



ORIGINAL ARTICLE

Received:2015/01/20

Accepted:2015/03/9

Study the Effect of Different Doses of Vitamin D Supplementation in Insulin Resistance During Women's Pregnancy with Missing Data

Hossein Fallahzadeh (Ph.D.)¹, Farzaneh Saberi (MS.c.)², Azadeh Najarzadeh(Ph.D.)³, Zahra Saberi(Ph.D.)⁴

1.Professor, Research center of prevention and epidemiology of non-communicable disease,School of Public Health,Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

2.Corresponding Author:M.Sc. Student in Biostatistics, School of Public Health,ShahidSadoughi University of Medical Sciences, Yazd, Iran. Email: Farzanehsaberi12@gmail.com Tel: 09131707202

3.Assistant Professor,Department of Nutrition,School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

4.Assistant Professor, Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, , Iran

Abstract

Introduction: The aim of this study was to impute missing data and to compare the effect of different doses of vitamin D supplementation on insulin resistance during pregnancy.

Methods: A clinical trial study was done on 104 women with diabetes and gestational age less than 12 weeks between 1391 and 1393. These subjects were randomly divided into three groups; pregnant women who received daily 200 IU vitamin D (group A), women where receiving monthly 50,000 IU vitamin D (Group B) and Group C are women who received 50,000 IU vitamins D every two weeks. In order to investigate the effect of missing data, the data were studied in two ways, with and without considering missing data. To analyze data in the presence of missing observations, the mechanism of MCAR is considered as the missing mechanism. Then, in order to impute the missing data, four methods including mean imputation, random overall hot-deck imputation, within-class random hot-deck imputation and nearest neighbor imputation was used.

Results: In this study, in random overall hot-deck imputation, the difference between blood sugar and insulin resistance variables are not normal, so median and their interquartile range were reported in the table. Furthermore, kruskal-wallis test was used to compare 3 groups variables. The difference insulin resistance variable was not normal in the nearest neighbor imputation method, so the median and interquartile range was reported in the table. In addition, the kruskal-wallis test was used to compare 3 groups of data. The delta index was calculated for all imputation methods.

Conclusion: In this study, delta index was calculated to evaluate and to compare imputation methods. The random overall hot-deck imputation was described as the best imputation method.

Keywords: Complete, Complete Imputation, Hot – deck Imputation, Random Missing, Non-Random Missing.

Conflict of interest: The authors declared that there is no Conflict interests.



This Paper Should be Cited as:

Hossein Fallahzadeh, Farzaneh Saberi, Zahra Saberi, Azadeh Najarzadeh. Study the Effect of Different Doses of Vitamin D Supplementation in Insulin Resistance Tolooebehdasht Journal .2017; 16(5): 65-76. [Persian]

**تحلیل مشاهدات گمشده در مطالعه اثر دوزهای مختلف مکمل ویتامین D بر مقاومت به****انسولین در دوران بارداری****نویسندگان: حسین فلاح زاده^۱، فرزانه صابری^۲، آزاده نجارزاده^۳، زهرا صابری^۴**

۱. استاد مرکز تحقیقات پیشگیری و اپیدمیولوژی بیماری های غیر واگیر، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد، یزد، ایران.

۲. نویسنده مسئول: دانشجوی کارشناس ارشد آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد، یزد، ایران. تلفن تماس: ۰۹۱۳۱۷۰۷۲۰۲

Farzanehsaberi12@gmail.com

۳. استادیار گروه تغذیه، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد، یزد، ایران.

۴. استادیار گروه آمار، دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان، اصفهان، ایران.

چکیده

مقدمه: مطالعه حاضر با هدف جانمایی داده های گمشده و همچنین مقایسه اثر دوزهای مختلف مکمل ویتامین D بر مقاومت به انسولین در دوران بارداری انجام شد.

روش بررسی: یک مطالعه کارآزمایی بالینی بر روی ۱۰۴ زن باردار مبتلا به دیابت با سن بارداری کمتر از ۱۲ هفته طی سالهای ۱۳۹۱ الی ۱۳۹۳ انجام شد. این افراد به طور تصادفی به سه گروه تقسیم شده اند. زنان بارداری که روزانه ۲۰۰ IU ویتامین D دریافت کرده اند (گروه A)، زنانی که ماهانه IU ۵۰۰۰۰ ویتامین D دریافت کرده اند (گروه B) و گروه C، زنانی هستند که هر دو هفته یکبار IU ۵۰۰۰۰ ویتامین D دریافت کرده اند. در روش تجزیه و تحلیل داده های مطالعه دو حالت یکبار در حضور داده های گمشده و بار دیگر بدون حضور داده های گمشده را مدنظر قرار دادیم. در این مطالعه با در نظر گرفتن مکانیسم MCAR، روش جانمایی، جانمایی میانگین، جانمایی تصادفی کلی بی درنگ، جانمایی تصادفی بی درنگ درون رده ای و نهایتاً "جانمایی نزدیکترین همسایگی بر روی داده های گمشده انجام شد.

یافته ها: در این مطالعه، در روش جانمایی تصادفی کلی بی درنگ، متغیرهای اختلاف قند خون و اختلاف مقاومت به انسولین نرمال نیستند و در جدول میانه و دامنه میان چارکی آنها گزارش شد. همچنین برای مقایسه ۳ گروه برای این متغیرها از آزمون کروسکال والیس استفاده شد. در روش جانمایی نزدیکترین همسایگی متغیر اختلاف مقاومت به انسولین نرمال نشد و برای این متغیر هم به همین ترتیب میانه و دامنه میان چارکی گزارش شد و برای مقایسه ۳ گروه آن از آزمون کروسکال والیس استفاده شد. همچنین شاخص دلنا برای تمامی روش های جانمایی محاسبه شد.

نتیجه گیری: در این مقاله به منظور ارزیابی و مقایسه روش های جانمایی، شاخص دلنا محاسبه شد. روش جانمایی تصادفی کلی بی درنگ به عنوان بهترین روش جانمایی بیان شد.

واژه های کلیدی: جانمایی، جانمایی بی درنگ، گم شدگی کاملاً تصادفی، گم شدگی تصادفی، گم شدگی غیر تصادفی

این مقاله بخشی از پایان نامه دانشجوی کارشناسی ارشد آمارزیستی، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد است.

طلوع بهداشت

دو ماهنامه علمی پژوهشی

دانشکده بهداشت یزد

سال شانزدهم

شماره: پنجم

آذر و دی ۱۳۹۶

شماره مسلسل: ۶۵

تاریخ وصول: ۱۳۹۳/۱۰/۳۰

تاریخ پذیرش: ۱۳۹۳/۱۲/۱۸



مقدمه

ترین روش، حذف داده‌های گمشده از مطالعه و تنها بر اساس

داده‌های کامل تحلیل‌های آماری را انجام دهیم.

دومین روش، جانهی Imputation می‌باشد که در این روش

جانهی مقادیر گمشده به طور مستقیم است. دو روش دیگر شامل

وزن‌دهی و روش‌های مبنی بر تابع درست‌نمایی می‌باشند که در

این روش‌ها هم به جانهی مقادیر گمشده با مقادیری خاص

پرداخته می‌شود ولی جانهی به طور غیرمستقیم انجام می‌شود (۸).

در ارتباط با داده‌های گمشده بیان دو نکته حائز اهمیت می‌باشد.

اولین نکته در رابطه با الگوهای مختلف گمشدگی می‌باشد. لازم

به ذکر است که در اینجا داده‌های کامل را به صورت یک

ماتریس در نظر می‌گیریم که سطرهای آن مربوط به واحدهای

آماری است و ستون‌های این ماتریس مربوط به متغیرهایی که

برای هر واحد اندازه‌گیری می‌شود (۱۵). موقعی که مجموعه داده -

ها کامل است یعنی همه واحدها شناخته شده هستند (۱۴).

فرض کنید $Y = (y_{ij})$ یک ماتریس داده‌های $(n \times k)$ بدون

داده گمشده باشد که سطر i ام آن را $y_i = (y_{i1}, \dots, y_{ik})$

تشکیل می‌دهد و y_{ij} مقدار متغیر j ام برای واحد i ام است. برای

حالتی که با داده‌های گمشده مواجه هستیم یک ماتریس نشانگر

گم شدن داده‌ها $M = (m_{ij})$ تعریف می‌کنیم به طوریکه اگر

y_{ij} گمشده باشد $m_{ij} = 0$ و اگر y_{ij} مشاهده شده باشد

$m_{ij} = 1$ باشد.

در این صورت ماتریس M الگوی گم شدن داده‌ها را تعریف

می‌کند. در آزمایش‌های بالینی ممکن است بعضی افراد به دلایل

نامعلوم در یک زمان خاص از مطالعه خارج شوند و دیگر

بازگردانده نمی‌شوند. به این الگو، الگوی یکنوا (انصراف) گفته

می‌شود به عبارت دیگر می‌توان گفت که اگر y_j گمشده باشد

مقاومت به انسولین یک حالتی از کاهش پاسخ بافت به اثر

انسولین در متابولیسم گلوکز است (۱). غلظت پائین ۲۵

هیدروکسی ویتامین D با دامنه‌ی تغییرات گلوکز و خطر افزایش

دیابت نوع ۲ همبسته می‌باشد، همچنین یک همبستگی بین سطح

ویتامین D و مقاومت به انسولین در زنان باردار شناخته شده

است (۲). در طول بارداری مقاومت به انسولین در افراد مبتلا به

دیابت حاملگی امری شناخته شده است. برخی از مطالعات

رابطه‌ی بین ۲۵ هیدروکسی ویتامین D و مقاومت به انسولین را

نشان داده‌اند (۳،۴). ویتامین D نقش مهمی در ترشح و عملکرد

انسولین دارد (۱،۵). زنان باردار مبتلا به دیابت در مقایسه با زنان

باردار طبیعی بیشتر به کمبود ویتامین D شناخته شده‌اند (۱،۶).

مطالعات طولی و کارآزمایی‌های بالینی از مهم‌ترین مطالعات در

تحقیقات علوم پزشکی به شمار می‌آید (۷). در مطالعات طولی،

در طول زمان هر فرد مورد آزمایش قرار می‌گیرد و متغیرهایی به

طور مکرر اندازه‌گیری می‌شود (۸). یکی از مسائلی که در چنین

مطالعاتی با آن مواجه می‌شویم و امری اجتناب‌پذیر است وجود

داده‌های گمشده می‌باشد ولی متأسفانه تاکنون توجه کمی به آن

شده است (۹،۱۰). در واقع می‌توان گفت که داده‌های گمشده

علاوه بر مطالعات بالینی در تمامی تحقیقات پزشکی حضور

دارند که چشم پوشی از آن‌ها می‌تواند باعث اریبی در نتایج

شود (۱۱،۱۲). اریبی، خطای سیستماتیک شناخته شده است و از

برخورد نامناسب با داده‌های گمشده نتیجه می‌شود. محققین باید

تلاش کنند تا اریبی را کاهش دهند و برآوردهای معتبر و مفید

را بدست بیاورند (۱۳). به طور کلی ۴ روش برای برخورد با

داده‌های گمشده در مطالعات طولی وجود دارد. اولین و ساده



$$f(Y, M | \theta, \emptyset) = f(Y | \theta) f(M | Y, \emptyset) \\ = \prod_{i=1}^n f(y_i | \theta) \prod_{i=1}^n f(M_i | y_i, \emptyset)$$

که در آن $f(y_i | \theta)$ چگالی y_i که به پارامتر نامعلوم θ وابسته است و $f(M_i | y_i, \emptyset)$ چگالی یک توزیع برنولی برای متغیر نشانگر M_i با احتمال $P(M_i = 1 | y_i, \emptyset)$ که y_i ممکن است دارای مقدار گمشده باشد (۱۵). جنبه‌ی یک روش کلی و انعطاف‌پذیر برای برخورد با مسائل داده‌های گمشده می‌باشد (۱۵، ۱). جنبه‌ی بی‌درنگ یک روش ناپارامتری که مقادیر گمشده را با پاسخ مشابه جایگزین می‌کند (۱۶) در واقع روش جنبه‌ی بی‌درنگ Hot dec imputation از واحدهای ثبت شده در نمونه برای جایگزین کردن مقادیر استفاده می‌کند. جنبه‌ی بی‌درنگ در عمل رایج است ولی برخلاف شهرت در عمل، تاکنون مباحث تئوری آن گسترش پیدا نکرده است (۱۵). رایج‌ترین شکل جنبه‌ی، جنبه‌ی میانگین Mean imputation و میانه می‌باشد که برای داده‌های غیرنرمال از میانه استفاده می‌شود. در این روش جنبه‌ی، میانگین مجموعه مقادیر ثبت شده برای جایگزین داده‌های گمشده استفاده می‌شود (۱۱، ۱۷). در روش جنبه‌ی تصادفی کلی بی‌درنگ Random overall hot-deck imputation یک فرد به صورت تصادفی از میان تمامی افراد نمونه انتخاب می‌شود و سپس مقدار این فرد به فرد گمشده تخصیص داده می‌شود. در روش جنبه‌ی بی‌درنگ درون رده‌ای Within-class random hot-deck imputation فرد اعطاکننده به صورت تصادفی از همان طبقه‌ای که فرد گمشده قرار دارد انتخاب می‌شود. مجموعه‌ی تمامی واحدهای مشاهده‌ی کاملی که برای این نوع جایگذاری استفاده می‌شوند را عموماً "مجموعه‌ی اعطاکننده می‌نامند (۱۸، ۱۹). جنبه‌ی

تمامی متغیرها $y_1, \dots, y_{j+1}, \dots, y_k$ نیز برای $j = 1, 2, \dots, k - 1$ گمشده باشند. در مقابل اگر اطلاعات مربوط به حداقل یکی از افراد به طور غیر یکنواخت ثبت شده باشد در این حالت مقادیر گمشده دارای الگوی عمومی (غیر یکنواخت یا متناوب) است. دومین نکته مربوط به مکانیسم‌های مختلف گم شدن داده‌ها می‌باشد. ماتریس داده‌های کامل را به صورت $Y = (y_{ij})$ و ماتریس نشانگر گم شدن داده‌ها $M = (m_{ij})$ را داریم. مکانیسم گم شدن داده‌ها به وسیله‌ی توزیع شرطی M به شرط y ، یعنی $f(M | y, \emptyset)$ مشخص می‌شود که پارامترهای نامعلوم مکانیسم گم شدن می‌باشد. اگر گم‌شدگی به مقادیر y مشاهده شده یا گمشده بستگی نداشته باشد یعنی اگر

$$\forall y \text{ and } \emptyset ; f(M | y, \emptyset) = f(M | \emptyset)$$

گم‌شدگی داده‌ها را کاملاً "تصادفی (Missing Completely at Random) که به روش (MCAR) معروف است. اگر گم‌شدگی فقط به اجزای y_{obs} از y وابسته نباشد یعنی اگر

$$\forall y_{mis} \text{ and } \emptyset ; f(M | y, \emptyset) = f(M | y_{obs}, \emptyset)$$

در این صورت این مکانیسم گم شدن داده‌ها را گم‌شدگی تصادفی Missing at Random که به روش معروف است (۸، ۱۳). در نهایت اگر گم‌شدگی حداقل به یکی از مقادیر گمشده وابسته باشد مکانیسم گم‌شدگی غیر تصادفی Missing Not at Random که به روش (MNAR) معروف است (۸). در این حالت فرض کنید که توزیع توأم (M_i, Y_i) را مستقل از واحدها در نظر بگیریم آنگاه احتمال اینکه یک واحد که مشاهده شده به مقادیر y یا M سایر واحدها بستگی نداشته باشد در این صورت داریم:



فراوانی X بعد از جانمایی باشد. به این ترتیب فاصله بالا که همان شاخص دلتا می‌باشد را تعریف می‌کنیم، که در آن جمع روی دسته‌های X است. در رابطه‌ی بالا ماکزیمم مقداری که مجموع در بالا می‌تواند داشته باشد مقدار ۲ است (۲۱).

روش بررسی

یک مطالعه کارآزمایی بالینی بر روی ۱۰۴ زن باردار مبتلا به دیابت با سن بارداری کمتر از ۱۲ هفته طی سال‌های ۱۳۸۸ تا ۱۳۹۰ انجام شده است. این افراد به طور تصادفی به ۳ گروه تقسیم شده‌اند. زنان بارداری که روزانه ۲۰۰ IU ویتامین D دریافت کرده‌اند (گروه A)، زنانی که ماهانه ۵۰۰۰۰ IU ویتامین D دریافت کرده‌اند (گروه B) و گروه C، گروهی هستند که ۲ هفته‌ای ۵۰۰۰۰ IU ویتامین D دریافت کرده‌اند. لازم به ذکر است که مصرف ویتامین D برای این ۳ گروه از دوازدهمین هفته‌ی بارداری تا زمان وضع حمل می‌باشد. در این مطالعه‌ی بالینی تصادفی شده ۴۰ نفر در گروه C، ۳۸ نفر در گروه B و نهایتاً ۲۶ نفر در گروه A (گروه کنترل) شرکت کرده‌اند. قابل توجه است که بگوئیم این افراد کسانی هستند که مشارکت را کامل کرده‌اند و افرادی که به دلایلی نامعلوم از مطالعه کناره‌گیری کرده‌اند از مطالعه حذف شده‌اند. سطح سرم گلوکز خون (FBS)، انسولین و ۲۵-هیدروکسی ویتامین D قبل و در پایان مداخله اندازه‌گیری شده است. مدل homestic را برای ارزیابی مقاومت به انسولین (HOMA-IR) استفاده شد. داده‌ها شامل ۳ متغیر مستقل و ۴ متغیر وابسته می‌باشد. متغیرهای مستقل سن، وزن و BMI و متغیرهای وابسته اختلاف سطح ۲۵-هیدروکسی ویتامین D قبل و بعد از مداخله، اختلاف FBS قبل و بعد از مداخله، اختلاف انسولین قبل و بعد از مداخله و نهایتاً

نزدیکترین همسایگی Nearest neighbor imputation که به روش (NNI) معروف است، یکی از روش‌های Hot deck که برای جبران ناپاسخی در نمونه‌ی مورد بررسی استفاده می‌شود. اگر چه کاربرد این روش قدمت طولانی دارد ولی تا قبل از سال ۲۰۰۰ میلادی تعداد کمی از ویژگی‌های تئوری این روش شناخته شده است. جانمایی NNI را در ساده‌ترین حالت شرح می‌دهیم. فرض کنید یک نمونه دومتغیره $(x_1, y_1) \dots (x_n, y_n)$ داریم و فرض کنید که r تا از n تا مقادیر y مشاهده شده‌اند (پاسخگر) و بقیه‌ی مقادیر y ($m=n-r$) گمشده هستند (ناپاسخگر) و همه‌ی مقادیر x مشاهده شده‌اند. برای سهولت فرض می‌کنیم که $y_{r+1} \dots y_n$ گمشده باشند. روش NNI یک گمشده y_j را که $r+1 \leq j \leq n$ توسط y_i که $1 \leq i \leq r$ و i نزدیکترین همسایگی j که توسط متغیر x اندازه گرفته شده است یعنی:

$$|x_i - x_j| = \min_{1 \leq l \leq r} |x_l - x_j|$$

اگر مقادیر x وجود داشته باشند، در اینصورت نزدیکترین همسایگی چندگانه‌ی j و i به طور تصادفی انتخاب شده است، بنابراین اگر x متغیری طبقه‌بندی باشد در اینصورت روش NNI ناپاسخگرها را در طبقه‌ی x توسط یک نمونه‌ی تصادفی از پاسخگرها در طبقه‌ی یکسان جانمایی می‌کند (۲۰).

معیار مقایسه روش‌های جانمایی یک شاخص توصیفی به نام شاخص دلتا می‌باشد که فرمول آن به صورت زیر می‌باشد:

$$\Delta = \frac{1}{2} \sum_x |f_x - \tilde{f}_x|$$

فرض کنید متغیر X را با n^* مورد گمشده در n واحد در نظر می‌گیریم. فرض کنید f_x فراوانی نسبی مربوط به رده‌ی x از X ، در مجموعه داده‌هایی با n^* مورد گمشده و \tilde{f}_x مربوط به



توصیف داده‌ها با میانگین و انحراف معیار انجام شد. در جدول ۱ نتایج این روش آورده شده است.

در آزمون ANOVA اگر P Value منجر به معنی داری شود نشان می‌دهد که حداقل بین یکی از دو گروه مورد مقایسه اختلاف معنی داری وجود دارد. در این صورت می‌توان با استفاده از آزمون‌های Post hoc تعیین کرد که کدام دو گروه با هم اختلاف معنی دار داشته‌اند. آزمون post hoc تعیین کرد که کدام دو گروه با هم اختلاف معنی دار داشته‌اند. آزمون Post hoc که برای متغیر اختلاف مقاومت به انسولین انجام دادیم نشان داد که گروه A با گروه B و گروه C با گروه B اختلاف معنی دار داشته‌اند.

در این قسمت داده‌های گمشده را با روش تصادفی کلی بی‌درنگ جهانی کردیم و توزیع داده‌ها را بدست آوردیم، توزیع دو متغیر اختلاف سطح ۲۵-هیدروکسی ویتامین D و اختلاف انسولین نرمال شد (با استفاده از آزمون کولموگروف اسمیرنوف) از آزمون‌های پارامتریک (t-test زوجی، ANOVA) استفاده کردیم و اختلاف معنی دار داشته‌اند. آزمون Post hoc که برای متغیر اختلاف مقاومت به انسولین انجام دادیم نشان داد که گروه A با گروه B و گروه C با گروه B اختلاف معنی دار داشته‌اند.

توصیف داده‌ها با میانگین و انحراف معیار انجام شد. برای دو متغیر اختلاف قندخون و اختلاف مقاومت به انسولین توزیع داده‌ها نرمال نشد بنابراین از آزمون‌های ناپارامتریک (ویلکاکسون و

اختلاف مقاومت به انسولین قبل و بعد از مداخله می‌باشند. برای تجزیه و تحلیل داده‌ها از دو نرم‌افزار SPSS و R (ورژن ۳.۱.۲) استفاده می‌کنیم. در روش تجزیه و تحلیل داده‌های مطالعه دو حالت یکبار در حضور داده‌های گمشده و بار دیگر بدون حضور داده‌های گمشده را مدنظر قرار دادیم. در این مطالعه ۴ روش جهانی، جهانی میانگین، جهانی تصادفی کلی بی‌درنگ، جهانی تصادفی بی‌درنگ درون رده‌یو نهایتاً "جهانی نزدیکترین همسایگی بر روی داده‌ها انجام شده است. لازم به ذکر است که از پکیج Hot deck imputation نیز برای تجزیه و تحلیل استفاده شده است. همچنین مکانیسم MCAR برای داده‌های گمشده در نظر گرفته شده است. درصد موارد گمشدگی برای تمام متغیرها ۲۰٪ در نظر گرفته شده است.

یافته‌ها

مطالعه حاضر با هدف جهانی داده‌های گمشده در مطالعه اثر دوزهای مختلف مکمل ویتامین D بر مقاومت به انسولین در دوران بارداری انجام شد. یکی از اهدافی که در این مطالعه دنبال می‌شود بررسی اختلاف میان متغیرهای مورد بررسی در گروه‌های مختلف است، بدین منظور لازم است ابتدا پس از جهانی، توزیع‌های متغیرهای مورد بررسی تعیین شود تا راهکار مورد نیاز برای انجام این استنباط‌ها مشخص گردد. پس از جهانی داده‌های گمشده توزیع داده‌ها را بدست آوردیم و نتایج را در جدول-هایی آوردیم. در ابتدا داده‌های گمشده را با میانگین جهانی کردیم و توزیع داده‌ها را بدست آوردیم، چون نرمال شد

(با استفاده از آزمون کولموگروف اسمیرنوف) از آزمون‌های پارامتریک (t-test زوجی، ANOVA) استفاده کردیم و



توزیع ۳ متغیر اختلاف سطح ۲۵-هیدروکسی ویتامین D، اختلاف انسولین و اختلاف قند خون نرمال شد از آزمون های پارامتریک (t-test زوجی، ANOVA) استفاده کردیم و توصیف داده ها با میانگین و انحراف معیار انجام شد. برای متغیر اختلاف مقاومت به انسولین توزیع داده ها نرمال نشد بنابراین از آزمون های ناپارامتریک (ویلکاکسون و کروسکال والیس) استفاده کردیم.

برای توصیف داده ها میانه و دامنه میان چارکی گزارش شد. نتایج این روش در جدول ۴ آورده شده است. به منظور ارزیابی و مقایسه روش های جانهای شاخص دلتا محاسبه شده است که در جدول ۵ مقادیر آن مشاهده می شود. همانطور که در این جدول می بینیم مقادیر شاخص دلتا برای روش جانهای تصادفی کلی بی درنگ برای تمامی متغیرها کوچک تر از بقیه می باشند و به همین دلیل این روش بهترین روش جانهای معرفی می شود.

کروسکال والیس) استفاده کردیم. برای توصیف داده ها میانه و دامنه میان چارکی گزارش شد. در جدول ۲ نتایج این روش آورده شده است.

در این قسمت ابتدا داده های گمشده را با روش تصادفی بی درنگ درون رده ای جانهای کردیم و توزیع داده ها را بدست آوردیم. چون توزیع تمام متغیرها نرمال شد (با استفاده از آزمون کولموگروف اسمیرنوف) از آزمون های پارامتریک (t-test زوجی، ANOVA) استفاده کردیم و توصیف داده ها با میانگین و انحراف معیار انجام شد. در جدول ۳ نتایج این روش آورده شده است.

چون آزمون ANOVA برای متغیر اختلاف قندخون معنی دار شده است آزمون post hoc را انجام دادیم. این آزمون نشان داد که گروه C با گروه B اختلاف معنی دار داشته اند. در این قسمت داده های گمشده را با روش نزدیکترین همسایگی جانهای کردیم و توزیع داده ها را بدست آوردیم،

جدول ۱: اختلاف میانگین در ۳ گروه مداخله با روش جانهای میانگین

| متغیرها | گروه درمانی | 200IU (A) | P Value | 50000IU/m (B) | P Value | 50000IU/w © | P Value | (نتیجه آزمون ANOVA) |
|-----------------------------------|-------------|---------------|---------|---------------|---------|---------------|---------|---------------------|
| اختلاف سطح ۲۵- D هیدروکسی ویتامین | | ۱۱/۲۰ ± ۶۵/۷۸ | <۰/۰۰۱ | ۲۳/۳۷ ± ۱۴/۰۹ | <۰/۰۰۱ | ۱۸/۴۳ ± ۱۰/۸۷ | <۰/۰۰۱ | ۰/۲۲ |
| اختلاف انسولین | | ۵/۶۳ ± ۵/۵۶ | <۰/۰۰۱ | ۴/۷۳ ± ۴/۷۲ | <۰/۰۰۱ | ۵/۳۵ ± ۶/۱۷ | <۰/۰۰۱ | ۰/۸ |
| اختلاف قندخون | | ۳/۲۳ ± ۱۳/۰۶ | ۰/۲۵ | ۱۱/۷۹ ± - 2 | ۰/۱۶ | ۲/۶۲ ± ۹/۵۲ | ۰/۰۹ | ۰/۰۵۸ |
| اختلاف مقاومت به انسولین | | ۱/۲۸ ± ۱/۲۴ | <۰/۰۰۱ | ۰/۹۹ ± ۱/۲۳ | <۰/۰۰۱ | ۱/۱۵ ± ۱/۴۶ | <۰/۰۰۱ | ۰/۰۰ |



جدول ۲: اختلاف میانگین در ۳ گروه مداخله با روش جانهای تصادفی کلی بی درنگ

| P Value (نتیجه آزمون ANOVA) | P Value | 50000IU/w © | P Value | 50000IU/m (B) | P Value | 200IU (A) | گروه درمانی |
|-----------------------------------|---------|----------------|---------|------------------|---------|-----------------|--------------------------------------|
| ۰/۸۸ | <۰/۰۰۱ | ۲۲/۹۸± ۱۲/۴۹ | <۰/۰۰۱ | ۲۱/۹۲± ۱۳/۰۳ | <۰/۰۰۱ | ۲۱/۴۹± ۱۲/۰۳ | اختلاف سطح ۲۵- D هیدروکسی ویتامین |
| ۰/۴۹ | <۰/۰۰۱ | ۶/۱۷± ۶/۳۳ | <۰/۰۰۱ | ۴/۷۲± ۵/۳۰ | <۰/۰۰۱ | ۵/۰۷± ۴/۲۵ | اختلاف انسولین |
| ۰/۴۸ | ۰/۵۱۳ | -۳(۱۳) | ۰/۳۲۷ | ۱(۱۳/۱۵) | ۰/۷۳۵ | ۰(۱۰) | اختلاف قندخون |
| ۰/۴۳ | <۰/۰۰۱ | ۰/۸۵(۱/۹۹) | <۰/۰۰۱ | ۰/۹۳(۱/۲۴) | <۰/۰۰۱ | ۰/۸۴(۱/۴۶) | اختلاف مقاومت به انسولین |

جدول ۳: اختلاف میانگین در ۳ گروه مداخله با روش جانهای تصادفی بی درنگ درون رده‌ای

| P Value (نتیجه آزمون ANOVA) | P Value | 50000IU/w © | P Value | 50000IU/m (B) | P Value | 200IU (A) | گروه درمانی |
|-----------------------------------|---------|----------------|---------|------------------|---------|--------------|---|
| ۰/۱۰۶ | <۰/۰۰۱ | ۱۸/۱۹± ۱۰/۹۸ | <۰/۰۰۱ | ۲۴/۰۹± ۱۴/۵۷ | <۰/۰۰۱ | ۱۸/۹۳± ۱۲/۵۱ | اختلاف سطح ۲۵- هیدروکسی ویتامین D |
| ۰/۹۴ | <۰/۰۰۱ | ۴/۹± ۶/۴۶ | <۰/۰۰۱ | ۴/۵۹± ۵/۱۳ | <۰/۰۰۱ | ۵/۰۷± ۵/۳۲ | اختلاف انسولین |
| ۰/۰۱ | ۰/۰۵۹ | ۳/۳۲± ۱۰/۷۹ | ۰/۰۱۶ | -۴/۴۳± ۱۰/۶۳ | ۰/۴۷ | ۲/۰۴± ۱۳/۳۸ | اختلاف قندخون |
| ۰/۸۱ | <۰/۰۰۱ | ۱/۱۵± ۱/۴۶ | <۰/۰۰۱ | ۰/۹۷± ۱/۲۲ | <۰/۰۰۱ | ۰/۹۸± ۱/۰۷ | اختلاف مقاومت به انسولین |

جدول ۴: اختلاف میانگین در ۳ گروه مداخله با روش جانهای نزدیکترین همسایگی

| P Value (نتیجه آزمون ANOVA) | P Value | 50000IU/w © | P Value | 50000IU/m (B) | P Value | 200IU (A) | گروه درمانی |
|-----------------------------------|---------|----------------|---------|------------------|---------|--------------|---|
| ۰/۱۸ | <۰/۰۰۱ | ۱۷/۹۶± ۱۰/۹۶ | <۰/۰۰۱ | ۲۴/۱۱± ۱۳/۸۳ | <۰/۰۰۱ | ۲۱/۳۳± ۱۲/۳۷ | اختلاف سطح ۲۵- هیدروکسی ویتامین D |
| ۰/۶۶ | ۰/۰۰۱ | ۴/۰۷± ۷/۰۲ | <۰/۰۰۱ | ۴/۱۵± ۵/۰۴ | <۰/۰۰۱ | ۵/۴۱± ۵/۶۵ | اختلاف انسولین |
| ۰/۹۳ | ۰/۰۲۸ | ۴/۲± ۱۱/۶۵ | ۰/۱۹۷ | ۳/۲۶± ۱۵/۰۶ | ۰/۳۰۴ | ۲/۶۳± ۱۱/۹۸ | اختلاف قندخون |
| ۰/۶۳ | <۰/۰۰۱ | ۰/۸۵ (۱/۵۸) | ۰/۰۱۸ | ۰/۸ (۱/۸۳) | ۰/۰۰۳ | ۰/۶۱ (۲/۲۴) | اختلاف مقاومت به انسولین |



جدول ۵: مقادیر شاخص دلتا برای ۴ روش جانهی با مکانیسم گمشدگی MCAR

| روش جانهی | تصادفی کلی بی درنگ | تصادفی بی درنگ درون | میانگین | نزدیکترین همسایگی |
|--------------------------|--------------------|---------------------|---------|-------------------|
| متغیر | | | | |
| اختلاف انسولین | ۰/۰۴۵ | ۰/۰۷ | ۰/۰۶۵ | ۰/۱۰۵ |
| اختلاف مقاومت به انسولین | ۰/۰۳۵ | ۰/۰۵۵ | ۰/۰۴ | ۰/۰۳۵ |
| D اختلاف ویتامین | ۰/۰۶ | ۰/۱۱۵ | ۰/۰۶۵ | ۰/۰۵ |
| اختلاف قند خون | ۰/۰۶ | ۰/۰۶ | ۰/۰۸ | ۰/۰۷ |

بحث و نتیجه گیری

مطالعات طولی و کارآزمایی‌های بالینی از کارآمدترین مطالعات در تحقیقات علوم پزشکی به شمار می‌آیند. به دلیل نیاز به پیگیری و ثبت اطلاعات افراد در طول زمان، وجود داده‌های گمشده در بین مجموعه داده‌ها امری اجتناب ناپذیر است.

داده‌های گمشده در بسیاری موارد نتایج تحلیل‌های آماری را به شدت تحت تأثیر قرار می‌دهند به طوری که این تحلیل‌ها منجر به نتایج نامعتبر و اریب می‌شوند. مشکل وجود داده‌های گمشده در مطالعات طولی و کارآزمایی‌های بالینی، در سال‌های اخیر به شدت مورد توجه قرار گرفته است. در نتیجه استفاده از روش‌های آماری پیشرفته در مطالعات طولی و کارآزمایی‌های بالینی برای حل این مشکل ضروری است. یکی از این روش‌ها جانهی داده‌های گمشده می‌باشد، که در این مقاله ۴ روش جانهی بیان شد.

در مطالعه حاضر با توجه به اهمیت داده‌های گمشده ابتدا مفاهیم اساسی درباره‌ی آن‌ها مطرح گردید و سپس مفاهیم جانهی بیان شد و بعد از آن ۴ روش جانهی انجام شده بیان گردیدند. آزمون‌های T زوج شده، ANOVA و کروسکال

والیس بر روی داده‌های جانهی شده انجام شد. نتایج تحلیل حاضر نشان داد که در روش جانهی تصادفی کلی بی درنگ، متغیرهای اختلاف قند خون و اختلاف مقاومت به انسولین نرمال نیستند و در جدول میانه و دامنه میان چارکی آن‌ها گزارش شد. همچنین برای مقایسه ۳ گروه برای این متغیرها از آزمون کروسکال والیس استفاده شد. در روش جانهی نزدیکترین همسایگی متغیر اختلاف مقاومت به انسولین نرمال نشد و برای این متغیر هم به همین ترتیب میانه و دامنه میان چارکی گزارش شد و برای مقایسه ۳ گروه آن از آزمون کروسکال والیس استفاده شد که مقادیر P-Value آن در جدول موجود می‌باشد. همچنین برای ارزیابی این روش‌ها از یک شاخص توصیفی به نام دلتا استفاده می‌شود. مقادیر دلتا برای متغیرها و روش‌های مختلف جانهی با در نظر گرفتن مکانیسم MCAR برای داده‌های گمشده محاسبه شده است. مقادیر دلتا باید اعدادی بین صفر و یک باشد که در اینجا نیز به همین صورت می‌باشد. همانطور که در جدول ۲ می‌بینیم مقادیر شاخص دلتا برای روش جانهی تصادفی کلی بی درنگ برای تمامی متغیرها کوچک‌تر از



تضاد منافع

نویسندگان این مقاله اعلام می‌دارند که هیچگونه تضاد منافی در این مقاله وجود ندارد.

تقدیر و تشکر

از همکاری و کمک کلیه اساتید و کارمندان دانشکده بهداشت دانشگاه علوم پزشکی شهید صدوقی یزد و افرادی که در جمع آوری داده‌ها همکاری داشتند کمال تشکر و سپاسگزاری را داریم.

بقیه می‌باشند و به همین دلیل این روش به عنوان بهترین روش جهانی انتخاب شده است.

با توجه به اینکه مقادیر گمشده همواره مشاهده نشده باقی خواهند ماند، تمام روش‌های بیان شده برای تحلیل این داده‌ها فرضیاتی غیر قابل اثبات هستند.

بنابراین بهتر است در برخورد با داده‌های گمشده صرفاً به نتایج حاصل از یک روش اتکا نکنیم و چندین روش را انجام دهیم و بهترین روش را انتخاب کنیم.

References

- 1-Mojibian M, Soheilykhah J, Jannati Moghadam M, Shojaoddiny Ardekani A. The effect of different doses of vitamin D supplementation on insulin resistance during pregnancy. Yazd Diabetes Research Center, Jomhuri, Blvd, Yazd, Iran April .2013;29(4): 396-99.
- 2- Pamela R, Welma S, Jane C. Vitamin D supplementation reduces insulin resistance in South Asian women living in New Zealand who are insulin resistant and vitamin D deficient a randomized, placebo-controlled trial .British Journal of Nutrition.2010;103: 549-55.
- 3- Maghbooli Z, Hosseinnezhad A, Rahmani M, Shafaei AR, Larijani B. Relationship between leptin concentration and insulin resistance. Hormone and metabolic research .2007;39(12): 903-7.
- 4-Forouhi NG, Luan J, Cooper A, Boucher BJ, Wareham NJ. Baseline serum 25-hydroxy vitamin D is predictive of future glycemc status and insulin resistance. Diabetes .2008;57(10):2619-26.
- 5- Scragg R, Holdaway I, Singh V, Meltcalf P, Baker J, Dryson E . Serum 25 hydroxy vitamin D3 levels decreased in impaired glucose tolerance and diabetes mellitus. Diabetes Res ClinPract.1995; 27:181-8.
- 6- Soheilykhah S, Mojibian M, Rashidi M, Rahimi-Saghand S, Jafari F. Maternal VitaminD Status in Gestational Diabetes Mellitus. Nutrition in Clinical Practice.2010; 25(5):524.
- 7-Zayeri F, Akbarzadehbaghban A, kazemzadeh M, Ahmadi H, Shahriari M. Using Selection Model for Missing Data in Comparing Intravit real Bevacizumab Injection Alone andin Combination with Triamcinolone Clinical Trial in Age-related Macular Degeneration Patients Scientific Journal of Ilam University of Medical Sciences.2013; 20(4):7-15.



- 8-Zayeri F ,Akbarzadeh Baghban AR , Kazemzadeh M , Yaseri M , Abbasi AM .Different Types of Missing in Longitudinal Data and the Likelihood-base Methods Applied in their Analysis. Scientific Journal of Ilam University of Medical Sciences.2010;20(4):208-22.
- 9-MirMohammadKhani M,HolakoueeNai C,Imputation multip.leapplicationsinmedical and epidemiological research. Journal of Public Health and Institute of Public Health Research.2011;9(4):77-88.
- 10- Roderick J. Little RJA.The Prevention and Treatment of Missing Data in Clinical Trials , The new england journal of medicine.2013 ;13(3):55-60.
- 11- Jason S. Haukoos,Craig D. Newgard,MPH. Advanced Statistics: Missing Data in Clinical Research—Part 1: An Introduction and Conceptual Framework . Society for Academic Emergency Medicine.2007;11(2): 662-68.
- 12-Tabatabai RodSeyed M, AlaviMajd H, KhademMabudi A.Weighted estimating equation analysis Meschede G values of the independent variables in Linear Regression and its Application to Data osteoporosis .2006;7(3,4) : 147-51.
- 13- Little RJA, Rubin DB. Statistical analysis with missing data,2002. 2nd edition. New York: Wiley & Sons
- 14- Alex A, Sergio RM. A conservative feature subset selection algorithm with missing data .2010;3(1):585-90
- 15-Andridge RR,Little RJA.A Review of Hot Deck Imputation for Survey Nonresponse.International Statistical Review.2010;78: 40–64.
- 16- Robin Mitra. Bayesian Methods to Impute Missing Covariates for Causal Inference And Model Selection , Department of Statistical Science,Duke University.2008:4-5
- 17- Geert JMG, Heijden V.Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example.2006; 59:1102-09
- 18-Kalton G ,Kasprzyk D . The treatment of missing survey data , survey methodology.1986; 12:1-16
- 19- Niloofar P, Ganjali M, FaridRohani MR. Performance evaluation of imputation based on Bayesian Networks , The Indian Journal of Statistics. 2013;75 (1): 90-111.



-
- 20- Jiahua C , Jun S , Nearest Neighbor Imputation for Survey Data , Jornal of Official Statistics .2000 ; 16(2): 113-131.
- 21- Marco D, Mauro S, Lucia C, Orietta L , Alessandra P. Bayesian networks for imputation, J. R. Statist Soc.2004; 67(2): 309-22.