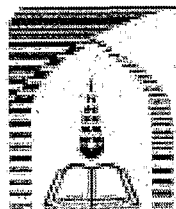


بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Signature

112073

۸۷/۱/۸۵۵۹  
۸۸/۱/۱۴



دانشگاه تربیت مدرس  
دانشکده علوم پزشکی

رساله دوره دکتری در رشته آمار زیستی

عنوان:

مدل رگرسیون کلاس پنهان چند سطحی با متغیرهای پیش بین پنهان  
و کاربرد آن در پزشکی و اپیدمیولوژی

نگارش:

آوات فیضی

استاد راهنما:

دکتر غلامرضا بابایی

استاد مشاور:

دکتر انوشیروان کاظم نژاد

کتابخانه تخصصی  
پزشکی

۱۳۸۸ / ۱ / ۱۸

بهمن ۱۳۸۷

۱۱۲۰۶۳



بسمه تعالی

تاییدیه اعضای هیات داوران حاضر در جلسه دفاع از رساله دکتری

آقای آوات فیضی رشته آمار زیستی رساله دکتری خود را با عنوان: "مدلهای رگرسیون کلاس پنهان چند سطحی با متغیرهای پیش بین پنهان و کاربرد آن در پزشکی و اپیدمیولوژی در تاریخ ۸۷/۱۱/۲۱ ارائه کردند. اعضای هیات داوران نسخه نهایی این رساله را از نظر فرم و محتوا تایید کرده است و پذیرش آنرا برای تکمیل درجه دکتری پیشنهاد می کنند.

امضاء	نام و نام خانوادگی	اعضای هیات داوران
	دکتر غلام رضا بابایی	۱- استاد راهنما
	دکتر انوشیروان کاظم نژاد	۲- استاد مشاور
-----	-----	۳- استاد مشاور
	دکتر ابراهیم حاجی زاده	۴- استاد ناظر
	دکتر علیرضا حیدر نیا	۵- استاد ناظر
	دکتر ید... محرابی	۶- استاد ناظر
	دکتر عباس گرامی	۷- استاد ناظر
	دکتر سقراط فقیه زاده	۸- نماینده تحصیلات تکمیلی

آئین نامه پایان نامه (رساله) های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله) های تحصیلی دانشجویان دانشگاه تربیت مدرس مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است. بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله) ی خود، مراتب را قبلا به طور کتبی به دفتر "دفتر نشر آثار علمی" دانشگاه اطلاع دهد.

ماده ۲: در صفحه سوم کتاب (پس از برگ شناسنامه) عبارت ذیل را چاپ کند:  
"کتاب حاضر، حاصل پایان نامه کارشناسی ارشد / رساله دکتری نگارنده در رشته ایمان‌ریزی...  
است که در سال ۱۳۸۷... در دانشکده علوم پزشکی دانشگاه تربیت مدرس به راهنمایی  
دکتر محترم آیت الله العظمی بروجردی، مشاوره دکتر آیت الله العظمی بروجردی / کتابم از آن دفاع شده است."

ماده ۳: به منظور جبران بخشی از هزینه های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به "دفتر نشر آثار علمی" دانشگاه اهداء کند. دانشگاه می تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.

ماده ۴: در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده را به عنوان خسارت به دانشگاه تربیت مدرس، تادیه کند.

ماده ۵: دانشجو تعهد و قبول می کند در صورت خودداری از پرداختهای بهای خسارت، دانشگاه مذکور را از طریق مراجع قضایی مطالبه و وصول کند، به علاوه به دانشگاه حق می دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶: اینجانب کوبان فیضی دانشجوی رشته ایمان‌ریزی ..... مقطع دکتری .....  
تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی

تاریخ و امضا  
۸۷/۱۱/۲۱

## آیین نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی دانشگاه

### تربیت مدرس

مقدمه: با عنایت به سیاست‌های پژوهشی و فناوری دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیأت علمی، دانشجویان، دانش‌آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان‌نامه، رساله و طرحهای تحقیقاتی با هماهنگی دانشگاه انجام شده است، موارد زیر را رعایت نمایند:

ماده ۱- حق نشر و تکثیر پایان‌نامه/ رساله و درآمدهای حاصل از آنها متعلق به دانشگاه می باشد ولی حقوق معنوی پدید آورندگان محفوظ خواهد بود.

ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان‌نامه/ رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و با تایید استاد راهنمای اصلی، یکی از اساتید راهنما، مشاور و یا دانشجوی مسئول مکاتبات مقاله باشد. ولی مسئولیت علمی مقاله مستخرج از پایان‌نامه و رساله به عهده اساتید راهنما و دانشجو می باشد.

تبصره: در مقالاتی که پس از دانش‌آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان‌نامه/ رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

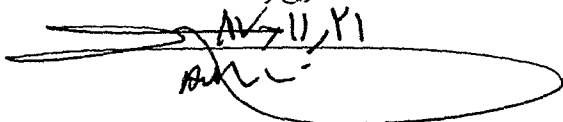
ماده ۳- انتشار کتاب و یا نرم افزار و یا آثار ویژه حاصل از نتایج پایان‌نامه/ رساله و تمامی طرحهای تحقیقاتی کلیه واحدهای دانشگاه اعم از دانشکده ها، مراکز تحقیقاتی، پژوهشکده ها، پارک علم و فناوری و دیگر واحدها باید با مجوز کتبی صادره از معاونت پژوهشی دانشگاه و براساس آئین نامه های مصوب انجام شود.

ماده ۴- ثبت اختراع و تدوین دانش فنی و یا ارائه یافته ها در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان‌نامه/ رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق معاونت پژوهشی دانشگاه انجام گیرد.

ماده ۵- این آیین‌نامه در ۵ ماده و یک تبصره در تاریخ ۸۷/۴/۱ در شورای پژوهشی و در تاریخ ۸۷/۴/۲۳ در هیأت رئیسه دانشگاه به تایید رسید و در جلسه مورخ ۸۷/۷/۱۵ شورای دانشگاه به تصویب رسیده و از تاریخ تصویب در شورای دانشگاه لازم‌الاجرا است.

نام و نام خانوادگی : آوات فیضی

تاریخ و امضاء

۸۷/۱۱/۲۱  


# تقدیم به روح پاک مادرہم

(رحمت و مغفرت الہی قرین روحشان باد - انشاء...)

9

به پاس زحمات پدرہم

## تقدیر و تشکر

سپاسگزار الطاف بی‌کران و بی‌منت خداوند بزرگ هستیم که توفیق تحصیل دانش را به من عطا فرمود.

مدیون ابدی لحظه لحظه مهربانی‌های خالصانه و دلگرمی‌های مادر عزیزم (رحمت الهی قرین روحشان باد) و سپاسگزار زحمات بیدریغ پدر بزرگوارم هستیم.

از حُسن نیت و مساعدت‌های همیشگی اساتید محترم گروه آمار زیستی جناب آقایان: دکتر کاظم‌نژاد، دکتر حاجی-زاده، دکتر فقیه زاده و دکتر بابایی در طول دوره تحصیلم در دانشگاه تربیت مدرس کمال تشکر و قدردانی را دارم.

از پروفسور Vermunt (بخش آمار و روش‌شناسی دانشگاه Tilburg هلند) که امکان استفاده از نرم‌افزار LatentGold برای اجرای مدل‌های گسترش داده شده در رساله را در طول دوره فرصت مطالعاتی اینجانب فراهم نمود تشکر می‌کنم.

از پروفسور Bockenholt (دانشگاه McGill-کانادا) و پروفسور Wall و Jia Guo (دانشگاه Minnesota-آمریکا) به واسطه راهنمایی‌های مؤثر و کارگشایشان متشکرم. از پروفسور Rabe-Hesketh (دانشگاه Berkeley-آمریکا) به خاطر پیشنهاداتشان و از پروفسور Steels (دانشگاه Bristol - انگلیس) به خاطر در اختیار قرار دادن داده‌های ارزشمند Bangladesh fertility survey تشکر می‌کنم.

از حُسن دقت و پیشنهادات آقایان دکتر گرامی و دکتر محرابی که زحمت داوری رساله را عهده دار بودند تشکر می‌کنم.

خوشوقت بودم از طی نمودن دوره تحصیلی در کنار همکلاسی‌های خوب و فهیم و آشنایی با دیگر دوستان دوره-های دکتری آمار زیستی در دانشگاه تربیت مدرس.

از همکاری کلیه کارکنان دانشکده علوم پزشکی بویژه گروه آمار زیستی، پژوهش و آموزش تشکر می‌کنم.

از مهربانی‌های دو خواهر عزیزم متشکرم.

در پایان متشکرم از صبوری، همدلی و همراهی همسرم.

## چکیده

در بسیاری از حوزه‌های تحقیقاتی از جمله پزشکی، اپیدمیولوژی، علوم رفتاری، روانشناسی و علوم اجتماعی مفاهیمی وجود دارند که مستقیم قابل مشاهده یا اندازه‌گیری نیستند (همانند ناتوانی، افسردگی، و...) بلکه بر مبنای یک سری از متغیرها و نشانه‌های مشهود مرتبط مورد ارزیابی و سنجش قرار می‌گیرند؛ مفاهیم و متغیرهای یاد شده در ادبیات آماری موسوم هستند به متغیرهای پنهان و نقش مهمی در مدل‌های نظری حوزه‌های مختلف علمی ایفاء می‌کنند. ارزیابی و اندازه‌گیری دقیق این متغیرها همواره مورد توجه بوده است و در این راستا مدل‌های متغیرهای پنهان بعنوان رهیافتی در مطالعه و ارزیابی جامع این مفاهیم مطرح گردیده‌اند. از طرفی اهمیت و گستردگی داده‌ها و ساختارهای چندسطحی موجب شکل‌گیری و گسترش مدل‌هایی شده است که مطالعه و مدل‌بندی متغیرهای پنهان را در ساختار مدل‌های آمیخته خطی تعمیم یافته امکان پذیر می‌نماید. مدل‌سازی در یک ساختار تلفیقی از داده‌های چندسطحی و متغیرهای پنهان در قالب "مدل‌های تعمیم‌یافته متغیرهای پنهان" صورت می‌گیرد. در پژوهش حاضر، در حوزه مدل‌های تعمیم‌یافته متغیرهای پنهان، مدل عمومی جدیدی تحت عنوان: "مدل رگرسیون کلاس پنهان چندسطحی با متغیرهای پیش‌بین پنهان" در سه حالت (بسته به تعلق متغیرهای پیش‌بین به سطوح مختلف) بسط داده شده است. مدل‌های جدید ارتباط متغیر(های) پاسخ پنهان گسسته را با متغیرهای پیش‌بین پنهان و در حضور اثر دیگر متغیرهای کمکی مشاهده شده در یک ساختار چند سطحی مدل‌سازی می‌کنند. مدل‌های یاد شده از یک طرف تعمیمی هستند از مدل رگرسیون لوژستیک با اثرات آمیخته بنحوی که هم متغیر پاسخ پنهان است و هم متغیرهای پنهان در جزء پیش‌بینی‌کننده‌ی خطی مدل وارد شده‌اند و از طرفی تعمیمی هستند از مدل‌های رگرسیون کلاس پنهان در وضعیت چند سطحی و با حضور متغیرهای پیش‌بین پنهان. وجود ساختار چندسطحی و حضور تعداد زیاد متغیرهای پنهان (اثرات تصادفی و عامل‌های پیش‌بین پنهان) در مدل‌های جدید موجب پیچیده‌گی و حجیم شدن چشم‌گیر محاسبات مربوط به برآوردیابی پارامترها در مقایسه با مدل‌های موجود در حوزه مدل‌های تعمیم‌یافته‌ی متغیرهای پنهان می‌گردد؛ از اینرو فرم‌های تعمیم‌یافته‌ای از الگوریتم‌های Upward-Downward EM و MCEM متناسب با شرایط نظری مدل‌های یاد شده جهت برآوردیابی مورد استفاده گرفت. پیوسته یا گسسته در نظر گرفتن توزیع اثرات تصادفی در مدل‌های چندسطحی متغیرهای پنهان منجر به دو رهیافت خاص برآوردیابی در این مدل‌ها می‌شود که به رهیافت‌های پارامتری و ناپارامتری موسوم هستند؛ نحوه‌ی برآوردیابی از طریق هر دو روش بصورت نظری بیان شده است و عملکرد رهیافت ناپارامتری بعنوان ابزاری برای ارزیابی اعتبار مدل‌های برازش شده در حالت پارامتری تشریح گردیده است.

مدل‌های جدید جهت تحلیل دو مجموعه داده‌ی چند سطحی در حوزه بهداشت روانی و علوم اجتماعی مورد استفاده قرار گرفتند. با توجه به تفاوت ماهیت داده‌های مورد تحلیل به لحاظ حجم خوشه‌ها، در قالب یک مطالعه شبیه‌سازی، در حجم‌های نمونه‌ای مختلف، اثر حجم خوشه‌ها و میزان اعتبار عامل‌های پیش‌بین پنهان بر میزان اریبی پارامترهای برآورد شده مورد مطالعه قرار گرفت. نتایج مطالعه شبیه‌سازی حاکی از آن بود که سرعت همگرایی برآوردها به مقادیر واقعی در ساختارهای دارای خوشه‌های حجیم بیشتر است. برآورد پارامترها و اثرات تصادفی و نیز مطالعه شبیه‌سازی در نرم افزار LatentGold syntax module انجام و الگوریتم MCEM در نرم افزار SAS IML برنامه نویسی شد.

**واژه‌ها و عبارات کلیدی:** مدل (رگرسیون) کلاس پنهان، ساختار چندسطحی، اثر تصادفی، رگرسیون کلاس پنهان چندسطحی، عامل‌های پیش‌بین پنهان، الگوریتم Upward-Downward EM، الگوریتم MCEM



## فهرست مطالب

صفحه	عنوان
<b>فصل اول: کلیات</b>	
۱	۱-۱- مقدمه .....
۲	۲-۱- مقدمه و کلیاتی در مورد مدل‌های متغیرهای پنهان و حوزه کاربری آنها .....
۹	۳-۱- کاربرد مدل‌های (رگرسیون) کلاس پنهان و .....
۱۱	۴-۱- موضوع و اهداف رساله .....
۱۵	۵-۱- محتوی فصول رساله .....
<b>فصل دوم: مرور مطالعات پیشین</b>	
۱۶	۱-۲- مقدمه .....
۱۷	۲-۲- مدل‌های متغیرهای پنهان .....
۱۸	۳-۲- مدل‌های کلاس پنهان معمولی .....
۲۱	۴-۲- مدل‌های کلاس پنهان همراه با کواریت یا مدل‌های رگرسیون کلاس پنهان .....
۲۳	۵-۲- مدل‌های عاملی .....
۲۴	۶-۲- مدل‌های رگرسیون چند سطحی .....
۲۷	۷-۲- مدل‌های چندسطحی متغیرهای پنهان .....
۲۹	۱-۷-۲- برآوردیابی در مدل‌های چند سطحی متغیرهای پنهان .....
<b>فصل سوم: مبانی نظری و کاربردها</b>	
۳۳	۱-۳- مقدمه .....
۳۷	۲-۳- رگرسیون لوژستیک با اثرات آمیخته .....
۳۸	۳-۳- مدل اول: مدل رگرسیون کلاس پنهان چندسطحی با متغیرهای پیش‌بین پنهان مربوط به آزمودنی‌های سطح یک .....
۳۹	۱-۳-۳- توزیع توأم پاسخ‌های مشاهده شده $X_{ij}$ و $Y_{ij}$ .....
۴۰	۲-۳-۳- جزء پیش‌بین خطی مدل .....
۴۰	۳-۳-۳- توزیع متغیرهای پنهان (فاکتورها و کلاسها) .....
۴۰	۱-۳-۳-۳- توزیع عضویت در کلاسها .....
۴۱	۲-۳-۳-۳- توزیع پاسخ به متغیرهای پاسخ مشاهده شده در هر کلاس .....
۴۱	۳-۳-۳-۳- ارتباط بین متغیرهای مشاهده شده $X_{ij}$ با عامل‌های پنهان .....

فصل سوم: مبانی نظری و کاربردها (ادامه)

صفحه

۴۲	..... فرض‌های توزیعی مدل جدید
۴۲	..... فرض استقلال موضعی
۴۲	..... توزیع عوامل پنهان
۴۴	..... توزیع توأم عاملهای پنهان و اثرات تصادفی
۴۳	..... بحث شناسایی پذیری مدلها
۴۳	..... شناسایی پذیری اجزاء مدل
۴۴	..... شناسایی پذیری کلی مدل
۴۴	..... برآوردیابی
۴۵	..... برآورد پارامترها
۴۷	..... MCEM
۵۰	..... Upward-Downward EM روش
۵۲	..... خطای استاندارد برآوردها
۵۳	..... استنباط در مورد اثرات تصادفی
۷-۳	..... مدل دوم: مدل رگرسیون کلاس پنهان چند سطحی با متغیرهای پیش بین پنهان مربوط به واحدهای
۵۳	..... سطح دوم
۵۵	..... جزء پیش‌بینی کننده‌ی خطی مدل
۵۵	..... توزیع توأم پاسخ‌های مشاهده شده
۵۵	..... تابع درستنمایی مدل دوم
۵۶	..... توزیع توأم متغیرهای پنهان (عوامل پنهان و اثرات تصادفی)
۸-۳	..... مدل سوم: رگرسیون کلاس پنهان چندسطحی با متغیرهای پیش‌بین پنهان مربوط به واحدهای سطح اول و
۵۷	..... دوم
۵۸	..... جزء پیش‌بینی کننده خطی مدل
۵۸	..... توزیع توأم پاسخهای مشاهده شده
۵۹	..... تابع درستنمایی مدل سوم
۶۰	..... توزیع توأم متغیرهای پنهان
۹-۳	..... رهیافت ناپارامتری برآوردیابی در مدل رگرسیون کلاس پنهان چند سطحی با متغیرهای پیش‌بین
۶۰	..... پنهان

## فصل چهارم: کاربرد مدلها و مطالعه شبیه‌سازی

۶۳	..... ۱-۴ مقدمه
۶۳	..... ۲-۴ مثال اول
۷۰	..... ۳-۴ مثال دوم
۷۵	..... ۴-۴ مثال سوم
۷۸	..... ۵-۴ مثال چهارم
۸۰	..... ۶-۴ نتایج مطالعه شبیه‌سازی

## فصل پنجم: بحث و نتیجه‌گیری

۸۳	..... ۱-۵ مقدمه
۸۴	..... ۲-۵ بحث، نتیجه‌گیری و پیشنهادات
۹۲	..... منابع و مأخذ
۱۰۲	..... فصل ضمایم
۱۰۳	..... ضمیمه (الف): پرسشنامه پرخاشگری
۱۰۵	..... ضمیمه (ب): بخشی از سؤالات پرسشنامه "خانواده اصلی" برای ارزیابی احساس امنیت فرزندان در خانواده
۱۰۶	..... ضمیمه (ج): پرسشنامه شیوه‌ی تربیتی والدین
۱۰۷	..... ضمیمه (د): برنامه مربوط به اجرای روش MCEM در SAS IML
۱۱۱	..... ضمیمه (ه): برنامه مربوط به پیاده‌سازی مدل اول در نرم افزار LatentGold Syntax module
۱۱۲	..... ضمیمه (و): برنامه مربوط به پیاده‌سازی مدل دوم در نرم افزار LatentGold Syntax module
۱۱۳	..... ضمیمه (ز): برنامه مربوط به پیاده‌سازی مدل دوم در نرم افزار LatentGold Syntax module
۱۱۴	..... ضمیمه (ح): برنامه مربوط به پیاده‌سازی مدل اول در حالت ناپارامتری در نرم افزار LatentGold Syntax module
۱۱۴	..... ضمیمه (ط): نمونه‌ای از برنامه شبیه‌سازی در نرم افزار LatentGold Syntax module مربوط به ارزیابی
۱۱۵	..... اثر حجم خوشه‌ها بر کیفیت برآورد پارامترها
۱۱۶	..... ضمیمه (ی) اطلاعات مربوط به داده‌های مثال دوم (Bangladesh fertility survey)

## فهرست جداول

صفحه

۳	جدول (۱-۱) دسته بندی کلی مدل‌های متغیرهای پنهان بر پایه مقیاس متغیرهای پنهان و مشاهده شده
۸	جدول (۲-۱) تعداد مدل‌های ممکن در یک ساختار دو سطحی بر حسب ترکیب‌های متغیرهای پنهان در سطوح اول و دوم
۶۷	جدول (۱-۴) : مدل‌های برازش یافته و ملاک‌های ارزیابی متناظر با آنها
۶۸	جدول (۲-۴) احتمالات شرطی پاسخ‌ها (الگوی بروز ابعاد مختلف پرخاشگری در کلاس‌ها) و حجم کلاس‌ها
۶۹	جدول (۳-۴) اثرات برآورد شده عامل پنهان و متغیرهای کمکی مشاهده شده در کلاس‌ها
۷۰	جدول (۴-۴) نتایج مربوط به تفاوت اثر متغیرهای پیش بین در کلاس‌ها با استفاده از آزمون Wald
۷۳	جدول (۴-۵) مدل‌های مختلف برازش شده و ملاک‌های ارزیابی متناظر با آنها
۷۴	جدول (۴-۶) احتمالات شرطی پاسخ‌ها در کلاس‌ها ، احتمالات حاشیه‌ای و برآورد اثرات عامل پنهان و متغیرهای کمکی مشاهده شده
۷۵	جدول (۴-۷) نتایج مربوط به تفاوت اثر متغیرهای پیش بین در کلاس‌ها با استفاده از آزمون Wald
۷۶	جدول (۴-۸) مدل‌های برازش شده و ملاک‌های ارزیابی متناظر با آنها
۷۷	جدول (۴-۹) اثرات برآورد شده عامل پنهان و متغیرهای کمکی مشاهده شده در کلاس‌ها
۷۹	جدول (۴-۱۰) مدل‌های مختلف برازش یافته، تعداد پارامترها و ملاک‌های برازندگی متناظر آنها
۸۰	جدول (۴-۱۱) حجم کلاس‌های سطح یک و احتمالات شرطی پاسخ‌ها در این کلاس‌ها، حجم کلاس‌های سطح دو و $P(c_{ij} = c   c_j = m)$
۸۱	جدول (۴-۱۲) نتایج مطالعه شبیه سازی برای ارزیابی میزان اریبی برآورد پارامترها وقتی حجم خوشه‌ها کم است.
۸۲	جدول (۴-۱۳) نتایج مطالعه شبیه سازی برای ارزیابی میزان اریبی برآورد پارامترها وقتی حجم خوشه‌ها زیاد است.

## فهرست نمودارها

### صفحه

- نمودار (۱-۳) مدل رگرسیون کلاس پنهان دو سطحی با متغیرهای پیش بین پنهان مربوط به آزمودنی های سطح یک ۳۸
- نمودار (۲-۳) مدل رگرسیون کلاس پنهان دو سطحی با متغیرهای پیش بین پنهان مربوط به آزمودنی های سطح دو ۵۴
- نمودار (۳-۳) مدل رگرسیون کلاس پنهان دو سطحی با متغیرهای پیش بین پنهان مربوط به آزمودنی های سطح یک و دو ۵۷

# فصل اول

## کلیات

### ۱-۱- مقدمه

در این فصل مفهوم متغیر پنهان ، انواع و دسته‌بندی مدل‌های متغیرهای پنهان توضیح داده شده است. مدل‌های متغیر پنهان که با مدل‌های تعمیم یافته در پژوهش حاضر ارتباط مستقیم دارند مورد تأکید قرار گرفته‌اند. چارچوب کلی مدل‌های چند سطحی متغیرهای پنهان و جایگاه مدل‌های جدیدی که در این رساله بسط داده شده‌اند در قالب ساختار کلی یاد شده مورد اشاره قرار گرفته است. کاربردهای مدل‌های کلاس پنهان، رگرسیون کلاس پنهان معمولی و چند سطحی بر اساس تحقیقات مختلفی که از این مدل‌ها استفاده کرده‌اند به اختصار مورد اشاره قرار گرفته است. اهداف رساله در قالب تبیین اهمیت نظری و کاربردی مدل‌های بسط یافته توضیح داده شده است. شیوه‌ی برآوردیابی پارامترها و داده‌های مورد استفاده جهت پیاده‌سازی عملی مدل‌ها به اختصار مورد اشاره قرار گرفته‌اند.

## ۱-۲- مقدمه و کلیاتی در مورد مدل‌های متغیرهای پنهان و حوزه کاربری آنها

## متغیر پنهان و مدل‌های متغیرهای پنهان

مفاهیم زیادی در حوزه تحقیقات علوم اجتماعی، رفتاری، پزشکی، بهداشت و اپیدمیولوژی وجود دارند که مستقیماً قابل اندازه‌گیری و مشاهده نیستند؛ به عنوان مثال می‌توان گرایشها، تنش‌های رفتاری (همانند استرس، افسردگی ...)، ویژگی‌های شخصیتی، ناتوانی، کیفیت زندگی و... را نام برد. چنین مفاهیمی را متغیرهای پنهان نامیده و این مفاهیم صرفاً می‌توانند به صورت غیر مستقیم از طریق متغیرهای و مشخصه‌های قابل مشاهده و اندازه‌گیری همانند سؤالات پرسشنامه، مشخصه‌ها و علائم بیماری و غیره ارزیابی شوند. تکنیک‌های مقیاس‌سازی<sup>۱</sup> متعددی به منظور سنجش و استخراج اطلاعات مربوط به مفاهیم پنهان از متغیرهای مشاهده شده بسط داده شده‌اند که مهمترین خانواده از این تکنیک‌ها عبارت است از مدل‌های متغیرهای پنهان<sup>۲</sup>. یک مدل متغیر پنهان (امکاناً غیر خطی) یک تحلیل مسیر یا مدل گرافیکی است که علاوه بر متغیرهای مشاهده شده، متغیرهای پنهان که معرف مفاهیم پنهان هستند را شامل می‌شود. در مدل‌های متغیرهای پنهان دو فرض اساسی مکانیسم علی مبتنی بر پاسخها را بیان میکنند: (۱) فرض می‌شود که (چگونگی) پاسخ دهی به متغیرهای مشاهده شده نتیجه‌ای از موقعیت آزمودنی بر حسب متغیر(های) پنهان است، (۲) متغیرهای مشاهده شده بعد از کنترل و استخراج متغیرهای پنهان چیز مشترکی برای بیان کردن ندارند؛ این فرض معروف است به استقلال موضعی<sup>۳</sup>. دو فرض دیگر در این مدلها معطوف به توزیع متغیرهای مشاهده شده و پنهان می‌باشند بطوریکه بر مبنای این فروض نوع مدل متغیرهای پنهان مشخص می‌شود. Bartholomew (۱۹۸۴) دسته بندی کلی زیر را برای مدل‌های متغیرهای پنهان ارائه کرده است [۱]:

---

۱ Scaling techniques

۲ Latent variables models

۳ Local independence assumption

جدول (۱-۱) دسته بندی کلی مدل‌های متغیرهای پنهان بر پایه مقیاس متغیرهای پنهان و مشاهده شده

متغیر پنهان		متغیر مشاهده شده
گسسته	پیوسته	
تحلیل نیم رخ پنهان (latent profile analysis-LPA)	تحلیل عاملی (Factor analysis-FA)	پیوسته
تحلیل کلاس پنهان (Latent class analysis-LCA)	تحلیل ویژگی پنهان (Latent trait analysis-LTA)	گسسته

در تحلیل عاملی<sup>۴</sup> و تحلیل ویژگی پنهان<sup>۵</sup>، متغیرهای پنهان بصورت پیوسته و نرمال توزیع شده‌اند لیکن در تحلیل نیم رخ پنهان<sup>۶</sup> و تحلیل کلاس پنهان<sup>۷</sup> متغیر پنهان گسسته بوده و بنابراین فرض می‌شود که از یک توزیع چند جمله‌ای تبعیت می‌کند. متغیرهای مشاهده شده در مدل‌های کلاسک FA و LPA پیوسته بوده و در غالب موارد توزیع شرطی آنها به شرط متغیرهای پنهان فرض می‌شود که نرمال هستند. در LCA و LTA متغیرهای مشاهده شده اسمی هستند و در نتیجه فرض می‌شود توزیع شرطی آنها دو جمله‌ای یا چند جمله‌ای می‌باشد. در حوزه‌ی مدل‌های متغیرهای پنهان تمایز قائل شدن در مقیاس متغیرهای مشاهده شده چندان حائز اهمیت نیست. تصریح توزیع‌های شرطی این متغیرها بطور طبیعی تابع نوع مقیاس آنها می‌باشد بنحویکه مدل‌هایی که اخیراً در این حوزه بسط یافته‌اند امکان مدل‌بندی همزمان متغیرهای مشاهده شده‌ی دارای توزیع‌های مختلف در یک ساختار واحد را فراهم می‌آورند. برای مثال این توزیعها در مورد متغیرهای پیوسته می‌تواند شامل نرمال، استیودنت، لگ-نرمال یا گاما و...، دو جمله‌ای برای متغیرهای دو حالتی، چندجمله‌ای برای متغیرهای رتبه‌ای و

۴ FA

۵ LTA

۶ LPA

۷ LCA



اسمی، و پواسن، دوجمله‌ای یا دوجمله‌ای منفی برای متغیرهای شمارشی. بسته به پیوسته یا گسسته (بر حسب کاربرد و مبانی نظری) در نظر گرفتن متغیر پنهان فرمهای تعمیم یافته LCA یا LTA بدست می‌آید.

### مدلهای کلاس پنهان

در مدل‌های کلاس پنهان (LC) بعنوان یکی از دسته‌های اصلی مدل‌های متغیرهای پنهان فرض می‌شود یک متغیر پنهان C رسته‌ای (کلاس) وجود دارد که پاسخ آزمودنی‌ها به متغیرهای مشاهده شده یا به عبارتی قرار گرفتن آنها در رسته‌های مختلف متغیر پنهان یادشده را متأثر می‌نماید و هر آزمودنی فرض می‌شود (با احتمال معین) فقط متعلق به یکی از این کلاس‌ها می‌باشد. معمولاً به لحاظ تکنیکی واژه تحلیل "کلاس پنهان" معطوف به حالتی است که متغیرهای مشاهده شده گسسته‌اند و واژه مدل‌های آمیزه متناهی<sup>۸</sup> پوشش دهنده دامنه وسیعتری از متغیرهای مشاهده شده با مقیاسهای مختلف می‌باشد. تحلیل کلاس پنهان به لحاظ نظری در مدل‌بندی رگرسیونی با چند متغیر پاسخ گسسته [۳ و ۲]، انواع مقیاس‌سازی<sup>۹</sup> [۵، ۴]، خوشه‌بندی [۷، ۶ و ۸] و مدل‌بندی ناپارامتری اثرات تصادفی<sup>۱۰</sup> [۱۱ و ۱۰، ۹] کاربرد دارد. تحلیل کلاس پنهان به لحاظ ماهیتی در دسته‌بندی آزمودنی‌ها همانند روشهای چند متغیره‌ی خوشه بندی، مقیاس سازی چند بعدی<sup>۱۱</sup> و تحلیل تناظر<sup>۱۲</sup> می‌باشد. لیکن تفاوت اصلی روشهای یاد شده با تحلیل کلاس پنهان آنست که روش اخیر مبتنی بر مدل است که امکان انجام آزمونهای آماری به منظور ارزیابی عملکرد مدل‌های مختلف برآزش شده را فراهم می‌کند اما دیگر روشها صرفاً بر مبنای ملاکهای تشابه (یا عدم تشابه) به صورت ریاضیاتی عمل می‌نمایند.

### مدل رگرسیون کلاس پنهان

مدل رگرسیون کلاس پنهان تعمیمی از مدل کلاس پنهان است بگونه‌ای که احتمال عضویت آزمودنی‌ها در

۸ Finite mixture models

۹ Scaling

۱۰ Nonparametric random effects modeling

۱۱ Multidimensional scaling

۱۲ Correspondence analysis

کلاسها به صورت تابعی از ویژگیهای آنها (متغیرهای همراه<sup>۱۲</sup> یا کمکی<sup>۱۴</sup>) مدلبندی می‌شود و تابع یاد شده از طریق نسبت بختها<sup>۱۵</sup> توصیف می‌شود (تفاوت تکنینی این دو نوع مدل در فصل دوم توضیح داده شده است).

### مدلهای عاملی

برداری از متغیرهای پنهان یا عاملهای مشترک پیوسته به صورت غیر مستقیم بر مبنای برداری از متغیرهای مشاهده شده‌ی پیوسته استخراج می‌شود. این عوامل مشترک می‌توانند مبین همان مفاهیم فرضی باشند که در تعریف متغیر پنهان به آنها اشاره شد. دو دسته اصلی مدل‌های عامل (عوامل) مشترک عبارتند از: تحلیل عاملی اکتشافی<sup>۱۶</sup> و تحلیل عاملی تأییدی<sup>۱۷</sup>. در تحلیل عاملی اکتشافی تعدادی از عوامل مشترک شناسایی می‌شود، برآوردیابی انجام و شناسایی پذیری مدل با اعمال کمترین قیدها صورت می‌گیرد. لیکن در تحلیل عاملی تأییدی اعمال قیدها به منظور استخراج تعداد مشخصی عامل بر مبنای نظریه‌های پایه‌ای یا طرحهای تحقیق صورت می‌گیرد. مقیاس عاملها یا از طریق استاندارد سازی عوامل<sup>۱۸</sup> که به مفهوم برابر یک قرار دادن واریانس عاملها است صورت می‌گیرد یا از طریق برابر قرار دادن یکی از بارهای عاملی برای هر یک از عاملها با عدد یک انجام می‌شود که به آن Factor anchoring گویند. استخراج عوامل مشترک پیوسته از متغیرهای مشاهده شده‌ی اسمی و رتبه‌ای از طریق مدل‌های Item-response theory و تحلیل ویژگی پنهان صورت می‌گیرد. چهار رهیافت پایه‌ای برای برآوردیابی پارامترها در این مدلها استفاده می‌شود که عبارتند از: روش مبتنی بر نرمال چند متغیره<sup>۱۹</sup> رهیافت مبتنی بر نرمال دو متغیره<sup>۲۰</sup> رهیافت مدل

<sup>۱۲</sup> Concomitant variables

<sup>۱۴</sup> Covariate

<sup>۱۵</sup> Odds Ratio

<sup>۱۶</sup> Exploratory factor analysis

<sup>۱۷</sup> Confirmatory factor analysis

<sup>۱۸</sup> Factor standardization

<sup>۱۹</sup> Underlying Multivariate Normal Approach

<sup>۲۰</sup> Underlying Bivariate Normal Approach

بخت‌های متناسب<sup>۲۱</sup> و رهیافت مبتنی بر توزیع تجمعی نرمال<sup>۲۲</sup>

### ساختارهای چند سطحی و مدل‌های رگرسیون چند سطحی

داده‌های چندسطحی<sup>۲۳</sup> اطلاعات بدست آمده از واحدهای آزمایشی یا آزمودنی‌هایی است که درون واحدهای آزمایشی سطوح بالاتر واقع شده‌اند. بعنوان مثال می‌توان مریضه‌های واقع در مراکز درمانی، مریضه‌های ارزیابی شده به وسیله پزشکان، بچه‌های واقع در خانواده‌ها، چشم چپ و راست و اندازه‌های مکرر را جزء ساختارهای چند سطحی در نظر گرفت. واحدهای اولیه آزمایشی را واحدهای سطح اول و خوشه‌ها یا واحدهای دربردارنده واحدهای اولیه را واحدهای سطح دوم و چنانچه واحدهای سطح دوم خود درون واحدهای سطح بالاتر قرار گرفته باشند ساختار حاصل یک ساختار سه سطحی خواهد بود. پاسخ‌های واحدهای واقع در خوشه‌ها تحت تأثیر ویژگی‌های خوشه‌ی مربوط هستند بطوریکه پاسخ آزمودنی‌های واقع در یک خوشه مشابه تر از پاسخ آزمودنی‌های واقع در خوشه‌های مختلف است لیکن به علت پاره‌ای از محدودیتها نمی‌توان انتظار داشت که کلیه متغیرهای کمکی و توضیحی سطح خوشه را در تحلیل دخالت داده شود که این امر موجب ایجاد غیر همگنی در سطح خوشه‌ها شده و این غیر همگنی (مشترک) موجب همبستگی (پاسخ) واحدهای درون خوشه‌ها حتی بعد از کنترل اثر متغیرهای کمکی دخیل شده می‌گردد. مدل‌های اثرات تصادفی (یا مدل‌های رگرسیون چند سطحی) [۱۴، ۱۳، ۱۲] همبستگی یاد شده یا غیر همگنی بین خوشه‌ها را از طریق واردن کردن ضرایب تصادفی رگرسیونی مدل‌بندی بندی می‌کنند. این ضرایب تصادفی شامل عرض از مبدأ و شیب‌ها هستند. عرض از مبدأ تصادفی بیانگر ناهمگنی غیرمشاهده شده در مقادیر کلی پاسخ می‌باشد در حالیکه شیب‌های تصادفی مبین ناهمگنی غیرمشاهده شده ناشی از تأثیر متغیرهای توضیحی بر متغیر(های) پاسخ می‌باشد.

### مدل‌های چند سطحی متغیرهای پنهان

<sup>۲۱</sup> The proportional odds model

<sup>۲۲</sup> The Normal Ogive Approach

<sup>۲۳</sup> Multilevel data

(۲۰۰۴) Skronidal & Rabe-Hesketh چارچوبی کلی تحت عنوان "مدلهای تعمیم یافته متغیرهای پنهان"<sup>۲۴</sup> را ارائه کردند که سه دسته کلی از مدلها را شامل می‌شود: "مدلهای اثرات تصادفی و عاملی"، "مدلهای دارای متغیرهای پنهان پیوسته و گسسته" و "مدلهای چند سطحی دارای متغیرهای پنهان واقع در سطوح مختلف"<sup>[۱۰]</sup>. عناصر یک مدل متغیر پنهان چند سطحی عبارتند از:

- مجموعه‌ای از پاسخ‌های مشاهده شده ( $Y$ ) که می‌توانند دو حالتی، اسمی، رتبه‌ای، پیوسته و یا ترکیبی از آنها باشد.
- مجموعه‌ای از متغیرهای پنهان ( $U$ ) که می‌توانند گسسته (اسمی یا رتبه‌ای)، پیوسته و یا ترکیبی از آنها باشد.
- مجموعه‌ای از متغیرهای پیش‌بینی‌کننده یا مستقل ( $Z, W$ ) و
- مشاهدات واقع در  $L$  سطح مختلف

ساختار رگرسیونی مدل متغیرهای پنهان چند سطحی به صورت زیر است:

$$g[E(Y_k)] = W_k^{(1)}\beta + Z_k^{(1)}\Lambda^{(1)}v_k \quad (1-1)$$

$$h[E(v_k^{(l)})] = W_k^{(l)}\beta^{(l)} + Z_k^{(l)}\Lambda^{(l)}v_k^{(l+)} \quad \text{برای } l=1,2,\dots,L \quad (2-1)$$

اندیس  $k$  مربوط به مشاهده‌های بدست آمده از واحدهای واقع در بالاترین سطح ساختار مورد مطالعه می‌باشد.  $h$  و  $g$  توابع پیوند (همانی، لوجیت، لگاریتم و...) هستند که وابسته به مقیاس متغیرهای پاسخ و پنهان واقع در سمت چپ رابطه‌های فوق هستند.  $v_k$  بیانگر متغیر پنهان مربوط به مشاهده  $k$  (در کلیه سطوح) است در حالیکه  $v_k^{(l)}$  و  $v_k^{(l+)}$  عبارتند از متغیرهای پنهان مربوط به آزمودنی  $k$  در سطوح  $l$  و بالاتر از  $l$ ، پارامترهای آزاد در این مدل شامل:  $\beta$  و  $\Lambda$  و  $\gamma$  و نیز (ک) واریانس (بین) متغیرهای پنهان و (بین) متغیرهای وابسته‌اند. باید توجه شود که ترکیبهای مختلف کمیت  $W_k^{(l)}\Lambda^{(l)}$  تعیین‌کننده مدل‌های اثرات تصادفی و مدل‌های عاملی خواهد بود. در واقع

<sup>۲۴</sup> Generalized latent variables models