دانشگاه تربیت مدرس

دانشکده علوم ریاضی

پایان نامه دوره کارشناسی ارشد آمار

روش برآورد بیزی در مدل‌های چندسطحی

توسط

عاطفه فرخی

استاد راهنما

دکتر موسی گل‌علیزاده

استاد مشاور

دکتر سید محمدابراهیم حسینی‌نسب

دی ۱۳۸۹

قدردانی

سپاس و ستایش معبود یگانه را که پرتو الطاف بی‌شمارش بر لحظه لحظهٔ زندگی‌ام ساطع و آشکار است. حمد و ثنا می‌گزارم او را که فکرت و اندیشه را در بستر روحم روان ساخت و بهره‌گیری از خوان گستردهٔ دانش اساتیدم را نصیب و روزی‌ام گردانید.

امتنان و سپاس می‌گزارم تلاشها، زحمات و راهنمایی‌های ظریف، ارزشمند و بی‌شائبهٔ استاد فرزانه و گرانمایه‌ام، جناب دکتر موسی گل‌علیزاده را که با حمیت و جدیت، مرا به دقت، اندیشه، درک و تعمق وامی‌داشتند.

عاطفه فرخی

دی ۱۳۸۹

تقدیم به بهترین واژگان حیات

پدر و مادر مهربانم

چکیده

یکی از فرض‌های اساسی در مدل‌های رگرسیون خطی ساده استقلال آماری بین مشاهدات است. گاهی اوقات این فرض برای موضوع مورد مطالعه صادق نیست و در نتیجه بکارگیری مدل‌های متداول رگرسیونی ممکن است مناسب نباشد. این حالت به‌ویژه برای داده‌هایی که دارای ساختار همبستگی درون گروهی بوده و به داده‌های چندسطحی یا خوشه‌ای معروف می‌باشند اتفاق می‌افتد. مدل مناسب برای تحلیل این‌گونه داده‌ها مدل‌های چندسطحی است. در مقایسه با برآورد پارامترها در مدل رگرسیون خطی ساده، مدل‌های چندسطحی با در نظر گرفتن همبستگی میان مشاهدات نتایج دقیق‌تری ارائه می‌دهند. روش‌های متفاوتی برای برآورد پارامترهای مدل‌های چندسطحی وجود دارد. توجه این پایان‌نامه بر روی روش‌های بسامدی و بیزی می‌باشد. به‌ویژه جهت بکارگیری روش بیزی از تعمیم الگوریتم مونت کارلوی زنجیر مارکوفی استفاده می‌شود که قالبی بسیار ساده داشته و باعث حذف همبستگی بین پارامترهای ثابت و خطای تصادفی منتسب به سطوح بالای مدل می‌شود. با این‌حال، افزایش بعد ماتریس واریانس—کواریانس بردار خطا در این حالت از کارایی آن می‌کاهد. لذا در این پایان‌نامه جهت بهبود سرعت همگرایی این روش دو راه‌کار پیشنهاد شده است که پایه آنها بر مبنای تجزیه چولسکی ماتریس واریانس—کواریانس است. عملکرد روش‌های پیشنهادی در مطالعه شبیه‌سازی و یک مثال کاربردی مورد ارزیابی قرار گرفته است.

واژه‌های کلیدی : مدل‌های چندسطحی، برآورد پارامترها، ماتریس کواریانس، روش ساختاری مونت کارلوی زنجیر مارکوفی، تجزیه چولسکی.

فهرست مندرجات

۱ آشنایی با تحلیل آماری چندسطحی ۱

۱.۱ مقدمه ۱

۲.۱ انگیزه استفاده از تحلیل آماری چندسطحی ۴

۳.۱ مفاهیم پایه‌ای تحلیل آماری چندسطحی ۸

۴.۱ بررسی تعدادی از مدل‌های آماری چندسطحی ۱۳

۱.۴.۱ مدل عرض از مبدا تصادفی ۱۴

۲.۴.۱ مدل شیب تصادفی ۱۵

۳.۴.۱ روش‌های مقایسه‌ی مدل‌های آماری ۱۹

۵.۱ تعمیم مدل‌های آماری چندسطحی ۲۱
۱.۵.۱ مدل چندسطحی لوژستیک ۲۳
۲.۵.۱ مدل چندسطحی پواسون ۲۵

۲ روش‌های بسامدی برآورد پارامترها در مدل‌های چندسطحی ۲۷

۱.۲ مقدمه ۲۷

۲.۲ روش کمترین توان‌های دوم ۲۹
۱.۲.۲ روش کمترین توان‌های دوم تعمیم یافته بازگشتی ۳۰
۲.۲.۲ روش کمترین توان‌های دوم تعمیم یافته بازگشتی محدود شده . . . ۳۳

۳.۲ روش ماکسیمم درستنمایی ۳۴
۱.۳.۲ روش ماکسیمم درستنمایی محدود شده ۳۵
۲.۳.۲ روش ماکسیمم درستنمایی شبیه‌سازی شده ۳۷

۳ روش بیزی برآورد پارامترها در مدل‌های چندسطحی ۳۹

۱.۳ مقدمه ۳۹

۲.۳ الگوریتم مونت کارلوی زنجیر مارکوف در مدل‌های چندسطحی ۴۱

۱.۲.۳ نمونه‌گیری گیبس در مدل‌های عرض از مبدا تصادفی ۴۳

۲.۲.۳ نمونه‌گیری گیبس در مدل‌های شیب تصادفی ۴۵

۳.۳ بهبود الگوریتم مونت کارلوی زنجیر مارکوف در مدل‌های چندسطحی ۴۷

۱.۳.۳ روش متعامدسازی ۴۷

۲.۳.۳ روش بسط پارامتر ۴۹

۳.۳.۳ مرکزی کردن سلسله مراتبی ۵۱

۴.۳.۳ روش ساختاری مونت کارلوی زنجیر مارکوف ۵۲

۴.۳ کاربرد روش $SMCMC$ در برخی مدل‌های چندسطحی ۵۴

۱.۴.۳ الگوریتم $SMCMC$ برای مدل‌های عرض از مبدا تصادفی ۵۶

۲.۴.۳ الگوریتم $SMCMC$ برای مدل‌های شیب تصادفی ۵۷

۴ بهبود روش ساختاری مونت کارلوی زنجیر مارکوفی ۶۰

۱.۴ مقدمه ۶۰

۲.۴ بهبود روش ساختاری مونت کارلوی زنجیر مارکوف ۶۲

۱.۲.۴ بهبود $SMCMC$ در مدل عرض از مبدا تصادفی ۶۳

۲.۲.۴ بهبود $SMCMC$ در تعمیم مدل عرض از مبدا تصادفی ۶۴
۳.۲.۴ بهبود $SMCMC$ در مدل شیب تصادفی ۶۶

۳.۴ تجزیه چولسکی در مدل‌های چندسطحی ۶۸
۱.۳.۴ تجزیه چولسکی در مدل عرض از مبدا تصادفی ۶۹
۲.۳.۴ تجزیه چولسکی در تعمیم مدل عرض از مبدا تصادفی ۷۰
۳.۳.۴ تجزیه چولسکی در مدل شیب تصادفی ۷۱

۵ مطالعات شبیه‌سازی و مثال‌های کاربردی ۷۴

۱.۵ مقدمه ۷۴

۲.۵ مثال کاربردی ۷۶
۱.۲.۵ برآورد بسامدی پارامترها ۷۸
۲.۲.۵ برآورد بیزی پارامترها ۸۱

۳.۵ مطالعه شبیه‌سازی ۸۹

لیست اشکال

۱.۳.۱ بررسی کلسترول خون بیماران تحت درمان ۴ پزشک. نمودار نشان می‌دهد سطح اول بیماران تحت نظر پزشکان متفاوت می‌باشند که در سطح دوم یعنی پزشکان آشیانه کرده‌اند و سطح دوم در فاکتور دیگری که بیمارستان است آشیانه کرده است. . . . ۹

۲.۳.۱ ضریب همبستگی درون گروهی. واضح است که هرگاه واریانس میان گروه‌ها بیشتر باشد ضریب همبستگی درون گروهی کمتر است و بالعکس. شکل (a) بیانگر استقلال می‌باشد. در شکل (b) هر پزشک بیماران متفاوتی را مورد معالجه قرار می‌دهد. اما تفاوتی بین پزشکان مشاهده نمی‌شود. این تفاوت در شکل (c) بیشتر و در شکل (d) بیشترین است. همبستگی درون‌گروهی در شکل (d) از تمامی شکل‌ها بیشتر است چون‌که بیماران در هر گروه با هم شباهت دارند اما بین گروه‌ها تفاوت زیادی ملاحظه می‌گردد. ۱۲

۳.۴.۱ رابطه رگرسیونی بین سن و میزان کلسترول خون بیماران تحت نظر ۱۲ پزشک مختلف. مدل در نظر گرفته شده مدل عرض از مبدا وشیب تصادفی است. عرض از مبدا و شیب های متفاوت در هر دسته نشانگر مدل رگرسیونی متفاوت برای هر گروه از پزشکان می‌باشد. ۱۸

فصل ۱

آشنایی با تحلیل آماری چندسطحی

۱.۱ مقدمه

مدل‌های رگرسیونی از پرکاربردترین مدل‌ها در تجزیه و تحلیل الگوهای خطی می‌باشند. معمولا بررسی آماری روابط بین متغیرهای مستقل و وابسته که به صورت ترکیبات مختلف از مشاهدات کمی و کیفی باشند در قالب رگرسیون، تحلیل واریانس و تحلیل کواریانس صورت می‌گیرد. یکی از فرض‌های اساسی در کاربرد اینگونه مدل‌ها استقلال آماری بین مشاهدات است. گاهی اوقات این فرض برای موضوع مورد مطالعه صادق نمی‌باشد و در نتیجه بکارگیری مدل‌های مرسوم رگرسیونی دارای اشکال خواهد بود. بعنوان مثال پینهریو و بیتس (۲۰۰۰) نشان دادند که نادیده گرفتن فرض همبستگی بین مشاهدات منجر به کم‌برآوردی خطای معیار برآوردگرهای ضرایب رگرسیونی می‌شود.

مثال‌های بیشماری در حوزه علوم کاربردی شامل علوم اجتماعی، علوم پزشکی، جامعه شناسی و ... وجود دارند که طبیعت همبستگی بین مشاهدات را بطور واضح نشان می‌دهند. به‌عنوان مثال در بررسی نمرات دانش‌آموزان از تعدادی مدارس که بطور تصادفی انتخاب شده‌اند، واضح است که

شباهت نسبی زیادی بین نمرات آن دسته از دانش‌آموزانی که از یک مدرسه خاص هستند وجود دارد (گلداستین، ۱۹۹۵). در مطالعه رابطه بین درآمد افراد، که در نواحی مختلفی از شهر سکونت دارند و میزان رضایت‌مندی آنها از یک سیستم خدماتی اجتماعی، باید اذعان داشت که نوع دیدگاه افراد ساکن در منطقه خاص تا حدودی به همدیگر شباهت دارد.

مدل مناسب برای تحلیل داده‌هایی مانند دو مثال فوق ، مدل چندسطحی است. البته از این مدل با نام‌های معادل دیگری نیز در مقالات و کتب آماری نام برده می‌شود. از آن جمله می‌توان به مدل‌های سلسله مراتبی، مدل‌های اثرهای تصادفی (گلمن و هیل، ۲۰۰۷)، مدل‌های پانلی، مدل‌های آشیانه‌ای و مدل‌های خوشه‌ای اشاره کرد.

ماسون و همکاران (۱۹۸۳) اولین کسانی بودند که مفاهیم و روش‌های مناسبی را برای تحلیل داده‌های چندسطحی گسترش داد. بریک و رادن‌باش (۱۹۹۲) و گلداستین (۱۹۸۶) مدل چندسطحی را برای داده‌های خطی مطرح کردند. استیراتلی و همکاران (۱۹۸۴) و وانگ و ماسون (۱۹۸۵) تحلیل چندسطحی را در مورد داده‌هایی که نتیجه آنها به صورت شکست یا موفقیت باشد بکار بردند.

پارامترهای مدل چندسطحی با استفاده از روش‌های بسامدی و بیزی برآورد می‌شوند. در چارچوب بسامدی پارامترها با استفاده از روش‌های کمترین توان‌های دوم و ماکسیمم درستنمایی [1] (ML) برآورد می‌شوند. هر یک از این دو روش دارای معایبی در برآورد پارامترهای مدل چندسطحی می‌باشند. جهت رفع مشکلات این روش‌ها از تعمیم آن‌ها استفاده می‌شود. تعمیم روش کمترین توان‌های دوم شامل کمترین توان‌های دوم تعمیم یافته بازگشتی [2] $(IGLS)$(گلداستین، ۱۹۹۵) و

[1] Maximum Likelihood

[2] Iterative Generalized Least Squares

کمترین توان‌های دوم تعمیم یافته بازگشتی محدود شده[۳] $(RIGLS)$(گلداستین و رسبش، ۱۹۹۶) می‌باشد. روش ماکسیمم درستنمایی در برآورد پارامترهای مدل چندسطحی دارای مشکل کم‌برآوردی بوده لذا جهت رفع این مشکل روش ماکسیمم درستنمایی محدود شده[۴] $(REML)$(پینهریو و بیتس، ۲۰۰۰) مورد استفاده قرار گرفته است. روش ماکسیمم درستنمایی شبیه‌سازی شده[۵](SML) نیز جهت تقریب درستنمایی‌هایی با ابعاد بالا بکار می‌رود.

برآوردهای بیزی پارامترها در مدل‌های چندسطحی نیز بسیار مورد توجه قرار گرفته است. نخستین تحقیقات در این زمینه توسط دراپیر و براون (۲۰۰۶) صورت گرفت. پس از آن محققین بیشماری در بهبود روش‌های پرطرفدار مونت کارلوی زنجیر مارکوفی[۶]$(MCMC)$ در برآورد پارامترها تحقیق نمودند. نکته قابل تامل در روش‌های بیزی بویژه $MCMC$، حجم بالای محاسبات ناشی از بزرگ بودن ماتریس واریانس—کواریانس مدل است. بعلاوه همبستگی زیاد پارامترها، بویژه پارامترهای اثرات ثابت و تصادفی، در بعضی مواقع باعث مشکلاتی در همگرایی می‌شود. روشی مانند پارامتری کردن[۷] (گلفند و همکاران، ۱۹۹۵) و روش ساختاری مونت کارلوی زنجیر مارکوفی[۸]$(SMCMC)$ (سرجنت و همکاران، ۲۰۰۰) برای رفع این مشکل پیشنهاد شد.

مسئله مهم از نقطه نظر استنباطی این است که تحلیل چندسطحی خطی تعمیمی از تحلیل رگرسیون خطی می‌باشد. بنابراین برای فهم بیشتر تحلیل چندسطحی می‌توان مدل متناظرش در حالت رگرسیون خطی را در نظر گرفت. در این فصل ابتدا انگیزه استفاده از مدل‌های آماری چندسطحی

---

[۳] Restricted Iterative Generalized Least Squares

[۴] Restricted Maximum Likelihood

[۵] Simulated Maximum Likelihood

[۶] Markov Chain Monte Carlo

[۷] Reparametrization

[۸] Structured Markov Chain Monte Carlo

و مفاهیم پایه‌ای تحلیل آماری مدل چندسطحی مطرح شده و سپس بررسی تعدادی از مدل‌های آماری چندسطحی، روش‌های مقایسه مدل‌های آماری و تعمیم مدل‌های آماری چندسطحی مرور می‌شوند. در فصل دوم از بین روش‌های بسامدی روش‌های کمترین توان‌های دوم،$IGLS$، $RIGLS$، ML، $REML$ و SML مورد بررسی قرار می‌گیرند. در فصل سوم الگوریتم $MCMC$ در مدل‌های چندسطحی و بهبود الگوریتم $MCMC$ در مدل‌های چندسطحی بررسی خواهد شد. در فصل چهارم جهت بهبود روش‌های بیزی برآورد پارامترها در مدل‌های چندسطحی روش‌های متعامدسازی[۹] (هیل و اسمیت، ۱۹۹۲)، بسط پارامتر[۱۰] (گلمن و همکاران، ۲۰۰۸)، روش‌های مرکزی کردن سلسله مراتبی[۱۱] (گلفند و شو، ۱۹۹۹) و ساختاری مونت کارلوی زنجیر مارکف ارائه شده است. اما هر یک از روش‌های مذکور دارای مشکلاتی می‌باشند و بمنظور رفع این مشکلات در این پایان‌نامه استفاده از تجزیه چولسکی ماتریس واریانس—کواریانس مدل ساختاری و تجزیه چولسکی[۱۲] ماتریس واریانس—کواریانس مدل اولیه پیشنهاد می‌شود. نهایتا در فصل پنجم به‌وسیله شبیه‌سازی روش‌های ارائه شده در فصل‌های قبل مورد مقایسه قرار می‌گیرد.

۲.۱ انگیزه استفاده از تحلیل آماری چندسطحی

به‌منظور آشنایی بیشتر با مفهوم تحلیل چندسطحی و کاربرد آن مسئله ساده رگرسیون خطی یک متغیره را در نظر بگیرید. فرض کنید هدف تحقیق بررسی تاثیر یک متغیر مستقل بر روی یک متغیر پاسخ باشد. بطور مشخص فرض کنید محققی در پی بررسی رابطه‌ی بین کلسترول خون بیماران و سن آنها

[۹] Orthogonalization

[۱۰] Parameter Expantion

[۱۱] Hierarchical-Centering

[۱۲] Cholesky decomposition

باشد. جهت توصیف این رابطه می‌توان از رگرسیون خطی به صورت

$$y_i = \beta_0 + x_i\beta_1 + \epsilon_i, \qquad i = 1, ..., n \qquad (۱.۲.۱)$$

استفاده نمود که y_i (متغیر وابسته) بیانگر میزان کلسترول خون بیماران، β_0 و β_1 پارامترهای ثابت رگرسیونی، x_i (متغیر مستقل) نشان دهنده سن بیماران و ϵ_i خطای تصادفی (خطای اندازه‌گیری) مدل بوده که دارای توزیع نرمال با میانگین صفر و واریانس مثبت σ_e^2 است بعلاوه برای $cov(\epsilon_i, \epsilon_j) = 0, i \neq j$.

اکنون فرض کنید هدف تحقیق بررسی تاثیر متغیر جنسیت بر میزان کلسترول خون بیمار در حضور متغیر مستقل سن باشد. از آنجا که متغیر جنسیت یک متغیر اسمی است لذا باید با قرار دادن برچسب‌های مناسب آن را به یک متغیر کمی تبدیل نمود. فرض کنید جنسیت زن و مرد به ترتیب با برچسب‌های یک و صفر نمایش داده شود. آن‌گاه با ورود متغیر جنسیت رابطه (۱.۲.۱) به‌صورت زیر تعمیم داده خواهد شد:

$$y_i = \beta_0 + x_i\beta_1 + x_{2i}\beta_2 + \epsilon_i, \qquad i = 1, ..., n \qquad (۲.۲.۱)$$

که x_i، x_{2i}، β_0 و $\beta_0 + \beta_2$ به ترتیب متغیر مستقل سن بیماران، جنسیت، عرض از مبدا برای مردان و عرض ازمبدا برای زنان را نشان می‌دهند. بنابراین هنگامی‌که متغیر جنسیت وارد مدل می‌شود عرض از مبدا خط رگرسیون برای زنان و مردان متفاوت خواهد بود.

حال فرض کنید در حضور متغیر مستقل سن علاقمند به بررسی تاثیر پزشک معالج بیماران نیز باشید. واضح است که گروهی از بیماران تحت نظر پزشک خاصی بوده و گروهی دیگر نیز با پزشک دیگری در ارتباط می‌باشند. توجه شود که نام پزشک نیز متغیری اسمی می‌باشد و لذا برای ورود آن به

مدل رگرسیونی لازم است از کدبندی جدیدی استفاده شود. بعبارت دیگر باید از متغیرهای مجازی[۱۳] استفاده نمود. می‌دانیم که برای این مثال تعداد متغیرهای مجازی یک عدد کمتر از تعداد پزشکان مورد مطالعه می‌باشد. لذا اگر m تعداد پزشکان و md_i برچسب (کد) پزشک iام باشد آن‌گاه مدل رگرسیونی خطی ساده به‌صورت زیر تغییر خواهد کرد:

$$y_i \;=\; \beta_0 + x_i\beta_1 + md_1\beta_2 + \ldots + md_{m-1}\beta_m + \epsilon_i, \qquad i = 1, ..., n \quad (۳.۲.۱)$$

براساس رابطه رگرسیونی (۳.۲.۱) می‌توان حدس زد که اگر تعداد پزشکان مورد مطالعه خیلی زیاد مثلا ۱۰۰ باشد آن‌گاه تعداد متغیرهای مجازی افزایش خواهند یافت. از آنجا که در آمار مدل‌های ساده‌تر به مدل‌های پیچیده‌تر با پارامترهای اضافی ترجیح داده می‌شوند در این‌گونه موارد باید به‌طریقی از افزایش بی‌مورد متغیرهای مجازی جلوگیری کرد (نتر و وایزرمن، ۱۹۸۳). نکته دیگر در مورد ساختار ماتریس طرح می باشد. می‌توان پیش‌بینی کرد که با افزایش متغیرهای مجازی ماتریس طرح شامل مجموعه زیادی از اعداد طبیعی (صحیح) خواهد شد بطوری‌که امکان بروز هم‌خطی بین آنها وجود خواهد داشت. در آن‌صورت برآورد پارامترهای مدل قابل اعتماد نخواهند بود (پاتریک و همکاران، ۲۰۰۶). اضافه بر این برآورد همه ضرایب در مدلی که تعداد زیادی متغیرهای مجازی دارد باعث کاهش کارایی مدل خواهد شد (گلمن و هیل، ۲۰۰۷).

نکته قابل تامل در مدل‌بندی فوق این است که معمولا هدف اصلی تحقیق بررسی تاثیر پزشک خاص روی کلسترول خون بیمار نیست بلکه هدف کشف تاثیر یا عدم تاثیر جامعه پزشکان بر روی کلسترول خون بیماران می‌باشد. جهت رفع مشکلات فوق و جایگزینی یک مدل مناسب می‌توان پزشکان مورد مطالعه را نمونه‌ای از جامعه پزشکان در نظر گرفت. در این‌صورت خطای تصادفی ناشی از انتخاب پزشک به عنوان متغیر تصادفی وارد مدل می‌شود و به‌جای برآورد تعداد زیاد پارامترها تنها

[۱۳] Dummy variables

کافیست واریانس این متغیر تصادفی برآورد شود. به این منظور مدل (۳.۲.۱) به‌ازاء $i = 1, ..., n_j$ و $j = 1, ..., m$ بصورت

$$y_{ij} = \beta_0 + u_j + x_{ij}\beta_1 + \epsilon_{ij} \qquad (۴.۲.۱)$$

تغییر خواهد کرد که u_j خطای تصادفی ناشی از فرآیند نمونه‌گیری از جامعه پزشکان است. این خطا مستقل از خطای اندازه‌گیری ϵ_{ij} در سنجش میزان کلسترول خون بیماران بوده و دارای توزیع نرمال با میانگین صفر و واریانس σ_u^2 می‌باشد. بعلاوه u_jها مستقل بوده و در این حالت برآورد σ_u^2 بعنوان معیار تغییرات بین پزشکان مدنظر بوده و معنی دار بودن آن نشانگر تاثیر پزشک بر میزان کلسترول خون است.

مدل ارائه شده در رابطه (۴.۲.۱) به مدل چندسطحی با عرض از مبدا تصادفی[۱۴] معروف است. قبل از تشریح بیشتر این مدل و ارزیابی معایب و مزایای آن نسبت به مدل‌های مرسوم رگرسیونی مفاهیم و اصطلاحات پایه‌ای مدل‌های چندسطحی مورد مطالعه قرار می‌گیرد. سپس جزییات بیشتر در مورد این مدل و همچنین مدل‌های دیگر مورد بحث واقع می‌شود. ضمنا توجه شود که مدل‌های دیگری مانند مدل‌های متقاطع—طبقه‌بندی شده[۱۵] نیز وجود دارند که از ساختار آشیانه‌ای پیروی نمی‌کنند اما بعنوان مدل‌های چندسطحی در نظر گرفته می‌شوند. بررسی چنین مدل‌هایی در این پایان‌نامه صورت نخواهد گرفت. خواننده علاقمند می‌تواند به منابع موجود مانند گلداستین (۱۹۹۵) و گلمن و هیل (۲۰۰۷) مراجعه کند.

[۱۴] Random intersept

[۱۵] Cross classified

۳.۱ مفاهیم پایه‌ای تحلیل آماری چندسطحی

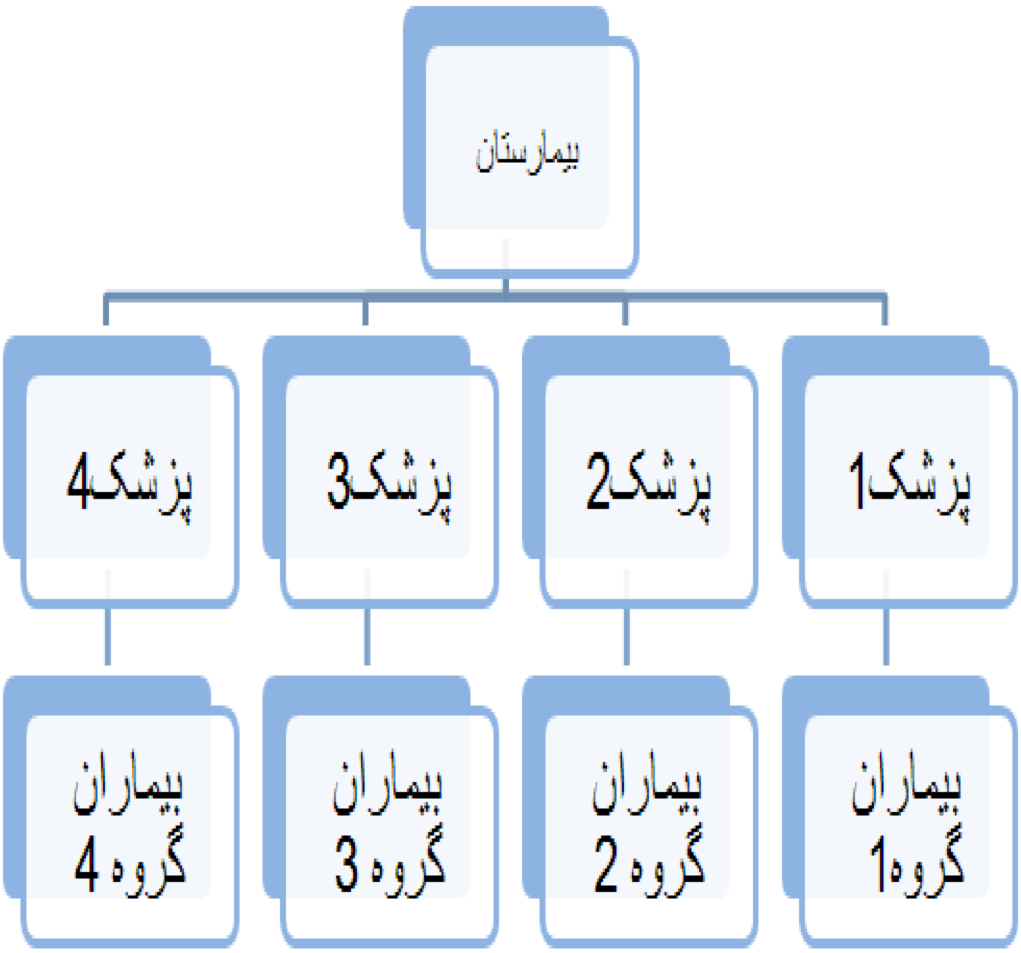
دلیل اصلی نامگذاری مدل‌های چندسطحی به وجود چند سطح متفاوت می‌باشد. درواقع سطوح مدل خوشه‌های مختلفی می‌باشند که به‌طریقی مرتب در یکدیگر آشیانه کرده‌اند. عنوان سطح فاکتورهای مورد مطالعه نامیده می‌شود. به اولین فاکتوری که در تمامی فاکتورهای دیگر آشیانه کرده سطح اول می‌گویند. فاکتوری که فاکتور اول درون آن آشیانه کرده و خود در فاکتور دیگری آشیانه کرده باشد سطح دوم نامیده می‌شود. بهمین ترتیب سطوح دیگر معرفی می‌گردند. واضح است که مدل دوسطحی[۱۶] دارای سطح اول و دوم است. مدل سه‌سطحی[۱۷] شامل سه‌سطح و مدل چهارسطحی شامل چهارسطح می‌باشند. شکل ۱.۳.۱ بصورت نمایشی ساختار یک مدل سه‌سطحی را نشان می‌دهد. بر اساس شکل، بیماران سطح اول، پزشکان سطح دوم و بیمارستان‌ها سطح سوم را تشکیل می‌دهند.

نامگذاری بیماران بصورت گروه ۱و ۲ و ... تنها با هدف آشکارسازی ساختار چندسطحی صورت گرفته است وگرنه دلیلی وجود ندارد که از قبل هر گروه خاصی از بیماران منتسب به پزشکان خاصی باشد. پزشکان انتخابی نمونه‌ای تصادفی از جامعه پزشکان و بیمارستان‌های انتخابی نیز نمونه‌ای تصادفی از کلیه بیمارستان‌های یک شهر می‌باشند.

---

[۱۶] 2-levels model

[۱۷] 3-levels model



شکل ۱.۳.۱: بررسی کلسترول خون بیماران تحت درمان ۴ پزشک. نمودار نشان می‌دهد سطح اول بیماران تحت نظر پزشکان متفاوت می‌باشند که در سطح دوم یعنی پزشکان آشیانه کرده‌اند و سطح دوم در فاکتور دیگری که بیمارستان است آشیانه کرده است.