرفته و در مدل‌های چند سطحی وارد شد. برای بررسی تأثیر سابقه آموزشی، جنسیت، مرتبه علمی و دانشکده مدرس بر امتیازات، دو مدل چند سطحی و رگرسیون خطی مورد مقایسه قرار گرفت. جهت بررسی معنی‌دار بودن عرض از مبدأ تصادفی، از آزمون نسبت درست‌نمایی و برای بررسی نکویی برازش دو مدل از شاخص اطلاع (Akaike)  $Akaike$  (Bayesian information criterion یا AIC)، شاخص اطلاع (Bayesian information criterion یا BIC) و میانگین مربعات خطا (Mean squared error یا MSE) استفاده گردید.

**یافته‌ها:** یافته‌ها نشان داد که گذشت زمان در بهبود نمره ارزشیابی اعضای هیأت علمی اثر مستقیم، مثبت و معنی‌داری داشت، اما تأثیر جنسیت معنی‌دار نبود. علاوه بر این، نمره مرتبه علمی استادان به طور معنی‌داری از استادیاران و همچنین نمره همکاران دانشکده دندان‌پزشکی به طور معنی‌داری از دانشکده پزشکی بالاتر بود. مقایسه نکویی برازش دو مدل مشخص کرد که روش مدل‌سازی چند سطحی در تحلیل داده‌های طولی عملکرد بهتری از خود نشان داد. در مدل رگرسیون خطی متغیرهایی همچون سابقه آموزشی، مرتبه علمی مربی نسبت به استادیار و نمرات همکاران دانشکده‌های پرستاری و داروسازی نسبت به پزشکی به علت در نظر نگرفتن شرط وابستگی مشاهدات طولی ارزشیابی و عدم تصحیح خطای معیار، به غلط معنی‌دار شد.

**نتیجه‌گیری:** مدل چند سطحی به دلیل در نظر گرفتن وابستگی مشاهدات طولی ارزشیابی، برازندگی بهتری دارد. علاوه بر این، استفاده ناصحیح از مدل رگرسیون خطی با فرض استقلال مشاهدات طولی منجر به استنتاج‌های ناصحیح شد. نتایج پژوهش از برازش هر دو مدل چند سطحی و رگرسیون خطی نشان داد که روند نمرات طولی ارزشیابی اعضای هیأت علمی دانشگاه علوم پزشکی کرمان توسط دانشجویان یک روند صعودی است و به عبارت دیگر، نمرات ارزشیابی در طول زمان بهتر شده است.

**کلیدواژه‌ها:** ارزشیابی استاد، دانشجو، ارزشیابی ارزش افزوده، داده‌های طولی، مدل‌سازی چند سطحی

\*نویسنده مسئول: مرکز تحقیقات مدل‌سازی در سلامت، پژوهشکده آینده‌پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران

## مقدمه

یکی از مفاهیمی که شناخت دقیق و علمی آن نقش مهمی در توسعه کمی و کیفی دانشگاه‌ها و مؤسسات آموزش عالی دارد، ارزشیابی آموزشی است (۱). منظور از ارزشیابی، بررسی سیستماتیک یک موضوع برای بهبود بخشیدن به برنامه‌ها می‌باشد (۲). ارزیابی عملکرد اعضای هیأت علمی نیز نوعی ارزشیابی با هدف اصلاح فعالیت‌های آموزشی، کیفیت تدریس و کمک به مدیران برای تصمیم‌گیری درباره مواردی همچون استخدام، ترفیع و... است (۳). از جمله روش‌های پیشرفته ارزشیابی اعضای هیأت علمی، ارزشیابی مدل ارزش افزوده (Value added models یا VAM) می‌باشد که اساس آن بر مشاهدات طولی (Longitudinal data) ارزشیابی اعضا و تحلیل آن با استفاده از تکنیک‌های آماری صورت می‌گیرد (۴). داده‌های طولی یا داده‌های اندازه‌گیری‌های مکرر مشاهداتی هستند که در آن یک موضوع یا یک متغیر در طول زمان به دفعات و بنا بر اهداف مشخص اندازه‌گیری می‌شود. این نوع داده‌ها مستقل از هم نیستند، بلکه مشاهدات مربوط به یک فرد یا یک موضوع به هم وابسته هستند (۵). نمرات ارزشیابی ثبت شده برای یک عضو هیأت علمی در طی چند ترم متوالی از نوع مشاهدات طولی است، بنابراین برای تحلیل این مشاهدات و رسیدن به نتایج درست و منطقی به تکنیک‌های پیشرفته آماری نیاز می‌باشد که در آن اصل رعایت شود. یکی از این تکنیک‌های پرتوان، روش آنالیز چند سطحی (Multilevel model) مشاهدات طولی است (۶).

روش استاندارد جهت بررسی تأثیر متغیرهای مستقل (مانند جنسیت و رتبه علمی) بر یک متغیر وابسته (مانند نمره اعضای هیأت علمی)، رگرسیون خطی (Linear regression) می‌باشد. روش مدل‌سازی چند سطحی خطی تعمیم یافته روش مدل‌سازی سنتی رگرسیون خطی است؛ با این تفاوت که در مدل‌سازی چند سطحی خطی نیازی به برقراری فرض استقلال مشاهدات جمع‌آوری شده نیست. برای تحلیل داده‌های طولی ارزشیابی اعضای هیأت علمی و تصحیح وابستگی بین

مشاهدات جمع‌آوری شده در این روش مدل‌سازی، از ضرایب رگرسیونی تصادفی (Random coefficient) در مدل استفاده می‌شود (۷). برای این که کمترین اریب در برآورد عوامل مدل تحلیل‌های چند سطحی وجود داشته باشد، لازم است به ساختار خوشه‌ای (Cluster) یا سلسله مراتبی (Hierarchical) مشاهدات توجه شود که در این‌جا نمرات ارزشیابی درون سطح اعضای هیأت علمی، اعضای هیأت علمی درون سطح گروه‌های آموزشی و گروه‌های آموزشی درون سطح دانشکده آشیانه شدند (۸) و ارتباطی که مشاهدات درون هر سطح با هم دارند، در تحلیل در نظر گرفته می‌شود (۹).

چنانچه محقق وابستگی مشاهدات طولی را در نظر نگیرد و برای تحلیل از آزمون‌های آماری استاندارد مانند رگرسیون خطی (که به شدت بر فرض استقلال مشاهدات تکیه دارد) استفاده کند، خطای معیار مربوط به ضرایب رگرسیونی نادرست تخمین زده می‌شود و به طور مستقیم بر روی معنی‌داری نتایج تأثیر می‌گذارد (۷).

در بررسی مجلات ایرانی همچون مجله آموزش پزشکی اصفهان و شهید بهشتی تهران مشاهده شد که در پنج سال اخیر هیچ مقاله‌ای در زمینه بررسی طولی یا مدل‌سازی چند سطحی داده‌های ارزشیابی اعضای هیأت علمی توسط دانشجو به چاپ نرسیده است، اما در مجلات آموزش پزشکی خارجی مانند Medical education و Medical teacher درباره استفاده از داده‌های طولی و مدل‌های چند سطحی مقالات فراوانی را می‌توان یافت.

ارزشیابی اعضای هیأت علمی در دانشگاه علوم پزشکی کرمان توسط دانشجو انجام می‌شود و هدف اصلی آن، ارزشیابی کیفیت تدریس اعضای هیأت علمی می‌باشد. این نوع ارزشیابی به صورت مدون طی هر نیم‌سال تحصیلی در این دانشگاه انجام و نتایج آن به صورت مقطعی و در هر نیم‌سال تحصیلی تحلیل و در اختیار اعضا قرار می‌گیرد.

در پژوهش حاضر سعی شد تا با تکیه بر اطلاعات طولی

نمرات ارزشیابی با استفاده از روش‌های استانداردسازی آماری به بازه بین ۱۰۰-۱ تغییر یافت و از این نمرات نهایی در تحلیل استفاده شد.

جهت بررسی تأثیر متغیرهایی مانند جنسیت، رتبه علمی استادان، سابقه کار و دانشکده بر چگونگی تغییرات ارزشیابی کیفیت تدریس استادان، این متغیرها به صورت متغیر مستقل در مدل وارد شد و مورد ارزیابی قرار گرفت. با توجه به این که متغیرهای رتبه علمی و دانشکده دارای چند حالت بود، رتبه علمی استادیار و دانشکده پزشکی به دلیل حجم نمونه بالا به عنوان گروه مبدأ در نظر گرفته شد.

قبل از شروع تحلیل‌های اصلی و به دلیل این که نمره ارزشیابی همه اعضا به طور کامل وجود نداشت، ابتدا با استفاده از روش الگوریتم EM (Expectation maximization) - که یک روش برآورد داده‌های گمشده بر مبنای روش ماکزیمم درست‌نمایی می‌باشد- مشاهدات گمشده برآورد و اطلاعات مربوط به استادان کامل گردید و سپس در تحلیل مورد استفاده قرار گرفت.

به طور مختصر منطق روش مدل‌سازی چند سطحی آن است که برای هر عضو هیأت علمی یک خط رگرسیون برآزش شود؛ بدین معنی که به ۳۳۶ خط نیاز بود. روش مدل‌سازی چند سطحی به جای برآورد تعداد بسیار زیاد عرض از مبدأ و شیب خط، اجازه می‌دهد اثر آن‌ها تصادفی باشد. به بیان دیگر، تنها یک عامل شامل واریانس عرض از مبدأها و یا شیب‌ها محاسبه و در مدل معنی‌داری آن بررسی می‌شود.

برای آزمون معنی‌داری عرض از مبدأ تصادفی در مدل چند سطحی در مقایسه با مدل رگرسیون خطی که عرض از مبدأها تصادفی نمی‌باشند، از آزمون نسبت درست‌نمایی (Likelihood ratio test یا LRT) استفاده شد (۶). در این آزمون مقدار حاصل از اختلاف لگاریتم درست‌نمایی دو مدل به دست آمده، در نهایت برای بررسی معنی‌داری با توزیع  $\chi^2$  مقایسه می‌گردد. همچنین برای بررسی معنی‌دار بودن ضرایب برآورد شده دو مدل، از مقدار احتمال (P) با سطح معنی‌داری

ارزشیابی اعضای هیأت علمی و روش تحلیل چند سطحی و رگرسیون خطی ساده، چگونگی تغییرات نمرات ارزشیابی اعضای هیأت علمی دانشگاه در طول زمان مورد ارزیابی قرار گیرد و نتایج حاصل از برآزش این دو مدل با هم مقایسه و تأثیر متغیرهایی مانند جنسیت، سابقه آموزشی، رتبه علمی و نوع دانشکده بر چگونگی روند نمرات ارزشیابی استادان در طول زمان سنجیده شود تا با استفاده از این نتایج، تأثیر فرایند ارزشیابی بر بهبود و ارتقای برنامه‌های آموزشی دانشگاه برای دست‌اندرکاران مشخص گردد.

## روش کار

در این پژوهش تحلیلی تمام اعضای هیأت علمی دانشگاه علوم پزشکی کرمان که بین سال‌های ۹۱-۱۳۸۷ در این دانشگاه مشغول به تدریس بودند، شرکت کردند. اعضای هیأت علمی در اواخر هر نیم‌سال تحصیلی از طریق پرسش‌نامه الکترونیکی که از طریق وب‌سایت پایش ارزشیابی اعضای هیأت علمی در اختیار دانشجویان قرار می‌گرفت، ارزشیابی می‌شدند. سؤالات این پرسش‌نامه و چند گزینه‌ای بودن آن بسته به نوع تئوری و عملی بودن دروس، متفاوت بود. بررسی مطالعات حاکی از آن است که روایی محتوا و پایایی سؤالات در حد مطلوب است (۱۰).

متغیر وابسته، نمره ارزشیابی بود که به هر عضو هیأت علمی در هر نیم‌سال تحصیلی تعلق می‌گرفت و نمرات از نیم‌سال دوم سال تحصیلی ۸۸-۱۳۸۷ تا نیم‌سال اول سال تحصیلی ۹۱-۱۳۹۰ به دلیل محرمانه بودن داده‌ها به صورت کدبندی شده و طی نامه‌نگاری با واحد ارزشیابی مرکز مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی کرمان (Education development center یا EDC) در اختیار محقق قرار گرفت. این متغیر از نوع کمی و به صورت میانگین نمراتی بود که هر استاد برای هر درس از فرم‌های ارزشیابی تکمیل شده توسط دانشجویان کسب می‌کرد و بازه آن بین ۱-۵ می‌باشد. برای داشتن نتایج ملموس‌تر و قابل مقایسه، بازه

۰/۰۵ استفاده شد.

Intra-class correlation coefficient) استفاده گردید (۱۱، ۶) که بالا بودن این شاخص نشان دهنده همبستگی شدید داده‌های درون هر خوشه یا گروه می‌باشد (۶). تمام فرایندها تحلیل و مدل‌سازی با استفاده از نرم‌افزار آماری R انجام گرفت.

### یافته‌ها

بر اساس یافته‌های به دست آمده، تعداد اعضای هیأت علمی مرد در حدود ۱/۵ برابر زنان بود (۲۰۴ در برابر ۱۳۲). دانشکده پزشکی بیشترین (۱۷۶ نفر) و دانشکده پیراپزشکی (۹ نفر) کمترین تعداد عضو هیأت علمی را داشت (جدول ۱).

برای بررسی نکویی برازش مدل‌های چند سطحی خطی و رگرسیون خطی بر روی مشاهدات پژوهش از سه معیار شاخص اطلاع Akaike information criterion) Akaike یا (AIC، شاخص اطلاع Bayesian Bayesian information criterion) یا (BIC و میانگین مربعات خطا (Mean squared error یا MSE) استفاده شد. کم بودن مقادیر این معیارها نشان دهنده بهتر بودن مدل برازش شده می‌باشد. علاوه بر این، برای تعیین میزان همبستگی مشاهداتی که متعلق به یک سطح یا خوشه هستند، از شاخص ضریب همبستگی درون خوشه‌ای (ICC) یا

جدول ۱: اطلاعات توصیفی اعضای هیأت علمی دانشگاه علوم پزشکی کرمان

متغیر	هیأت علمی [تعداد (درصد)]
جنسیت	مرد (۶۰/۷) ۲۰۴
	زن (۳۹/۳) ۱۳۲
مرتبه علمی	استاد (۸/۰) ۲۷
	استادیار (۴۵/۵) ۱۵۳
	دانشیار (۲۰/۸) ۷۰
	مربی (۱۲/۸) ۴۳
نام دانشکده	پزشکی (۵۲/۴) ۱۷۶
	بهداشت (۶/۵) ۲۲
	مدیریت (۴/۲) ۱۴
	پیراپزشکی (۲/۷) ۹
	پرستاری (۱۱/۰) ۳۷
	داروسازی (۵/۴) ۱۸
	دندان‌پزشکی (۱۷/۹) ۶۰

دانشکده مستقل است. شباهت اندکی بین نمرات اعضا در درون گروه مشاهده شد.

با مقایسه مدل رگرسیون خطی و مدل چند سطحی، مقدار LRT محاسبه شده برابر با ۲۲۷/۷۶ و به شدت معنی‌دار بود. مقدار MSE در مدل رگرسیون خطی ۱۹۲/۸ به دست آمد. پس از معنی‌دار شدن عرض از مبدأ تصادفی و در برآورد اثرات

مقدار ICC در سطح استادان، گروه‌های آموزشی و دانشکده‌ها به ترتیب ۰/۲۸، ۰/۱۰ و ۰/۰۱ محاسبه شد. با توجه به مقادیر محاسبه شده می‌توان گفت که شدت مشابهت نمرات ارزشیابی در سطح استادان ۲۸ درصد، در سطح گروه‌های آموزشی ۱۰ درصد و در سطح دانشکده‌ها تنها ۱ درصد می‌باشد و این به معنی مستقل بودن نمرات افراد در یک

ابتدا نتایج مدل رگرسیون خطی ارایه می‌گردد. نتایج این مدل حاکی از آن بود که متغیرهای زمان و سابقه آموزشی به طور معنی‌داری با نمره ارزشیابی استادان مرتبط بود (جدول ۲). نمره ارزشیابی استادان با گذشت هر نیم‌سال، حدود یک نمره افزایش داشته است؛ در حالی که با افزایش هر یک سال، سابقه نمره حدود ۰/۰۱ کاهش یافت. علاوه بر این، نمره استادان و مربیان تفاوت معنی‌داری با استادپاران نشان داد و به اندازه ۴ و ۲/۵ نمره بیشتر بود. در نهایت نمرات در دانشکده‌های پرستاری، داروسازی و دندان‌پزشکی نسبت به دانشکده پزشکی به ترتیب ۲/۵۸، ۳/۵۷ و ۵/۲ واحد بیشتر بود (جدول ۲).

تصادفی مربوط به مدل چند سطحی، مقدار واریانس برآورد شده در سطح استادان برابر با ۵۰/۴۵ و در بخش خطای مدل، ۱۴۲/۴۵ بود؛ به عبارت دیگر حدود ۲۶ درصد از تغییراتی که در مدل رگرسیون خطی قابل توجیه نبود، به دلیل تفاوت‌های بین استادان می‌باشد. همچنین کم بودن مقادیر AIC و BIC (۱۶۱۳۱/۲۸) و (۱۶۲۱۵/۴۲) در مدل چند سطحی در مقایسه با مدل رگرسیون خطی ( $AIC = 16359/04$ )،  $BIC = 16443/18$ )، تأییدی بر کارایی و توانایی بالای این مدل در برازش مشاهدات طولی ارزشیابی با کنترل سطح استادان است. با وجود برتری معنی‌دار مدل چند سطحی، در

جدول ۲: عوامل مؤثر بر نمره ارزشیابی استادان توسط دانشجویان با استفاده از دو مدل رگرسیون خطی و چند سطحی

متغیر	طبقات متغیر	مدل رگرسیون خطی			مدل چند سطحی		
		P	SE	B	P	SE	B
زمان		< ۰/۰۰۱	۰/۱۸	۱/۰۰۳	< ۰/۰۰۱	۰/۱۵	۰/۹۸
سابقه آموزش		۰/۰۰۶	۰/۰۴	-۰/۰۱۱	۰/۱۹۰	۰/۰۶	-۰/۰۸
جنسیت	مرد (۰ نفر)			گروه مبدأ			
	زن (۱ نفر)	۰/۹۷۰	۰/۶۹	۰/۰۲	۰/۹۳۰	۱/۰۵	۰/۰۷
	استادیار (۰ نفر)			گروه مبدأ			
مرتبه علمی	استاد (۱ نفر)	< ۰/۰۰۱	۰/۹۰	۴/۱۲	۰/۰۰۵	۱/۳۸	۳/۹۰
	دانشیار (۳ نفر)	۰/۰۹۰	۰/۸۵	۱/۴۳	۰/۳۲۰	۱/۲۹	۱/۲۷
	مربی (۴ نفر)	۰/۰۳۰	۱/۱۵	۲/۳۷	۰/۲۱۰	۱/۷۵	۲/۱۶
	پزشکی (۰ نفر)			گروه مبدأ			
	بهداشت (۱ نفر)	۰/۵۸۰	۱/۳۰	۰/۷۰	۰/۷۴۰	۱/۹۸	۰/۶۴
	مدیریت (۲ نفر)	۰/۲۱۰	۱/۶۰	۱/۹۹	۰/۴۰۰	۲/۴۳	۲/۰۲
دانشکده	پیراپزشکی (۳ نفر)	۰/۸۷۰	۲/۰۲	-۰/۳۰	۰/۹۰۰	۳/۰۷	-۰/۳۳
	پرستاری (۴ نفر)	۰/۰۳۰	۱/۲۱	۲/۵۸	۰/۱۶۰	۱/۸۴	۲/۷۵
	داروسازی (۵ نفر)	۰/۰۱۰	۱/۴۲	۳/۵۷	۰/۰۸۰	۲/۱۶	۳/۷۱
	دندان‌پزشکی (۶ نفر)	< ۰/۰۰۱	۰/۸۹	۵/۳۰	۰/۰۰۰۱	۱/۳۶	۵/۵۸

SE: Standard error

است، مقدار خطای معیار در مدل چند سطحی بیشتر از مدل رگرسیون خطی می‌باشد؛ در حالی که در متغیر زمان این مقدار در مدل رگرسیونی ۰/۰۳ بیشتر از مدل چند سطحی بود.

در مقایسه خطای معیار متغیرهای دو مدل مشخص شد که در متغیرهایی مانند جنسیت، درجه علمی و دانشکده‌ها که مستقل از زمان می‌باشند و سابقه آموزشی که وابسته به زمان

در بررسی تأثیر جنسیت بر چگونگی تغییرات نمرات ارزشیابی ثبت شده، گروه زنان نسبت به مردان در مدل رگرسیون خطی  $0/02$  و مدل چند سطحی  $0/07$  بیشتر بود که این متغیر در هیچ کدام از مدل‌های برازش شده معنی‌دار نبود (جدول ۲). نتایج به دست آمده از برازش مدل چند سطحی نشان داد که تنها متغیرهای معنی‌دار مدل شامل «زمان، مرتبه علمی استاد نسبت به استادیار و دانشکده دندان‌پزشکی نسبت به پزشکی» بودند. در این مدل با هر واحد تغییر در متغیر زمان، به میزان  $0/98$  بهبودی در نمره ارزشیابی استادان مشاهده شد. همچنین تغییرات نمره ارزشیابی مرتبه علمی استاد در مقایسه با مبدأ مدل یعنی استادیار،  $3/9$  واحد و این تغییرات در مقایسه دانشکده دندان‌پزشکی با دانشکده پزشکی،  $5/58$  واحد بیشتر بود (جدول ۲).

### بحث و نتیجه‌گیری

با توجه به اهمیت ارزشیابی کیفیت تدریس اعضای هیأت علمی در دانشگاه‌ها و مراکز آموزشی و توسعه برنامه‌های ارزشیابی، در مطالعه حاضر چگونگی تغییرات نمرات ارزشیابی اعضای هیأت علمی دانشگاه علوم پزشکی کرمان با در نظر گرفتن متغیرهای جنسیت، سابقه آموزشی، مرتبه علمی و دانشکده با استفاده از روش مدل‌سازی چند سطحی مورد بررسی قرار گرفت و نتایج حاصل با مدل‌سازی رگرسیون خطی مقایسه شد.

نتایج حاصل از تحلیل چند سطحی مشاهدات حاکی از آن است که روند نمرات ارزشیابی اعضای هیأت علمی دانشگاه علوم پزشکی کرمان توسط دانشجو در طول شش ترم مورد بررسی در پژوهش یک روند صعودی دارد. نتایج حاصل با نتایج پژوهش‌های فتاحی و همکاران (۱۲)، لسانی (۱۳) و Maker و همکاران (۱۴) همسو بود. در پژوهش آنان نیز نمرات ارزشیابی مورد بررسی روندی افزایشی داشت. همچنین یافته‌های مطالعه با نتایج مطالعه Cohan و همکاران که به بررسی تغییرات عملکرد آموزشی اعضای هیأت علمی مرکز

پزشکی دانشگاه میشیگان با مقایسه نمرات ارزشیابی اعضا توسط دانشجویان برای دو سال متوالی پرداخت (۱۵)، همسو می‌باشد. میانگین نمرات ارزشیابی اعضای هیأت علمی در مطالعه Cohan و همکاران نیز یک روند صعودی داشت (۱۵)، اما با نتایج مطالعه توصیفی سلمان‌زاده و همکاران در دانشگاه علوم پزشکی ایران (۱۶)، مطابقت ندارد. با وجود این که نتایج ارزشیابی تحقیق سلمان‌زاده و همکاران به اطلاع اعضای هیأت علمی رسیده بود، اما تغییر قابل توجهی در عملکرد آنان مشاهده نشد (۱۶).

آنچه از بررسی و مقایسه روند نمرات ارزشیابی کیفیت تدریس اعضای هیأت علمی در پژوهش حاضر و سایر پژوهش‌های ارائه شده به دست آمد، این بود که در بیشتر مطالعات روند رو به رشدی در نمرات ارزشیابی در طول زمان وجود داشت که این خود حاکی از مؤثر بودن فرایند ارزشیابی کیفیت تدریس اعضای هیأت علمی بود (۱۲).

بررسی معنی‌داری متغیرهای مستقل در دو مدل نشان داد که تعداد متغیرهای مستقل معنی‌دار در برازش مدل رگرسیون خطی بیشتر از مدل چند سطحی خطی می‌باشد و این خود به دلیل عدم تصحیح خطای استاندارد و برآورد نادرست آن در مدل رگرسیون خطی است که باعث معنی‌داری غلط نتایج می‌شود. با مقایسه ضرایب و خطای معیار استاندارد مشخص شد که خطای معیار متغیرهای دانشکده محل کار و رتبه علمی در مدل رگرسیون خطی کمتر از مدل چند سطحی بود. همین امر دلیل معنی‌دار شدن تفاوت نمرات تعداد بیشتری از دانشکده‌ها و مرتبه‌های علمی شده بود (جدول ۲). به عبارت دیگر، خطاهای معیار محاسبه شده در رگرسیون خطی معتبر نیست و این امر به طور مستقیم معنی‌داری متغیرها را تحت تأثیر قرار می‌دهد.

نتایج حاصل از مقایسه این دو مدل‌سازی با نتایج حاصل از مطالعه Moerbeek و همکاران (۱۷) مطابقت داشت. آنان به مقایسه مدل‌های چند سطحی و رگرسیون خطی مطالعات مداخله‌ای پرداختند و عدم تصحیح SE در مدل رگرسیون

مقطعی جوادی و همکاران در بررسی نمرات ارزشیابی استادان دانشکده‌های مختلف گزارش کرد که امتیاز استادان در دانشکده مدارک پزشکی نسبت به دو دانشکده کتابداری و مدیریت خدمات بهداشتی-درمانی بیشتر از همه می‌باشد، با وجود این که تعداد دانشجویان و معدل آنان در دانشکده مدیریت از دو دانشکده دیگر بیشتر بود (۲۲). با توجه به نتایج پژوهش جوادی و همکاران (۲۲) و پژوهش حاضر که در آن تغییرات نمرات دانشکده دندان‌پزشکی نسبت به سایر دانشکده‌ها بیشتر است، باید دید که در بررسی متغیر دانشکده محل کار چه موضوعات تأثیرگذار دیگری دخیل هستند که آن‌ها نیز در مدل وارد و بررسی شوند تا نتایج هماهنگی به دست آید.

یکی از نقاط ضعف مطالعه حاضر، طولانی نبودن فرایند ارزشیابی توسط دانشجویان بود.

مطالعه‌ای در دانشگاه تگزاس آمریکا نشان داد که ارزشیابی از استاد باید ۷-۵ سال انجام شود تا بتوان از آن در جهت ارتقای اعضای هیأت علمی استفاده نمود (۲). در دانشگاه علوم پزشکی کرمان نمرات ارزشیابی ثبت شده استادان از سال ۱۳۸۶ در اختیار محقق قرار گرفت. هرچند در چند سال اخیر این روش ارزشیابی با اهتمام و جدیت بیشتر و با بهره‌گیری از نظرات همه دانشجویان انجام می‌شود، اما اطلاعات ارزشیابی برای همه استادان به صورت کامل ثبت نشده است که از جمله دلایل آن می‌توان به دستی بودن فرایند ارزشیابی در سال‌های آغازین و مشارکت تعدادی از دانشجویان به جای همه دانشجویان در این فرایند اشاره کرد.

به علت فراوانی بسیار زیاد مشاهدات گمشده در نمره ارزشیابی استادان در سال‌های اول ثبت نمرات ارزشیابی، پژوهش تنها بر روی اطلاعات ثبت شده در سه سال اخیر انجام گردید.

از جنبه‌های قابل توجه مطالعه حاضر، بهره بردن از روش‌های پیشرفته آماری برای تحلیل نمرات ارزشیابی استاد توسط دانشجو می‌باشد. این امر برای جامعه دانشگاهی که علاقمند به بهبود کیفیت تدریس استادان و توانمندسازی

خطی مشاهده شد (۱۷). همچنین نتایج پژوهش حاضر در مقایسه برازش دو مدل، با یافته‌های تحقیقات Hox (۳)، Twisk (۵)، Snijders TABAB و Bosker (۹) و Goldstein (۱۱) در زمینه برازش مدل چند سطحی بر داده‌های طولی همسو می‌باشد.

بررسی مطالعات نشان می‌دهد که اتفاق نظری درباره موضوع تأثیر جنسیت اعضای هیأت علمی بر نمره ارزشیابی وجود ندارد (۱۸، ۱۴). نتایج حاصل از بررسی معنی‌داری اثر جنسیت در پژوهش حاضر مشخص کرد که این متغیر در هر دو مدل آماری برازش شده معنی‌دار نیست که این نتیجه با نتایج پژوهش شکورنیا و همکاران (۱۹) مغایرت دارد. آنان در پژوهش خود به بررسی مقطعی ارتباط بین نمره ارزشیابی استاد با معدل تحصیلی دانشجویان پرداختند که تأثیر جنسیت بر چگونگی نمره ارزشیابی استادان توسط دانشجویان معنی‌دار بود و دانشجویان به استادان مرد نمره ارزشیابی بالاتری داده بودند و این نتایج با نتایج مطالعات فتاحی و همکاران (۱۲) و لسانی (۱۳) همسو می‌باشد. در مطالعه آنان نیز مشخص شد که بین نمرات ارزشیابی کیفیت تدریس استادان زن و مرد تفاوت معنی‌داری وجود ندارد (۱۳، ۱۲).

بررسی پژوهش علی‌اصغرپور و همکاران نشان داد که موضوعاتی همچون درجه علمی و سابقه کار از نظر مدرسین و دانشجویان جزء موضوعات پر اهمیت و تأثیرگذار بر ارزشیابی دانشجویان از استاد می‌باشد (۲۰). بر اساس نتایج پژوهش حاضر، تغییرات نمره ارزشیابی مرتبه علمی استاد نسبت به سایر مرتبه‌های مورد بررسی در پژوهش بیشتر بود، اما سابقه آموزشی تنها متغیر معنی‌دار در مدل رگرسیون خطی مشاهده شد و در مدل چند سطحی معنی‌دار نبود. نتیجه حاصل از بررسی سابقه تدریس با نتیجه مطالعه Marsh (۲۱) همسو می‌باشد. در مطالعه او مشخص شد که نمرات ارزشیابی با افزایش تجربه مدرس تمایل به کاهش دارد (۲۱).

در بررسی متغیر دانشکده و تأثیر آن بر نمره ارزشیابی استادان مطالعات اندکی (۲۲، ۱۷) وجود دارد، اما نتایج مطالعه

## نتیجه‌گیری

نتایج پژوهش نشان داد که نمره ارزشیابی استادان دانشگاه علوم پزشکی کرمان توسط دانشجو در طی زمان بهبود می‌یابد. همچنین در مقایسه برآزش دو مدل رگرسیون خطی و چند سطحی بر داده‌های پژوهش، مدل چند سطحی خطی عملکرد بهتری دارد.

اعضای هیأت علمی خویش هستند، این امکان را فراهم می‌کند تا از برنامه‌های پیشرفته‌تر و نوین‌تری برای ارزشیابی استادان استفاده کنند. علاوه بر این، چنین مطالعه‌ای در سایر دانشگاه‌های علوم پزشکی و مؤسسات آموزشی وجود ندارد. بنابراین پیشنهاد می‌گردد مطالعات مشابهی در سایر دانشگاه‌ها و در صورت امکان با دوره طولانی‌تر انجام شود تا امکان مقایسه نتایج مهیا باشد.

## References:

1. 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.
2. 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]
3. Raissifar A, Khaghani zadeh M, Ebadi A, Masomi M. A review of current methods for the evaluation of faculties in universities, with emphasis on evaluation of faculties by students. *Sci J Educ Strategies Med Sci*. 2008;1(1):10-8. [In Persian]
4. Goe L, Bell C, Little O. Approaches to evaluating teacher effectiveness: A research synthesis. Washington, DC: National Comprehensive Center for Teacher Quality; 2008.
5. Twisk JWR. Applied longitudinal data analysis for epidemiology: a practical guide. New York: Cambridge University Press; 2003.
6. Twisk JWR. Applied multilevel analysis: a practical guide for medical researchers. New York: Cambridge University Press; 2006.
7. Hox JJ. Applied multilevel analysis. Amsterdam: TT-publikaties; 1995.
8. Amir Kafi M. The importance and logic of multilevel models in social research. *Iran J Sociol*. 2007;7(4): 38-71. [In Persian]
9. Snijders TAB, Bosker RJ. Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: SAGE Publications Limited; 2011.
10. Admami A, Nakhaee N, Fasihi T, Fatahi Z. Preliminary assessment of the validity and reliability of the evaluation questionnaires by the students regarding teaching methods of the faculty members of Kerman University of Medical Sciences in 2002-2003. *J Strides Dev Med Educ*. 2004;1(2):121-9. [In Persian]
11. Goldstein H. Multilevel statistical models. London: Arnold Publishers; 1999.
12. Fatahi Z, Musapur N, Haghdoost AA. The trend of alterations in the quality of educational performance in faculty members of Kerman University of Medical Sciences. *J Strides Dev Med Educ*. 2006;2(2):63-71. [In Persian]
13. Lesani M. Study of trend the quality of academic staff teaching mathematics and Computer Bahonar University of Kerman. Proceedings of the 5th Conference on "Evaluation of the quality of the education system; 2012 Sep 6-10; Tehran, Iran. 2012. [In Persian]
14. Maker VK, Lewis MJ, Donnelly MB. Ongoing faculty evaluations: developmental gain or just more pain? *Curr Surg*. 2006;63(1):80-4.
15. Cohan RH, Dunnick NR, Blane CE, Fitzgerald JT. Improvement of faculty teaching performance: efficacy of resident evaluations. *Acad Radiol*. 1996;3(1):63-7.
16. Salmanzadeh H, Ghadamiyan M, Maleki M. Evaluation, the Path to the Quality Improvement. *J Health Administration*. 2001;4(10):41-50. [In Persian]
17. Moerbeek M, van Breukelen GJ, Berger MP. A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. *J Clin Epidemiol*. 2003;56(4):341-50.



18. Karamdoost NA. School of Psychology and Educational Sciences examined the relationship between students evaluation of the professorship teaching them a lesson in the mean scores years of education. *J Psychol Educ.* 2004;34(1):57-76. [In Persian]
19. Shakurnia A, Malayeri AR, Torabpour M, Elhampour H. correlation between student evaluation of teaching and students' grades. *Iran J Med Educ.* 2006; 6(1): 51-8. [In Persian]
20. Aliasgharpoor M, Monjamed Z, Bohrani N. Factors affecting students' evaluation of teachers: comparing viewpoints of teachers and students. *Iran J Med Educ.* 2010;10(2):186-95. [In Persian]
21. Marsh HW. Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *J Educ Psychol.* 2007;99(4):775.
22. Javadi M, Yaghoobi M, Tavakolli M, Afkar A. A study on relation between teachers evaluation score & students mean score of special courses in Isfahan college of management & informatic sciences(original article). *Hospital.* 2009;8(2): 45-50. [In Persian]

## Multilevel Modeling of Longitudinal Data on Academic Evaluation of the Academic Staff of Kerman University of Medical Sciences, Iran

Somayeh Shahroudi<sup>1</sup>, Aliakbar Haghdoost<sup>2</sup>, Farzaneh Zolala<sup>3</sup>, Maryam Okhovati<sup>4</sup>,  
Mohammadreza Baneshi<sup>5\*</sup>

1. M.Sc. in Biostatistics, Student Research Committee, Kerman University of Medical Sciences, Kerman, Iran

2. Ph.D. in Epidemiology, Professor, Regional Center of Care HIV, World Health Organization Collaborating Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

3. Ph.D. in Epidemiology, Assistant Professor, Regional Center of Care HIV, World Health Organization Collaborating Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

4. Ph.D. in Library and Information Sciences, Assistant Professor, Physiology Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman, Iran

5. Ph.D. in Biostatistics, Associate Professor, Research Center for Modeling in Health, Institute of Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

• Received: 18 Apr, 2014

• Received Corrected Version: 6 Sep, 2014

• Accepted: 7 Sep, 2014

**Background & Objective:** In order to promote quality of education, teaching performances of all academic staff are evaluated by students every semester. This study aims to address the changing trend of staff scores in different semesters, and factors affecting it, in Kerman University of Medical Sciences, Iran, using multilevel modeling. Moreover, the results were compared to that of simple linear regression modeling.

**Methods:** In the present analytical study, the scores of 336 academic staff of Kerman University of Medical Sciences during 2008 to 2012 were extracted from the students' monitoring website. The tutor's code was used for the identification of their scores in different semesters and was entered into the multilevel models. In order to investigate the effects of gender, work experience (in years), school (seven schools), and academic rank (master, assistant professor, associate professor, professor), the simple regression and multilevel models were compared. To study the significance of the random intercepts, the likelihood ratio test was used. In addition, to study the goodness of fit of the models, the Akaike information criteria (AIC), Bayesian information criteria (BIC), and the mean squared error (MSE) were used.

**Results:** The results showed that time had a significant positive impact on the improvement of staff scores. However, the scores of male and female staff were not significantly different. The scores of professors were significantly higher than assistant professors. In addition, the score of dental school staff was significantly higher than that of medical school staff. The comparison of the goodness of fit of models showed that the multilevel modeling provided a better fit to the longitudinal data. In the linear regression model, variables such as work experience and academic rank (professors in comparison to assistant professors and nursing school staff in comparison to pharmacology school staff) were falsely considered significant, due to the lack of consideration of the dependence of longitudinal observations of the evaluation and correction of standard errors.

**Conclusion:** The multilevel model, due to the consideration of the dependence of longitudinal observations of the evaluation, provides a better fit to data. Moreover, the incorrect use of the linear regression model, considering longitudinal observations to be independent, led to erroneous conclusions. The results of the present study, in terms of the goodness of fit of the models, showed a positive trend in scores of academic staff of Kerman University of Medical Sciences. In other words, the scores have improved over time.

**Key Words:** Teacher evaluation, Student, Added value Evaluation, Longitudinal data, Multilevel modeling

\*Correspondence: Modeling in Health Research Centre, Institute for Futures Studies in Health, Medical University, beginning seven Gardens Blvd., Kerman, Iran

• Tel: (+98) 34 3132 5403

• Fax: (+98) 34 3132 5403

• Email: M\_baneshi@kmu.ac.ir