پیش‌بینی سن تقویمی بر اساس سن دندانی دمیرجین با استفاده از روش رگرسیون ریج استوار

مهدی روزبه [۱*](Ph.D)، سید محمد ملک جعفریان [۲](Ph.D)، منیره معنوی [۱](M.Sc)، ملیحه سادات ملک جعفریان [۳](M.Sc)

۱- دانشکده علوم ریاضی، آمار و علوم کامپیوتر، دانشگاه سمنان، سمنان، ایران

۲ - دانشگاه علوم پزشکی شهید صدوقی یزد، یزد، ایران

۳- شرکت آب و فاضلاب شهری استان سمنان، سمنان، ایران

* نویسنده مسئول، تلفن: ۰۲۳-۳۱۵۳۳۲۰۶ mahdi.roozbeh@semnan.ac.ir

تاریخ دریافت: ۱۳۹۷/۱۰/۱۳ تاریخ پذیرش: ۱۳۹۸/۷/۳

چکیده

هدف: تخمین سن، نقش مهمی در پزشکی قانونی، بیماری‌های غدد و دندان‌پزشکی بالینی دارد. هم‌چنین ارزیابی مراحل تکاملی دندانی نسبت به رویش دندان با ارزش‌تر است. در این پژوهش به مدل سازی سن تقویمی با استفاده از روش‌های نوین و غنی آماری پرداخته شده است به طوری‌که می‌تواند به عنوان یک روش عملی که تلفیقی از چند روش نوین آماری است، در علم پزشکی مورد بررسی قرار گیرد.

مواد و روش ها: در بین روش‌هایی که برای تعیین سن استفاده می شود، رایج‌ترین روش مورد استفاده در سراسر دنیا، روش نوین دمیرجین مودیفای است که بر اساس کلسیفیکاسیون دندان دائمی در رادیوگرافی پانورامیک می‌باشد. جامعه مورد بررسی در این تحقیق تعداد ۸۷ نفر از بیماران مراجعه‌کننده به صورت روش تصادفی ساده در طی دوره ۱۲ماهه سال ۹۴-۱۳۹۳ به درمانگاه خاتم‌الانبیا استان یزد مراجعه نموده‌اند، می‌باشد. با استفاده از سن تخمینی از دندان مولر سوم و متغیر جنسیت به ارزیابی سن تقویمی پرداختیم. در تجزیه و تحلیل مسائل رگرسیونی و به‌ویژه مدل‌بندی آماری بسیاری از داده‌ها مانند داده‌های اقتصادی، روانشناسی، علوم اجتماعی، علوم پزشکی، مهندسی و ... با مشکل هم‌خطی در میان متغیرهای پیشگو و حضور نقاط دورافتاده در مجموعه داده‌ها مواجه هستیم. روش رگرسیون کم‌ترین توان‌های دوم خطا در برآورد پارامترهای مدل رگرسیونی نسبت به داده‌های پرت بسیار حساس است. اغلب روش‌های موجود برآوردیابی پارامترهای این مدل‌ها با رویکرد کم‌ترین توان‌های دوم خطا، تحت تأثیر داده‌های پرت، برآوردهایی نامناسب، دور از انتظار و با نرخ خطای زیاد ارائه می‌دهند. برای غلبه بر مشکل مشاهده‌های دورافتاده از روش‌های استوار استفاده می شود. هم‌چنین برای حل مشکل هم‌خطی چندگانه استفاده از رگرسیون ریج توصیه می شود. لذا در این مطالعه برآوردگرهای رگرسیونی ریج استوار در مدل سازی متغیرهای وابسته معرفی خواهد شد که نسبت به داده‌های پرت حساسیت کم‌تری دارند.

یافته‌ها: میانگین سنی افراد مورد مطالعه ۱۷,۲۱ سال به تفکیک جنسیتی ۶۷ درصد زن و ۳۳ درصد مرد بوده است. هم‌چنین در بررسی ارتباط بین سن تخمینی ۴ دندان به ترتیب بین دندان‌های بالا-راست و بالا-چپ و دندان‌های پایین-راست و پایین-چپ با ضریب هم‌بستگی بالای ۷۰ درصد بوده است. میزان همبستگی بین دندان‌های فک بالا و پایین حدود ۳۰ درصد و بین دندان‌های چپ و راست ۶۰ درصد بوده‌است. دلیل استفاده از مدل رگرسیون استوار ریج در این پژوهش، وجود داده‌های پرت و هم‌خطی بین متغیرهای مستقل است.

نتیجه گیری: لزوم استفاده از روش‌های پیشرفته‌ی آماری در علوم پزشکی در تحقیقات حاضر حائز اهمیت است که جهت برگزیدن بهترین مدل نیاز به بررسی دقیق داده‌ها داریم. لزوم استفاده از روش‌های پیشرفته‌ی آماری در علوم پزشکی در تحقیقات اخیر بسیار حائز اهمیت است که جهت برگزیدن بهترین مدل نیاز به بررسی دقیق داده‌ها داریم. در این تحقیق مدل برازش داده شده برای تخمین سن بر اساس روش رگرسیون استوار ریج، نسبت به سایر روش‌ها کاراتر است.

واژه‌های کلیدی: برآورد ریج، داده پرت، دندان مولر سوم، تحلیل رگرسیون استوار، سن تقویمی، سن دندانی دمیرجین، هم‌خطی

مقدمه

تخمین سن، نقش مهمی در پزشکی قانونی، بیماری‌های غدد، ارتودنسی، دندان‌پزشکی اطفال، باستان‌شناسی و

[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]

[DOI: 10.29252/koomesh.22.2.228]

دندان‌پزشکی بالینی دارد. در کشورهای در حال توسعه که ثبت دقیق زمان تولد افراد وجود ندارد، این روش راه بسیار مناسبی برای تخمین سن تقویمی افراد است [۱]. اغلب بین بلوغ فیزیکی و سن فرد ارتباط وجود دارد [۲]. برخی از تظاهرات بلوغ فیزیکی مانند سن اسکلتی، قاعدگی، اندازه و قد افراد و کلسیفیکاسیون دندانی در نبود اطلاعات دموگرافیک دقیق افراد، می‌تواند در تعیین سن مورد استفاده قرار گیرد [۳]. ارزیابی مراحل تکاملی دندانی نسبت به رویش دندان با ارزش‌تر است زیرا زمان رویش دندان کوتاه می‌باشد و با زمان ظهور دندان در دهان تعیین می شود و این رخداد با عوامل موضعی مانند کمبود فضا و عوامل سیستمیک مانند وضعیت تغذیه‌ای تغییر می‌کند [۱].

سن دندانی می‌تواند بر اساس رویش یا مراحل شکل‌گیری دندان در رادیوگرافی ارزیابی شود. در این بین مدل‌سازی جهت تخمین سن تقویمی با استفاده از روش‌های نوین و غنی آماری حائز اهمیت است. در بین روش‌هایی که برای تعیین سن استفاده می شود، رایج‌ترین روش مورد استفاده در سراسر دنیا، روش دمیرجین است. این روش بر اساس کلسیفیکاسیون دندان دائمی در رادیوگرافی پانورامیک می‌باشد. مشکلی که در بررسی سن دندانی با روش دمیرجین وجود دارد زمانی است که تقریباً تمام ریشه‌های دندان‌های دائمی بیمار تکامل یافته و ریشه‌های دندان های عقل بیمار باید به عنوان ملاک مورد بررسی قرار گیرند. دمیرجین و همکاران در سال ۲۰۰۴ روش جدیدی به نام دمیرجین مودیفای (Modified Demirjian) ابداع کردند [۳]. لازم به ذکر است استفاده از این روش برای جامعه مورد مطالعه در کشور ایران و مدل‌سازی سن افراد برای اولین بار صورت گرفته است.

پس از رویش مولر دوم دائمی در حدود سنین ۱۳- ۱۲ سالگی تعیین سن با مشکل مواجه می‌گردد. این امر باعث می‌شود که در این سنین مولر سوم مورد توجه فراوان قرار گیرد، چرا که هنوز در حال شکل‌گیری است. این دندان در سنین ۲۱- ۱۷ سالگی در دهان ظاهر می‌گردد و با توجه به فرآیند تشکیل ریشه می‌توان مراحل تکاملی آن را تعریف کرد که تقریباً از ۱۴ سالگی آغاز می‌گردد [۴،۵]. در کنار استفاده از روش‌هایی هم‌چون نولا و دمیرجین در تعیین سن با کمک مولر سوم [۴،۵]، روش‌های دیگری نیز وجود دارد که طبق تحقیقات صورت گرفته روش دمیرجین بر سایر روش‌ها عملکرد بهتری دارد.

عزالدینی اردکانی و همکاران [۶] مطالعه‌ای با هدف مطالعه مقطعی از نوع تشخیصی با همکاری سازمان پزشکی قانونی و دانشکده دندان‌پزشکی شهید صدوقی یزد در سال ۸۴-۱۳۸۳ بر روی ۵۸ بیمار با محدوده سنی ۱۵ تا ۲۵ سال که به منظور جراحی دندان عقل به مطب خصوصی مراجعه کرده بودند انجام دادند. محاسبات آماری از طریق آزمون‌های تی زوجی، ضریب هم‌بستگی، ویل کاکسون (Wilcoxon) و کلموگروف اسمیرنف (Kolmogorov-Smirnov) انجام شد. کم‌ترین خطاهای به دست آمده در تخمین سن با سن تقویمی بیمار، مربوط به دندان عقل کشیده شده سمت راست فک پایین بود.

جوادی‌نژاد و همکاران در سال ۲۰۰۸ مطالعه‌ای با هدف تعیین سن دندانی به کمک رادیوگرافی پانورامیک به روش دمیرجین و تعیین دقت روش دمیرجین در تخمین سن تقویمی کودکان شهر اصفهان انجام دادند. که به این نتیجه رسیدند که از این روش می‌توان برای تعیین سن در مواردی که نیاز به تعیین سن وجود دارد (از جمله در پزشکی قانونی) استفاده کرد [۷].

بر اساس بررسی صورت گرفته در این پژوهش که با نتایج مطالعات متعددی همخوانی دارد، وجود هم‌بستگی بین سن کرونولوژیک و سن دندانی هر چهار دندان مولر سوم (بر اساس روش دمیرجین) می‌باشد [۴-۱۰].

هم‌چنین در مطالعه‌ی تفکری و همکاران نیز با روش دمیرجین، بیش‌ترین هم‌بستگی به ترتیب بین دندان‌های بالا-راست (UR) و بالا-چپ (UL) و دندان‌های پایین-راست (LR) و پایین-چپ (LL) گزارش شده که با نتایج سایر مطالعات متعدد نیز مطابقت دارد. لازم به ذکر است که در این پژوهش نیز همین نتیجه بازگو شد [۱۱،۱۲].

علاوه بر این در مطالعه باقرپور بر روی جمعیت ایرانی با استفاده از روش گلیزر و هانت مودیفای (Gleiser & Hunt modified)، سن دندانی UR و UL بیش‌ترین ضریب هم‌بستگی را داشتند [۱۳]. اما مطالعه او که با روش دمیرجین بر روی جمعیت ایرانی ۲۷-۱۰ ساله انجام شده است، بیش‌ترین ضریب هم‌بستگی را بین دندان‌های LL و UL و به میزان ۹۴درصد نشان داد علت این تفاوت می‌تواند مربوط به تفاوت در رده سنی مورد بررسی در مطالعه باشد، چنان که در مطالعه حاضر رده سنی ۱۹-۱۴ سال مورد بررسی قرار گرفت [۸].

طبق مطالعه و بررسی سوابق پژوهشی در امر تخمین سن تقویمی، تاکنون مدلی آماری جهت تخمین سن تقویمی بر اساس بررسی هم‌زمان هر چهار دندان عقل تدوین نشده است. از این حیث نگارش و تحقیقات صورت گرفته‌ی این پژوهش، نوآوری لازم را دارا می‌باشد.

مواد و روش‌ها

بیماران مراجعه‌کننده در طی دوره ۱۲ ماهه سال ۹۴-۱۳۹۳به درمانگاه خاتم‌الانبیا استان یزد، جامعه مورد بررسی

[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]

[DOI: 10.29252/koomesh.22.2.228]

تحقیق را تشکیل داده‌است. تعداد ۸۷ نفر از بیماران مراجعه‌کننده که به صورت روش تصادفی ساده انتخاب شده‌اند.

معیارهای ورود و خروج نمونه‌ها در مطالعه:

۱. عدم وجود مشکلات رشدی – تکاملی

۲. کیفیت قابل قبول رادیوگرافی پانورامیک

۳. وجود حداقل یک دندان عقل

۴. عدم وجود شکستگی و یا ضایعاتی پیرامون دندان عقل

تعداد ۸۷ رادیوگرافی پانورامیک دیجیتال مربوط به بیماران نوجوان ارتودنسی مراجعه‌کننده به کلینیک خاتم‌الانبیا از آرشیو کلینیک بررسی شد. جنسیت و سن کرونولوژیک دقیق افراد (دارای روز و ماه تولد علاوه برسال) از روی پرونده بیماران ثبت شد، در صورت عدم وجود در پرونده از طریق تماس با بیمار به طور مستقیم پرسیده شد. رادیوگرافی‌ها دارای دندان‌های عقل بوده و بین سنین ۱۴ تا ۱۹ قرار داشت.

روش‌های رگرسیونی یکی از پرکاربردترین روش‌های آنالیز آماری در علوم پزشکی است. در اولویت بالاتر انتخاب بهترین زیرمدل یکی از مباحث مهم در ساختن مدل‌های رگرسیونی می‌باشد که شرط آن شناخت و مطالعه‌ی کامل متغیرها می‌باشد. در حالت کلی مدل رگرسیون خطی چندگانه با p متغیر پیش‌بین به‌صورت زیر تعریف می‌گردد: (1)

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, i = 1, \dots, n$$

که در آن y_iها و x_{ij}ها به ترتیب مشاهدات وابسته و مستقل معلوم، β_iها پارامترهای نامعلوم و ε_iها خطاهای تصادفی غیر قابل مشاهده هستند که معمولاً ناهمبسته با میانگین صفر و واریانس ثابت σ^2 فرض می‌شوند.

در مواقعی که بین متغیرها رابطه‌ی خطی (هم‌خطی) وجود داشته باشد، معمولاً استفاده از روش‌های کلاسیک رگرسیونی مفید واقع نمی‌شود که در این مطالعه از روش ریج برای رفع این مشکل استفاده شده‌است. هم‌چنین وجود داده‌های پرت در رگرسیون کم‌ترین توان‌های دوم مشکل ایجاد کرده و نتایج حاصل را تحت تاثیر قرار می‌دهد، لذا از روش رگرسیون ریج استوار استفاده می‌شود.

رویکرد رگرسیون استوار

به طور کلی هدف از رگرسیون، تعیین روابط بین متغیرها، پیش‌بینی و تحلیل روابط به‌دست آمده است. فرم ماتریسی مدل رگرسیونی (۲-۱) به صورت زیر است:

$$y = X\beta + \varepsilon \qquad (2)$$

که در آن y بردار n×1 از مشاهدات متغیر وابسته، $X = (x_1, \dots, x_n)'$ ماتریس n×p از مشاهدات متغیرهای مستقل، β بردار p×1 از ضرایب نامعلوم و ε بردار n×1 خطای تصادفی می‌باشد. هدف برآورد بردار نامعلوم β به کمک مشاهدات است.

برآورد کم‌ترین توان‌های دوم پارامتر β در مدل بالا عبارت است از: (3)

$$\hat{\beta} = \underset{\beta}{\operatorname{Arg\,min}} \sum_{i=1}^{n} e_i^2 = \underset{\beta}{\operatorname{Arg\,min}} (y - X\beta)'(y - X\beta) = (X'X)^{-1}X'y$$

ادگارد در سال ۱۸۸۷ استدلال نمود که نقاط دورافتاده تاثیر زیادی بر روش کم‌ترین توان‌های دوم دارند، زیرا روش کم‌ترین توان‌های دوم در برآورد پارامترها به هر مشاهده وزن یکسان می‌دهد، یعنی وضعیت مقادیر نقش مهمی در برازش کم‌ترین توان‌های دوم بازی می‌کند [۱۴]. به بیان دیگر همه نقاط در برآورد، وزن یکسانی دارند. به طور کلی اگر در تحلیل به این نکته برخورد کنیم که خطاها توزیع غیرنرمال دارند، مخصوصاً در مواقعی که توزیع خطا دارای دم‌های پهن‌تری نسبت به توزیع نرمال باشد، یعنی از احتمال‌های بیش‌تری در دم‌ها نسبت به توزیع نرمال برخوردار باشد، بایستی یک روش رگرسیون استوار در نظر بگیریم.

روش‌های استوار قادرند به مشاهدات وزن‌های نابرابر اختصاص دهند. به طور کلی مشاهداتی که باقی‌مانده‌های بزرگ‌تری را تولید می‌کنند به وسیله برآورد استوار کم وزن‌ترند. در حالت کلی یک برآوردگر استوار نامیده می‌شود هر گاه با به‌کار بردن آن بتوان نتایج قابل اطمینانی حداقل در شرایط زیر گرفت:

۱- برخی فرضیه‌های زیربنایی روش برقرار نباشد.

۲- در داده‌ها مشاهدات پرت وجود داشته باشد [۱۵،۱۶].

روش‌های متنوعی جهت مدل‌بندی رگرسیون استوار وجود دارد که در این مطالعه ما از روش کم‌ترین توان‌های دوم پیراسته (LTS: Least Trimmed Squares) استفاده کرده‌ایم. برآورد استوار پارامتر β در مدل (۲) به روش رگرسیون LTS عبارت است از: (4)

$$\hat{\beta}^{LTS} = \underset{\beta}{\operatorname{Arg\,min}} \sum_{i=1}^{h} e_{(i)}^2 = \underset{\beta}{\operatorname{Arg\,min}} (y - X\beta)' Z (y - X\beta) = (X'ZX)^{-1}X'Zy$$

که در آن $e_{(1)}^2 \le \cdots \le e_{(n)}^2$ و h $(n/2 \le h \le n)$ تعداد مشاهدات خوب در مجموعه n تایی مشاهدات است که ممکن است معلوم یا مجهول باشد. این روش دارای الگوریتمی است که با جستجو در مجموعه مشاهدات به برآورد بهترین پارامترهای مدل بر اساس ترکیب‌های مختلف انتخاب h مشاهده خوب از مجموعه n تایی مشاهدات، می‌پردازد. هم‌چنین Z ماتریس قطری با عناصر قطر اصلی $z = (z_1, \dots, z_n)'$ می‌باشد، که در آن z_i مقدار یک اختیار می‌کنند اگر مشاهده iام خوب باشد و مقدار صفر اختیار می‌کند هر گاه مشاهده iام پرت باشد و داشته باشیم $\sum_{i=1}^{n} z_i = h$ در توان دوم باقی‌مانده‌ها مرتب شده و تفاوت این روش با روش کم‌ترین توان‌های دوم این است که

[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]

[DOI: 10.29252/koomesh.22.2.228]

باقی‌مانده‌های توان دوم بزرگ را به‌کار نمی‌بریم و در نتیجه برازش دور از نقاط پرت باقی می‌ماند [۱۷].

رویکرد رگرسیون ریج

وجود هم‌بستگی بین متغیرهای مستقل مدل، باعث به‌وجود آمدن هم‌خطی می شود. اگر یکی از متغیرهای پیش‌بین یک تابع دقیق خطی از یک یا چند متغیر پیش‌بین دیگر باشد گوئیم رگرسیون دارای هم‌خطی کامل است. هم‌خطی ناقص وقتی اتفاق می‌افتد که یکی از متغیرهای پیش‌بین به‌طور تقریبی یک تابع خطی از یک یا چند متغیر پیش‌بین دیگر باشد. در نظر نگرفتن پیش‌فرض‌ها در رگرسیون خطی می‌تواند روی نتیجه‌گیری‌ها اثر نامطلوب گذاشته و نتایجی نادرست به‌دست آید. یکی از این پذیره‌ها وجود عدم وابستگی در ماتریس داده‌ها ست که در صورت وجود باید از روش‌ها و تکنیک‌های مناسب رگرسیونی استفاده کرد. هم‌خطی چندگانه یکی از مشکلات متداول در مدل‌های رگرسیون خطی چندگانه است که به دلیل ایجاد تورم در ماتریس کوار یانس، باعث عملکرد ضعیف برآوردگرهای کلاسیک و در نهایت پیش‌بینی نادقیق می‌شود. روش مورد استفاده در این پژوهش برآوردگر ریج است که طبق تحقیقات به‌دست آمده میانگین توان‌های دوم خطا یا MSEی این روش حتی از روش کم‌ترین توان‌های دوم کم‌تر است و دقت مدل بیش‌تر می‌شود [۱۸]. هورل و کنارد رگرسیون ریج را با افزودن تابع جریمه $\sum_{i=1}^{n}\beta_i^2$ به تابع زیان توان دوم خطا و بهینه سازی آن حاصل معرفی کردند [۱۹]. در واقع آن‌ها برای اولین بار از رگرسیون ریج برای رفع مشکل هم‌خطی استفاده نمودند [۲۰].

برآورد ضرایب رگرسیونی ریج مدل (۲) به صورت زیر است: (5)

$$\hat{\beta}(k) = (X'X + kI_p)^{-1}X'y, k > 0$$

که در آن I_p ماتریس همانی از مرتبه p و k پارامتر ریج نامیده شده و می‌بایست بر اساس مشاهدات با مینیمم سازی MSE محاسبه گردد. به طور کلی در رگرسیون ریج به دنبال k بهینه هستیم تا از یک سو اثر هم‌خطی متغیرها را از بین ببریم و از سوی دیگر MSE را مینیمم نماییم.

رویکرد رگرسیون ریج استوار

اکنون با ترکیب دو روش کم‌ترین توان‌های دوم پیراسته و ریج می‌توان روش رگرسیون ریج استوار را برای حل هم‌زمان هم‌خطی چندگانه و وجود مشاهدات پرت به صورت زیر تعریف کرد: (6)

$$\hat{\beta}^{LTS}(k) = (X'ZX + kI_p)^{-1}X'Zy, k > 0$$

برای استفاده از برآوردگر فوق می‌بایست پارامتر ریج بهینه با مینیمم سازی MSE محاسبه و مشخص گردد [۲۱]. که این کار در نرم‌افزار R و با استفاده از کتابخانه ltsbase به‌راحتی انجام‌پذیر است.

نتایج

در ابتدای تحلیل، به بررسی داده‌ها از منظر آماره‌های توصیفی می‌پردازیم. در جدول ۱ درصد فراوانی بیماران به تفکیک سن تقویمی و تخمینی پارامترها به نمایش گذاشته شده است. در خصوص روش دمیرجین لازم به توضیح است که در این روش از کلسیفیکاسیون دندان‌های دایمی (بافت و تجمع کلسیومی دندان) در رادیوگرافی پانورامیک، برای تقریب سن استفاده می‌شود که جزء رایج‌ترین روش‌ها در سراسر دنیا می‌باشد.

جدول۱. آماره های توصیفی به تفکیک سن تقویمی و آزمون تخمین دمیرجین

	مینیمم	ماکزیمم	میانگین	انحراف معیار
سن تقویمی	۱۳/۵	۲۰/۵۸	۱۷/۲۱	۱/۱۶۶
آزمون دمیرجین	۱۳/۷۵	۱۹/۵	۱۶/۸۸	۱/۰۸

در شکل ۱ نگاه کلی به ارتباط و میزان هم‌بستگی بین تک تک‌متغیرها که شامل دندان‌های مولر سوم فک بالا و فک پایین و سن تقویمی می‌باشد، هستیم. همان‌طور که ملاحظه می‌گردد بر اساس شکل‌های ستونی رسم شده، متغیرها نرمال نمی‌باشد. هم‌چنین بر اساس مقادیر ضرایب همبستگی، ملاحظه می‌گردد که بین متغیرهای سن تخمینی همبستگی بالایی وجود دارد که این مقدار برای سن تخمینی LR و LL به میزان ۸۶ درصد می‌باشد. پس استفاده از رگرسیون ریج ضروریست و رگرسیون کم‌ترین توان‌های دوم معمولی کارساز نخواهد بود. هم‌چنین نرمال نبودن توزیع متغیرهای مستقل در شکل قابل ملاحظه می‌باشد. البته لازم به ذکر است که نرمال بودن متغیرهای مستقل جزء فرضیات اولیه و لازم در پیاده‌سازی رگرسیون خطی نمی‌باشد ولی با انجام تبدیلات نرمال‌سازی متغیرها شامل متغیرهای مستقل و وابسته، دست‌یابی به یک مدل رگرسیون قوی‌تری را امکان‌پذیر می‌کند.

تبدیل باکس کاکس که توسط باکس و کاکس در سال ۱۹۶۴ معرفی شده است، روی داده‌های اکیداً مثبت انجام می‌شود و در واقع این تبدیل نزدیک‌کننده توزیع متغیرها به توزیع نرمال چندمتغیره می‌باشد [۲۲]. هم‌چنین این تبدیل باعث هموار نمودن ارتباط خطی موجود بین متغیر مستقل و وابسته می‌شود. با انجام این تبدیل که به صورت شکل‌های باکس–کاکس در شکل ۲ به نمایش گذاشته شده است، فرایند نرمال شدن متغیرها قبل از برازش مدل رگرسیونی مشهود می‌باشد.

[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]

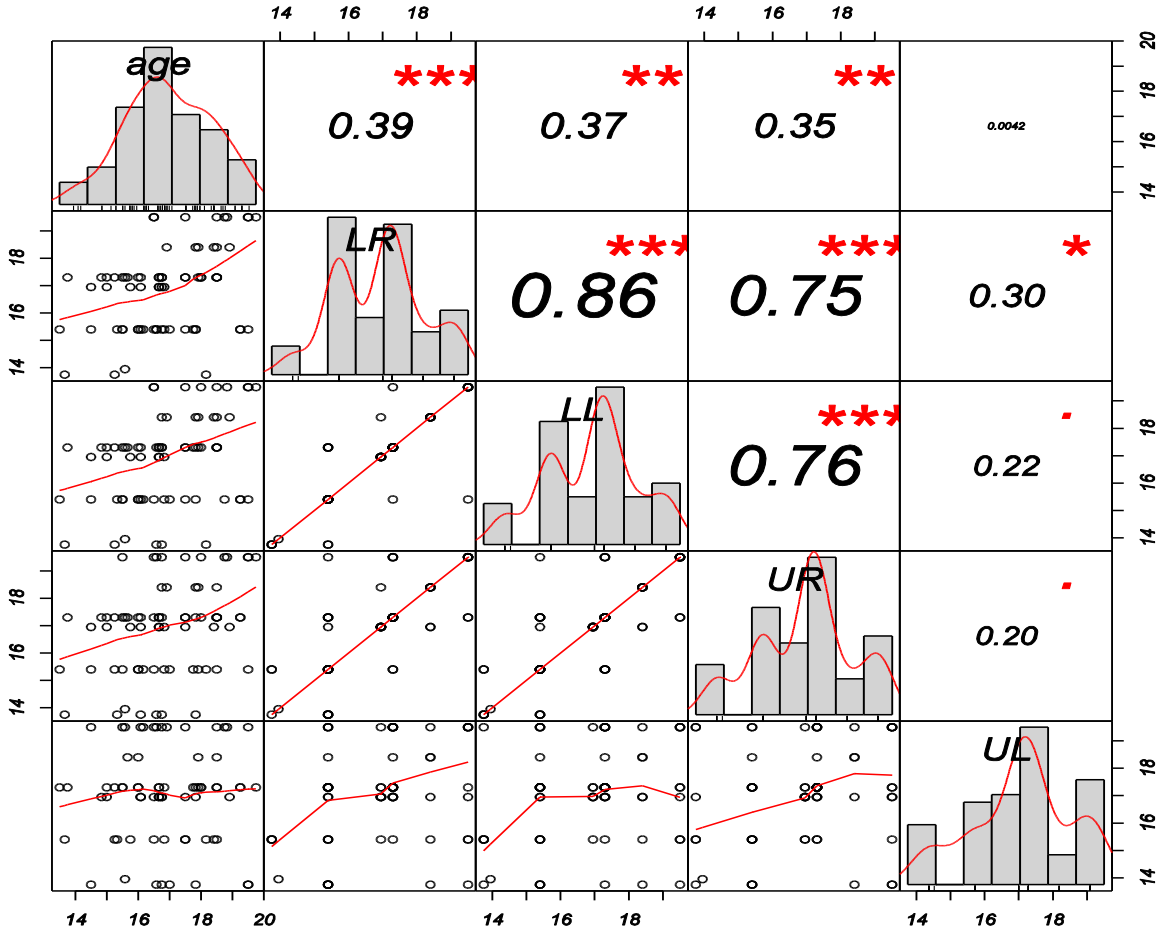
[DOI: 10.29252/koomesh.22.2.228]

برای انجام دادن این تبدیلات با توجه به عملکرد بهتر نرم‌افزار Minitab در این بحث, این تبدیلات در این نرم‌افزار اجرا شد.

در این مرحله برای بررسی وجود داده‌های پرت از رسم شکل باقی‌مانده‌های استانداردشده و با در نظر گرفتن بازه (۲و۲-) استفاده نمودیم. همان‌طور که در شکل ۳ ملاحظه می‌شود، چهار داده پرت در داده‌ها دیده می‌شود و بر این اساس لزوم استفاده از رگرسیون استوار آشکار می‌شود. لذا مدل نهایی به‌صورت زیر خواهد بود: (8)

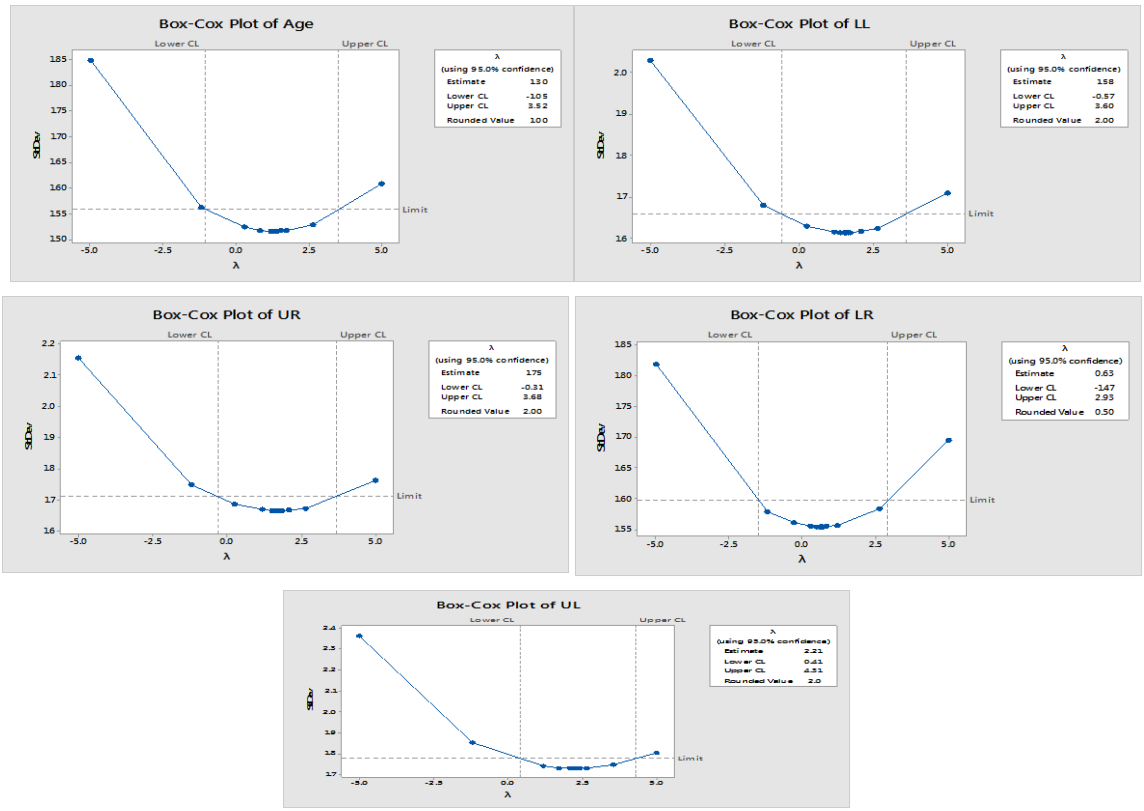
$$Tage = \beta_1 TLR + \beta_2\, TLL + \beta_3\, TUR + \beta_4\, TUL + \epsilon$$

که در آن منظور از Tage ,TUL ,TUR ,TLL ,TLR, متغیر های تبدیل یافته طبق تبدیل باکس-کاکس جدول ۲ می‌باشد. در جدول ۲ چهار روش رگرسیونی به ترتیب کم‌ترین توان‌های دوم، ریج، استوار و ریج استوار مقایسه شده‌اند که با توجه به میزان خطای حاصل (MSE) ملاحظه می‌گردد که روش ریج استوار کم‌ترین خطا را داراست که نشان از برازنده‌ترین مدل است. به عبارت دیگر می‌توان گفت که روش ریج استوار منجر به دقیق‌ترین پیش‌بینی برای سن تقویمی خواهد شد. بر حسب فرمول (۶) می‌توان با جای‌گذاری ضرایب به بهترین مدل رسید. هم‌چنین در شکل ۴ همان‌طور که ملاحظه می‌شود مقدار بهینه برای پارامتر k در مدل رگرسیونی ریج استوار مقدار ۰/۳۱ می‌باشد. در شکل ۵ به منظور درک شهودی کارایی و دقت بالای روش ریج استوار، نقاط پیش‌بینی شده در مقابل نقاط مشاهده شده برای متغیر وابسته تبدیل یافته برای برآورگرهای پیشنهادی رسم شده‌اند. بدیهی است که هر چه شکل پراکنش رسم شده انحراف کم‌تری از نیم‌ساز ربع اول داشته باشد، مدل مورد استفاده دارای پیش‌بینی دقیق‌تری است. بر این اساس می‌توان فهمید که رویکرد رگرسیونی ریج استوار دارای دقت پیش‌بینی بیش‌تری در مقایسه با سایر روش‌های پیشنهادی است.

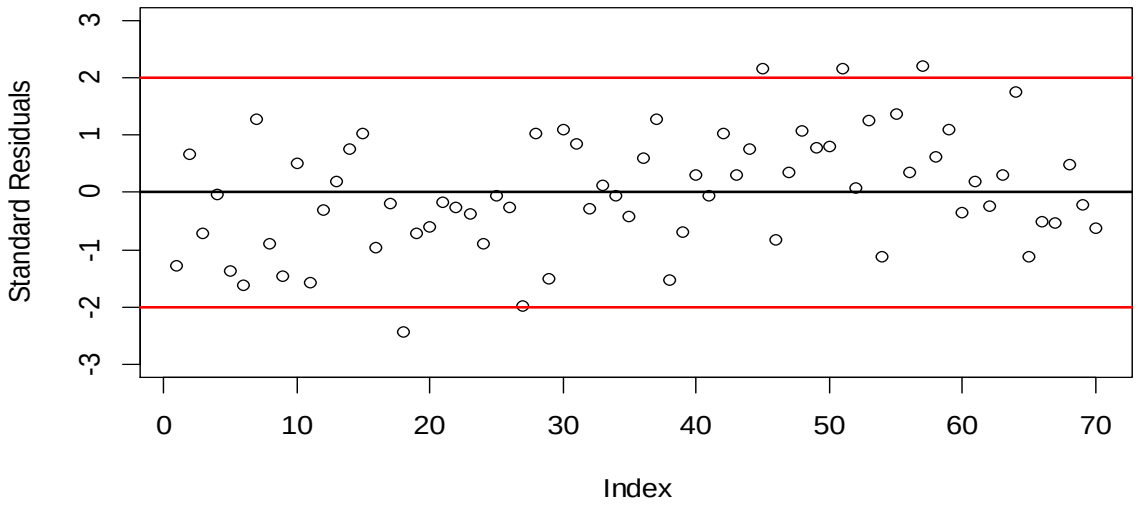


شکل ۱. نگاه کلی به کلیه متغیرهای وابسته و مستقل. به ترتیب بر حسب ستون و ردیف شکل بالا شامل سن تقویمی (age)، که به عنوان متغیر وابسته در نظر گرفته می‌شود و متغیرهای مستقل یا پیش‌بین عبارتند از: سن تخمینی دمیرجین دندان راست مولر فک پایین (LR)، سن تخمینی دمیرجین دندان چپ مولر فک پایین (LL)، سن تخمینی دمیرجین دندان راست مولر فک بالا (UR) ، سن تخمینی دمیرجین دندان چپ مولر فک بالا (UL). لذا مدل رگرسیون اولیه عبارت است از: $age = \beta_1 LR + \beta_2\, LL + \beta_3\, UR + \beta_4\, UL + \epsilon$

[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]
[DOI: 10.29252/koomesh.22.2.228]



شکل ۲. تبدیلات باکس – کاکس بر روی هر



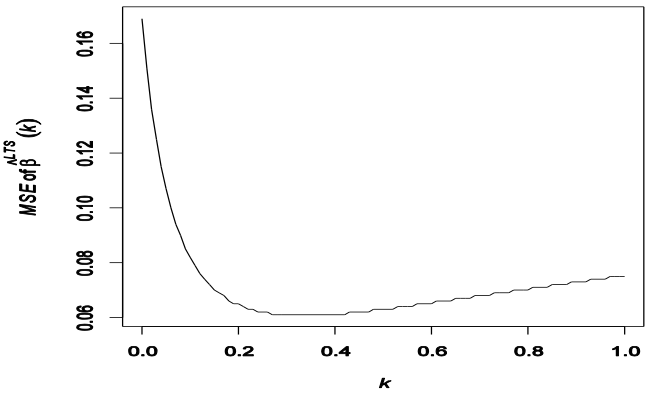
شکل ۳. بررسی وجود داده های پرت

جدول ۲. مقایسه برآوردگرهای پیشنهادی در داده‌های واقعی

	روش کمترین توان‌های دوم	روش ریج	روش استوار	روش ریج استوار
$\hat{\beta}_1$	۰/۲۸۴۷۸	-۰/۱۹۱۴۷	۰/۳۸۴۴۵	-۰/۲۵۳۸۷
$\hat{\beta}_2$	۰/۰۷۱۹۹	۰/۰۹۳۹۱	۰/۱۷۴۰۴	۰/۰۹۶۵۵
$\hat{\beta}_3$	۰/۱۰۴۵۹	-۰/۰۲۰۰۳	-۰/۰۲۹۸۴	-۰/۱۲۰۲۹
$\hat{\beta}_4$	۰/۱۱۳۳۹-	۰/۰۲۸۷۷-	۰/۰۸۸۰۶-	۰/۰۴۷۱۳-

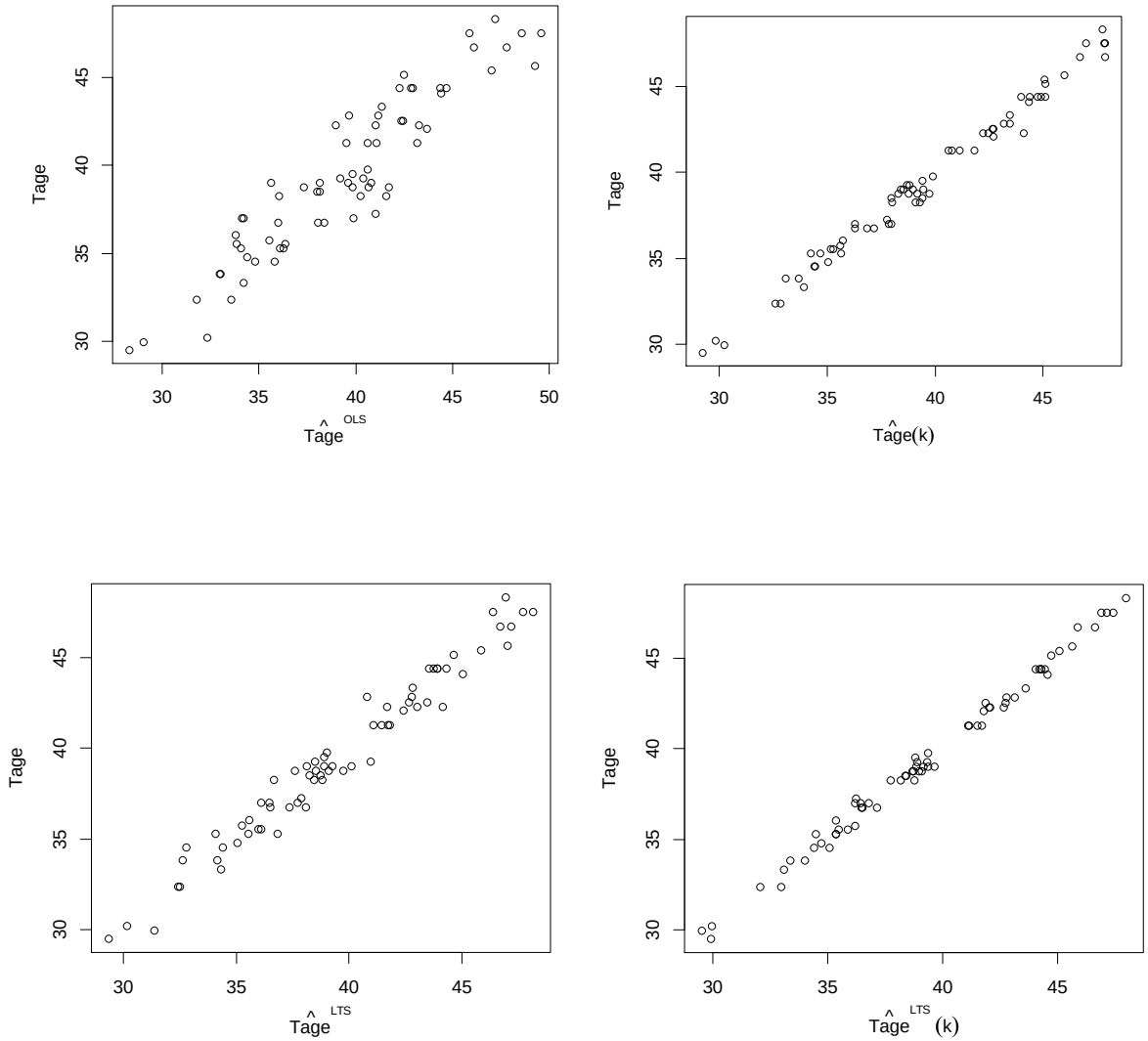
[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]

[DOI: 10.29252/koomesh.22.2.228]



MSE	۱/۱۷۰۹۱	۰/۰۸۱۰۲	۰/۱۶۷۳۱	۰/۰۶۳۰۰

شکل ۴. بررسی مقدار k بهینه با استفاده از کمترین مقدار MSE در مدل رگرسیونی ریج استوار



شکل ۵. شکل پراکنش مقادیر پیش_بینی شده در مقابل متغیر وابسته در روش‌های مورد بررسی

بحث و نتیجه‌گیری

[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]

[DOI: 10.29252/koomesh.22.2.228]

تخمین سن، به کمک روش‌های نوین و غنی آماری نقش مهمی در پزشکی قانونی، بیماری‌های غدد، ارتودنسی، دندان پزشکی اطفال، باستان شناسی و دندان‌پزشکی بالینی دارد. در بین روش‌هایی که برای تعیین سن استفاده می شود، رایج‌ترین روش مورد استفاده در سراسر دنیا، روش دمیرجین است که بر اساس کلسیفیکاسیون دندان دائمی در رادیوگرافی پانورامیک می‌باشد. جوادی‌نژاد و همکاران مطالعه‌ای با هدف تعیین سن دندانی به کمک رادیوگرافی پانورامیک سن تقویمی کودکان شهر اصفهان را به روش دیمیرجین تخمین زدند و به این نتیجه رسیدند که از این روش می‌توان برای تعیین سن در مواردی که نیاز به تعیین سن وجود دارد (از جمله در پزشکی قانونی) استفاده کرد [۱۲]. لازم به ذکر است که در حین این مطالعه علی‌رغم هدف اصلی که برازش مدل رگرسیونی جهت تخمین سن تقویمی بر اساس روش دمیرجین بوده است، به صورت هم‌زمان سن حاصل از روش‌های هاریس (Harris) و مورس (Moorees) نیز تخمین زده شده است. نتایج نشان داد که از بین روش‌های مختلف تعیین سن دندانی، روش دمیرجین بالاترین دقت را در جمعیت مورد مطالعه داشت و کم‌ترین اختلاف را با سن تقویمی نشان داد.

با توجه به کاربرد گسترده مدل‌های رگرسیونی در علوم پزشکی، بررسی دقیق و عمیق داده‌ها و متغیرها حائز اهمیت است زیرا عدم بررسی دقیق داده‌ها منجر به تحلیل و مدل‌بندی نادقیق رگرسیونی می شود. از طرفی روش رگرسیونی کم‌ترین توان‌های دوم با وجود داشتن خواص خوبی از جمله ناریبی و کارایی بالا در شرایط برقراری فرضیات زیربنایی مانند نرمال بودن جملات خطا و عدم وجود همبستگی بین متغیرهای پیشگو، چنان‌چه یکی از فرضیات اصلی بالا نقض گردد به شدت کارایی خود را از دست داده و حتی ممکن است ضرایب مدل را کاملاً اشتباه برآورد کند. به ویژه این‌که این روش نسبت به وجود داده‌های پرت بسیار حساس بوده و با وجود حتی یک داده پرت در مجموعه مشاهدات، دیگر قابل استفاده نبوده و دقت خود را از دست می‌دهد. در چنین شرایطی باید از روش‌های نوین بنام رگرسیون استوار استفاده نمود. هم‌چنین، در صورت وجود هم‌خطی بین متغیرهای پیشگو، رویکرد رگرسیون ریج جایگزین روش رگرسیونی کم‌ترین توان‌های دوم می‌شود.

در این مطالعه داده‌های پرت و هم‌خطی بین متغیرهای مورد بررسی وجود داشت و بنابراین به دلیل کاربردی نبودن روش‌های رگرسیونی کلاسیک، مجبور به استفاده از مدل رگرسیونی ریج استوار شده و در نهایت مدل حاصل سن تقویمی را بر اساس سن تخمینی چهار دندان مولر تخمین زد. این مدل می‌تواند سن تقویمی دقیق‌تری را بر اساس سن تخمینی دمیرجین مولر برای پزشکان و پژوهشگران این امر حاصل نماید. یکی از مواردی که در این موضوع حائز اهمیت است عدم وجود هر یک از دندان‌ها است.

در انتها به‌عنوان پیشنهادات برای تحقیقات آینده می‌توان به استفاده از مدل نیمه‌پارامتری شکنندگی به‌کار رفته توسط حسینی‌فرد و همکاران [۲۳] و مدل شبکه عصبی و رگرسیون کاکس معرفی شده توسط بیگلریان و همکاران [۲۴] برای پیش‌بینی سن تقویمی اشاره نمود.

تشکر و قدردانی

نویسندگان این مقاله برخود لازم می‌دانند از داوران محترم و سردبیر محترم مجله کومش که نظرات سازنده‌ای در ویرایش نهایی این مقاله داشته‌اند، تشکر نمایند.

منابع

[1] McKenna CJ, James H, Taylor JA, Townsend GC. Tooth development standards for South Australia. Aust Dent J 2002; 47: 223-227.

[2] Hegde RJ, Sood PB. Dental maturity as an indicator of chronological age: radiographic evaluation of dental age in 6 to 13 years children of Belgaum using Demirjian methods. J Indian Soc Pedod Prev Dent 2002; 20: 132-138.

[3] Demirjian A, Goldstein H. New systems for dental maturity based on seven and four teeth. Ann Hum Biol 1976; 3: 411-421.

[4] Darji JA, Govekar G, Kalele SD, Hariyani H. Age Estimation from Third Molar Development A Radiological Study. J Indian Acad Forensic Med 2011; 33: 130-134.

[5] Mesotten K, Gunst K, Carbonez A, Willems G. Dental age estimation and third molars: a preliminary study. Forensic Sci Int 2002; 129: 110-115.

[6] Ezoddini Ardakani F, Navab Aazam A, Bashardoost N, Mansoorian H, Ahmadieh MH, Sadat Hosseni SA. Correlation between chronological, skeletal, and dental age on panoramic radiography in patients referred to Yazd dental clinics on 2004-05. J Dental School Shahid Beheshti Univ Med Sci 2007; 24: 474-484. (Persian).

[7] Javadi Nejad S, Ghodousi A, Baharlouei M. Accuracy of age estimation from orthopantomograph using Demirjian's method. IJFM 2008; 14: 137-142.

[8] Arany S, Iino M, Yoshioka N. Radiographic survey of third molar development in relation to chronological age among Japanese juveniles. J Forensic Sci 2004; 49: 534-538.

[9] Schmeling A, Olze A, Pynn BR, Kraul V, Schulz R, Heinecke A, et al. Dental age estimation based on third molar eruption in First Nation people of Canada. J Forensic Odontostomatol 2010; 28: 32-38.

[10] Thevissen PW, Kaur J, Willems G. Human age estimation combining third molar and skeletal development. Int J Legal Med 2012; 126: 285-292.

[11] Liversidge HM. The assessment and interpretation of Demirjian, Goldstein and Tanner's dental maturity. Ann Hum Biol 2012; 39: 412-431.

[12] Javadi Nejad S, Ghafari R, Memar Ardestani M. Accuracy of two methods of age estimation in children based on radiography of developing teeth. IJFM 2011; 17: 141-147. (Persian).

[13] Naseh R, Padisar P, Rahmani J. Correlation between dental and chronological age in 6-15 years old orthodontic patients in Qazvin. J Qazvin Univ Med Sci 2011; 15: 1-18. (Persian).

[14] Edgewth FY. One observation relating to several quantities. J Hermathena 1887; 6: 279-228.

[15] Roozbe M, Chachi J. semiparametric robust estimation. J Statistic Sci 2015; 1: 95-112.

[16] Bagheri H, Rasekh A, Akhund M. Detecting outlier in Liu regression model. J Statistic Sci 2013; 8: 19-36.

[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]

[DOI: 10.29252/koomesh.22.2.228]

[17] Huber PJ. Robust Statistics, Harvard University Cambridge, Massachusetts, 1981.

[18] Amini M, Roozbeh M. Optimal partial ridge estimation in restricted semiparametric regression models. J Multivariate Anal 2015; 136: 26-40.

[19] Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. Thechnometrics 1970; 12: 69-82.

[20] Hoerl AE, Kennard RW. Ridge regression: Some simulation. Comm. Statist. Simulation Comput 1975; 4: 105-123.

[21] Roozbeh M. Robust ridge estimator in restricted semiparametric regression models. J Multivariate Anal 2016; 147: 127-144.

[22] Box GE, Cox DR. An Analysis of Transformed Data. J Royal Statistical Society 1964; 39: 211-252.

[23] Hosseinifard H, Baghestani AR, Jafarian M, Bayat M, Shamszadeh S, Akbarzadeh Baghban AR. Comparison of parametric and semi-parametric methods for estimation of the parameters in frailty models in order to investigation effective factors in survival of the dental implants placement. Koomesh 2017; 19: 371-379. (Persian)

[24] Biglarian A, Hajizadeh E, Anoshirvan Kazemnejad A. Comparison of artificial neural network and Cox regression models in survival prediction of gastric cancer patients. Koomesh 2010; 11: 205-210. (Persian).

[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]

[DOI: 10.29252/koomesh.22.2.228]

Volume 22, Issue 2 (Spring 2020), 228- 236

Prediction of chronological age based on Demirjian dental age using robust ridge regression method

Mahdi Roozbeh (Ph.D)[*1], Seyed Mohammad Malekjafarian (Ph.D)[2], Monireh Manavi (M.Sc)[1], Malihe Sadat Malekjafarian (M.Sc) [3]

1 - Faculty of of Mathematics, Statistics and Computer Sciences, Semnan University, Semnan, Iran
2 - Shahid Sadoughi University of Medical Sciences, Yazd, Iran
3- Facing the Governor of Semnan Province City Water and Wastewater Company, Semnan, Iran

* Corresponding author. +98 2331533206 mahdi.roozbeh@semnan.ac.ir
Received:3 Jan 2019; Accepted: 25 Sep 2019

Introduction: Estimation of age has an important role in legal medicine, endocrine diseases and clinical dentistry. Correspondingly, evaluation of dental development stages is more valuable than tooth erosion. In this research, the modeling of calendar age has been done using new and rich statistical methods. Considerably, it can be considering as a practicable method in medical science that is a combination of some new statistical methods.

Materials and Methods: Among the methods used to determine age, the most commonly used method in the world is the modern modified Demirjian's method based on the calcification of the permanent tooth in panoramic radiography. The study population is consisted of 87 patients who referred to Khatam-ol-Anbia Clinic of Yazd in a simple randomized method during the 12 months of the 2014-2015 year. Using the estimated age of third molar tooth and gender variables, we evaluated the calendar age. In the analysis of regression issues and especially the statistical modeling of many data such as economic data, psychology, social sciences, medical sciences, engineering, etc., we faced with the problem of collinearity among the predictor variables and the presence of remote areas in the data set. The least squares error method in estimation of the parameters of regression model was very sensitive to the outliers. Most of the existing methods for estimating the parameters of these models based on the least squares error approach, affected by the outliers, were yielded to inappropriate estimates, unexpected and high error rates. Robust methods were used to overcome the problem of the outlier observations. It is also recommended the ridge regression to fix the multicollinearity problem. Therefore, in this study, the robust ridge regression estimators will be introduced in the modeling of dependent variables that are less sensitive to the outliers.

Results: The mean age of the subjects was 17.21 years old, with a gender difference of 67% female and 33% male. Additionally, in the relationship between the estimated age of 4 teeth lower right (LR), lower left (LL), upper right (UR), upper left (UL) with a correlation coefficient were above 70%. Correlation between upper and lower jaw teeth was about 30% and between the left and right teeth was 60%. The reason of using robust ridge regression model in this study is the existence of outlier data and collinearity between independent variables.

Conclusion: The necessity of using advanced statistical methods in medical sciences in the recent research is very important. In order to choose the best model, we need to study the data carefully. In this research, the fitted model for prediction of age based on the robust ridge regression method was more efficient with respect to the other methods.

Keywords: Chronological Age, Collinearity, Demirjian Dental Age, Outlier Data, Robust Regression Analysis, Ridge Estimation, Third molar Tooth.

[Downloaded from koomeshjournal.semums.ac.ir on 2024-08-05]

[DOI: 10.29252/koomesh.22.2.228]