



دانشگاه علوم پزشکی کرمان

دانشکده بهداشت

پایان نامه مقطع کارشناسی ارشد رشته آمار زیستی

عنوان

تعیین مدل بهینه آمیخته خطی در پیش بینی روند تغییرات هموگلوبین A1C و قند خون  
ناشتا افراد دیابتی نوع ۲ با استفاده از داده‌های مرکز تحقیقات غدد و متابولیسم صدیقه طاهره  
اصفهان

توسط: الهه کاظمی

استاد راهنمای اول: دکتر عباس بهرامپور

استاد راهنمای دوم: دکتر سید محسن حسینی

استاد مشاور: دکتر الهام فقیه ایمانی

سال ۱۳۹۲

**Determination of Linear Mixed effect models in predictive of trend of Hemoglobin A1C and fasting blood sugar on type2 diabetes using Isfahan Endocrine and Metabolism Research Center data (I.E.M.R.C), Isfahan**

A Thesis

Presented to

The Graduated Staudies

By

**Elahe Kazemi**

Supervisors

**Dr abbas bahrampur and Dr. seyed mohsen hosseini**

Advisor

**Dr elham faghiih imani**

In Partial Fulfillment

Of the Requirements for the Degree

Master of science in:

**Biostatistics**

**Kerman University of Medical Scieces**

**2014**

## چکیده

**مقدمه:** بیماری دیابت یکی از شایع‌ترین بیماری‌های متابولیک در دنیا و یک معضل برای سلامت عمومی است. شواهد دال بر آن است که کنترل فندخون بیمار می‌تواند خطر عوارض ناتوان کننده و حتی کشنده دیابت را کاهش دهد. در این راستا مطالعات بسیاری در مورد تعیین عوامل در بهبود و کنترل این بیماری، جهت پیشگیری از بروز و پیشرفت عوارض مزمن آن صورت گرفته است؛ اما مطالعات صورت گرفته اکثراً به صورت مقطعی بوده و تنها عوامل مؤثر بر کنترل آن را سنجیده‌اند و ماهیت همراهی، نامعلوم است که این، ضرورت اجرا در یک مطالعه طولی را مشخص می‌کند. مطالعه طولی در مقایسه با مطالعه مقطعی کارایی بیشتری داشته و می‌تواند توان آماری را افزایش دهد. صرف نظر از وقت‌گیر بودن، پرهزینه بودن، نیاز به سازماندهی، پیگیری مداوم و پیچیده‌تر بودن تحلیل مطالعات طولی نسبت به مطالعات مقطعی، به جهت امکان بررسی روابط همراهی در این نوع مطالعات نسبت به مطالعات مقطعی امروزه استفاده از این مدل‌ها در مطالعات پزشکی افزایش چشمگیری یافته است. اغلب مطالعات طولی با توجه به تغییر یا عدم تغییر بیمارها در طی زمان دو هدف مقایسه بین اثرات این بیمارها یا تعریف دقیق روند پاسخ در طول زمان و مقایسه بین گروه‌های مختلف را دنبال می‌کنند. مهمترین مزیت مطالعه طولی تفکیک نمودن اثر زمان از تفاوت‌های فردی است. که در مطالعه مقطعی امکان پذیر نمی‌باشد. در داده‌های طولی مشاهدات درون هر آزمودنی به هم وابسته‌اند و به منظور استنباط علمی و معتبر باید این همبستگی در هنگام تحلیل در نظر گرفته شود، و نادیده گرفتن این همبستگی می‌تواند تأثیر نامطلوبی بر برآورد پارامترها، آزمون فرضیه‌ها و کارایی طرح مطالعه داشته باشد. اکثر روش‌های آماری جهت تحلیل اندازه‌های تکرار شده، شامل مدل‌های چندمتغیره یا چندمرحله‌ای (چند سطحی) می‌باشند. اگر طرح متعادل و داده‌های گمشده به صورت تصادفی باشند می‌توان از مدل‌های چندمتغیره استفاده کرد ولیکن یکی از ضعف‌های این روش‌ها آن است که فقط عیانگین‌های گروهی را برآورد و مقایسه می‌کنند و هیچ اطلاعی در مورد تغییرات فردی نمی‌دهند. با توجه به اهمیت داده‌های گمشده در مطالعات طولی که اغلب محققان در روند اجرای پژوهش خود با آن مواجه می‌شوند یک روش عمده و پرکاربرد تحلیل و مدل‌سازی داده‌های طولی و اندازه‌های تکرار شده، تحلیل مدل‌های آمیخته با اثرات تصادفی می‌باشد که مدعی سازگاری با داده‌های گمشده در داده‌های طولی است. سایر روش‌های مطرح برای کار با مقادیر گمشده، روش الگوریتم EM و روش چانه‌ی چندگانه می‌باشند.

این پژوهش به منظور معرفی مدل آمیخته خطی و کاربرد آن در بررسی روند تغییرات هموگلوبین A1C و فند خون ناشتا افراد دیابتی نوع ۲ با استفاده از کلیه عامل‌های دخیل در آن و بررسی روش‌های برخورد با داده‌های گمشده در مطالعات طولی طراحی گردیده است.

**روش‌ها:** در این مطالعه طولی از داده‌های موجود در بانک اطلاعاتی مرکز غدد و متابولیسم صدیقه‌طاهره اصفهان، در سال ۹۱-۷۹ شامل بیماران مراجعه کننده به مرکز، استفاده گردید؛ و با بکارگیری مدل آمیخته خطی و نرم‌افزار stata به بررسی روند

تغییرات هموگلوبین A1C افراد دیابتی نوع ۲ در حضور کلبه عامل‌های دخیل در آن پرداخته شد. در مرحله بعد با توجه به اهداف پژوهش، جهت بررسی اثر داده‌های گمشده از میان داده‌های موجود، مجموعه داده کامل و بدون داده گمشده‌گی در هیچ یک از متغیرها انتخاب شد. مجموعه داده جدید، داده کامل و مدل حاصل از آن مدل کامل نامیده شد. در داده کامل، به ترتیب ۱۰، ۲۰، ۳۰، ۵۰ و ۷۰ درصد داده‌های موجود در ستون متغیر HbA1C (متغیر پاسخ) به صورت کاملاً تصادفی حذف شد. در مرحله بعدی داده‌های گمشده توسط نرم‌افزار Spss و به دو روش الگوریتم EM و جایگزینی چندگانه MI برآورد شدند. در نهایت مدل‌های بدست آمده به روش مقایسه معیارهای برازش با مدل کامل مقایسه شدند.

**یافته‌ها:** پس از برازش مدل، تعیین شد کنترل تری‌گلسرید، کلسترول تام، LDL بالای نرم، نوع درمان بکارگرفته شده، فشارخون دیاستولیک، دوره ابتلا، زمان ورود به مطالعه (در واقع بیانگر پیشرفت دانش پزشکی)، همچنین ابتلا به بیماری رتینوپاتی یا پروتینوری، به طور همزمان عواملی تعیین کننده بر روند بهبود و کنترل قندخون افراد می‌باشند. یافته‌های این مطالعه نشان داد در روش حذف موارد دارای داده گمشده، با افزایش میزان داده گمشده، برازش مدل کاهش می‌یابد، هنگامی که داده‌های گمشده توسط روش‌های مختلف انتخاب متغیر برآورد می‌شوند، معیارهای اطلاع و تابع احتمال مدل نسبت به حالتی که موارد دارای داده گمشده حذف شوند افزایش قابل ملاحظه‌ای نمی‌یابد. به‌طور کلی نتایج نشان از مناسب‌تر بودن روش برآورد چندگانه نسبت به روش جایگزینی نکی توسط الگوریتم EM و بهتر بودن روش‌های جایگزینی نسبت به روش‌های حذفی دارد. نتیجه‌گیری: مدل آمیخته خطی مدلی مناسب جهت برازش به داده‌های یک طرح طولی با میزان داده گمشده زیاد نبوده و روش جانهی چندگانه علاوه بر پیچیدگی‌های محاسباتی روشی مناسب جهت برخورد با داده‌های گمشده در یک مدل طولی می‌باشد. نتایج بر اهمیت کنترل شاخصهای خونی و تجویز درمان مناسب توسط پزشک جهت پیشگیری از عوارض این بیماری تاکید می‌کند.