مقاله آموزش مداوم
Continuing Education Article

مروری بر داده‌های گم‌شده

الهه کاظمی[1]، *مسعود کریملو[2]، مهدی رهگذر[2]

Missing Data

*Kazemi E. (M.Sc.)[1], *Karimlo M.(Ph.D.)[2], Rahgozar M.(Ph.D.)[2]*

MEJDS, Vol 1, No. 1, Aut 2011&Win 2012, 47-52
مجله مطالعات ناتوانی، دوره ۱، شماره ۱، پاییز و زمستان ۱۳۹۰، ۵۲-۴۷
Received: 2/7/2011
Accepted: 25/7/2011
دریافت: ۱۳۹۰/۴/۱۱
پذیرش: ۱۳۹۰/۵/۳

چکیده

در این مقاله سعی شده است که مفاهیم گم‌شدگی داده به صورت ساده و کاربردی توضیح داده شود.

گم‌شدگی داده در تمامی پژوهش‌های علوم اجتماعی، رفتاری، پزشکی وجود دارد. در آمار ، گم‌شدن داده به وضعیتی گفته می‌شود که بخشی از مجموعه داده‌ها گزارش نشده باشند. گم‌شدگی داده باعث کاهش تطابق جامعه نمونه با جامعه کل شده و می‌تواند منجر به نتیجه‌گیری اشتباه در مورد جمعیت اصلی شود.

گم‌شدگی داده یک اتفاق معمول بوده و بسته به میزان آن، می‌تواند اثر قابل توجهی در نتیجه‌گیری به دست آمده از داده‌ها داشته باشد. تمامی روش‌های برآورد پارامترها بر پایه فرض کامل بودن مجموعه داده‌ها استوار است و تحت برقراری این شرایط منجر به برآوردهایی ناریب می‌شوند؛ و البته با افزایش نسبت گم‌شدگی، مقدار اریبی نیز افزایش خواهد یافت.

برای دهه‌ها، محققین از روش‌های قدیمی استفاده می‌کرده‌اند، این روش‌ها متکی به تصحیح مجموعه داده‌ها با صرف‌نظر کردن از موردهای دارای مقادیر گم‌شده و یا جایگزینی مقادیری تخمینی با مقادیر گم‌شده بودند. متأسفانه اکثر این روش‌ها وابسته به برقراربودن فرض دلایل گم‌شدگی داده و نوع سازوکار گم‌شدگی است؛ و در صورت عدم برقراری این فرض منجر به اریبی نتایج می‌شود.

کلیدواژه‌ها: داده گمشده، سازوکار گم‌شدگی کاملاً تصادفی، سازوکار گم‌شدگی تصادفی، سازوکار گم‌شدگی غیر قابل اغماض، سازوکار گم‌شدگی به علت ذات طرح

Abstract

In this paper, we are presenting the basic concepts of missing data in a very simple but practical approach.

Missing data are ubiquitous throughout the social, behavioral, and medical sciences. In Statistics, missing data occur when no data value is stored for the variable in the current observation. Missing data reduce the representativeness of the sample and can therefore distort inferences about the population.

Missing data are a common occurrence and can have a significant effect on the conclusions that can be drawn from the data. All the methods of parameters estimation are based on the completion of data set assumption and only in this case the result will be a non- biased one, and with the increase of missing proportion, the rate of biased results increase too.

For decades, researchers have relied on a variety of old techniques that attempt to "fix" the data by discarding incomplete cases or by filling in the missing values. Unfortunately, most of these techniques require a relatively strict assumption about the cause of missing data and are prone to substantial bias.

Keywords: Missing Data, Missing Completely at Random, Missing at Random, Non Ignorable, Missing by Natural Design

۱. کارشناس ارشد آمار حیاتی، دانشگاه علوم بهزیستی و توانبخشی؛ ۲. دکترای آمار حیاتی، دانشیار دانشگاه علوم بهزیستی و توانبخشی

***آدرس نویسنده مسئول:** تهران، اوین، بلوار دانشجو، خیابان کودکیار، دانشگاه علوم بهزیستی و توانبخشی، گروه آموزشی آمار حیاتی؛ تلفن: ۲۲۱۸۰۱۴۶؛ رایانامه: mkarimlo@yahoo.com

1. M.Sc. of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; 2. Biostatistician, Associate Professor of University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

**Correspondent Address:* Statistics and Computer Department, University of Social Welfare and Rehabilitation Sciences, Koodakyar Alley, Daneshjoo Blv., Evin, Tehran, Iran; Tel: +98 21 22180 146; E-mail: mkarimlo@yahoo.com

www.SID.ir

مقدمه

علی رغم این که در اکثر مثال‌های کتاب‌های درسی داده‌ها به طور کامل موجود می‌باشند، جمع آوری داده‌ها به طور کامل در تحقیقات عملی امکان پذیر نمی‌باشد. داده گم‌شده یک مشکل عمومی در تحقیقات علوم پزشکی، اپیدمیولوژیکی، توانبخشی، اجتماعی و رفتاری می‌باشد و عملاً تجزیه و تحلیل آماری را به سوی نتایج اریب سوق داده و نهایتاً دستیابی به یک نتیجه‌گیری مفید از داده‌های جمع آوری شده را با مشکل مواجه می‌سازد. با وجود تمامی این مشکلات گم‌شدگی بهتر از جواب اشتباه می‌باشد.

در ادبیات آماری اصطلاحات مختلف و غالباً مترادفی برای این مفهوم وجود دارد. این اصطلاحات عبارتند از مقادیر گم‌شده[1]، داده‌های گم‌شده[2]، داده‌های ناقص[3] و بی پاسخ[4]. گم‌شدگی هم در متغیر پاسخ و هم در متغیرهای مستقل رخ می‌دهد. دو نوع گم‌شدگی (بی پاسخی) داریم , بی پاسخی آزمودنی[5] که تمامی اطلاعات برای یک نمونه گم‌شده باشد (آزمودنی گم‌شده است). بی پاسخی در یک سئوال[6] که برخی از اطلاعات یک نمونه موجود نباشد (گم‌شده باشد).

به علاوه گم‌شدن در متغیرها به صورت‌های مختلفی می‌تواند رخ دهد. به عنوان مثال برخی از افراد شرکت‌کننده در مطالعه از ادامه همکاری انصراف می‌دهند، یا از پاسخ دادن به برخی از سئوالات اجتناب می‌کنند. محققین، تکنسین‌ها، جمع آوری کننده داده‌ها ممکن است اشتباهاتی را انجام دهند. در برخی مطالعات گم‌شدگی ممکن است به علت نوع طرح پژوهشی رخ داده باشد، به طور مثال در مطالعه‌ای برای جمع آوری داده‌های مورد نظر یک نمونه‌گیری دو مرحله‌ای انجام می‌شود، در مرحله اول مقادیر متغیرهایی که اندازه گیری آن‌ها آسان و ارزان است برای تمامی افراد شرکت‌کننده در مطالعه جمع آوری می‌شود و سپس در مرحله دوم مقادیر متغیرهایی که اندازه گیری آن‌ها پر هزینه و پیچیده می‌باشد برای زیر مجموعه‌ای از افراد شرکت‌کننده در مطالعه جمع آوری می‌شود، پس بنابراین جزئیات برای کل افراد شرکت‌کننده در مطالعه موجود نمی‌باشد. در یک مطالعه گذشته نگر به علت نقص مدارک و سوابق ممکن است برخی از اطلاعات در دسترس نباشد. همچنین این احتمال هست که به علت نقص یا ضعف دستگاه و تجهیزات، امکان مشاهده و اندازه گیری وجود نداشته باشد و یا در بعضی از مطالعات نظر سنجی، افرادی قادر به اظهار نظر دقیق نباشند؛ و یا به هر دلیلی داده جمع آوری شده برای برخی از افراد شرکت‌کننده در مطالعه گم شود.

آن چه که در قدم اول در برخورد با داده‌های گم‌شده حائز اهمیت است بازنگری و مشاهده مجدد آزمودنی‌های مورد مطالعه و اقدام به تکمیل مقادیر گم‌شده است (۳-۱).

به طور کلی سه روش در نحوه بررسی داده‌های گم‌شده مورد استفاده قرار می‌گیرد (۴):

۱. روش‌های مبتنی بر واحدهای کامل[7]

۲. روش‌های مبتنی بر جانهی[8]

۳. روش‌های مبتنی بر مدل

هر روشی را که برای تحلیل داده‌های گم‌شده به کار می‌بریم نقاط ضعف و قوت خودش را دارد، که وابسته است به:

۱. نسبت و الگوی گم‌شدگی

۲. مدلی که برای تحلیل پرسش‌های تحقیق انتخاب شده است

۳. تعداد متغیرهای مورد استفاده در تحقیق

۴. نوع متغیرهایی که دارای مقادیر گم‌شده می‌باشند

۵. سازوکار[9] گم‌شدگی

امروزه بسیاری از محققین روش‌هایی را برای مقابله با مشکل داده گم‌شده ارائه کرده‌اند که اکثر این روش‌ها وابسته به نوع سازوکار گم‌شدگی داده‌ها می‌باشند؛ لذا باید نوع سازوکار گم‌شدن معلوم باشد تا بتوان روش تحلیل مناسبی را انتخاب نمود. متاسفانه برخی از محققین بین الگوی گم‌شدگی و سازوکار گم‌شدگی تمایزی قائل نمی‌باشند و این دو تعریف را به جای هم استفاده می‌کنند که اشتباه است. یک الگوی گم‌شدگی داده نشان دهنده ترکیب داده‌های گم‌شده و مشاهده شده در مجموعه داده‌ها می‌باشد، در حالی که سازوکار گم‌شدگی داده‌ها، رابطه ممکن بین مقادیر گم‌شده و مشاهده شده را توضیح می‌دهد (۳).به همین دلیل در ادامه در رابطه با الگوی گم‌شدگی و سازوکار گم‌شدگی توضیحات بیشتری را ارائه می‌دهیم.

1. Missing Values
2. Missing Data
3. Incomplete Data
4. Non Response
5. Unit Non Response
6. Item Non Response
7. Complete Cases ,Complete Records
8. Imputation
9. Mechanism

Archive of SID

www.SID.ir

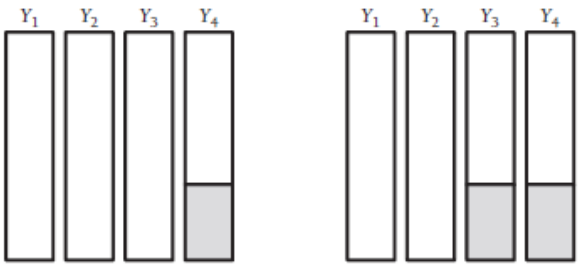
الگوی گم‌شدگی

الگوی گم‌شدگی داده‌ها تنها مکان گم‌شدگی در داده‌ها را به ما نشان می‌دهد. شکل ۱ شش طرح الگوی گم‌شدگی داده‌ها را نشان می‌دهد، نواحی سایه زده شده نشان دهنده مکان مقادیر گم‌شده در مجموعه داده‌ها می‌باشند. در قسمت (الف) شکل ۱, الگوی تک متغیره[1] نشان داده شده است که مقادیر گم‌شده در یک متغیر رخ داده است. در قسمت (ب) شکل ۱ الگوی گم‌شدگی بی پاسخی واحد[2] نشان داده شده است که این نوع الگوی گم‌شدگی اغلب در تحقیقات پیمایشی[3] رخ می‌دهد. الگوی گم‌شدگی یکنواخت[4] در قسمت (ج) شکل ۱ نشان داده شده است که معمولاً در مطالعات طولی[5] رخ می‌دهد، که فرد شرکت‌کننده از مطالعه خارج می‌شود و دیگر به مطالعه بر نمی‌گردد (در برخی متون به آن ساییدگی[6] می‌گویند). الگوی گم‌شدگی داده عمومی (کلی) در قسمت (د) شکل ۱ نشان داده شده است که مقادیر گم‌شده در سرتاسر مجموعه داده‌ها به روش تصادفی پخش شده‌اند. الگوی گم‌شدگی با برنامه[7] در بخش (ه) شکل ۱ نشان داده شده است، که برای جمع آوری تعداد زیاد پرسشنامه‌ها با سئوالات زیاد مورد استفاده قرار می‌گیرد که به طور همزمان هزینه پاسخ‌دهنده‌ها را نیز کاهش می‌دهد. در آخر، الگوی متغیر پنهان[8] در قسمت (ی) شکل ۱ نشان داده شده است که فقط در تحلیل متغیرهای پنهان مانند مدل‌های معادلات ساختاری[9] دیده می‌شود.

از دیدگاه عملی تمایز بین الگوهای گم‌شدگی داده‌ها امروزه دیگر مهم نیست زیرا برآوردهای ماکزیمم درست‌نمایی و جانهی چندگانه برای تمامی الگوهای گم‌شدگی داده مناسب هستند (۳).

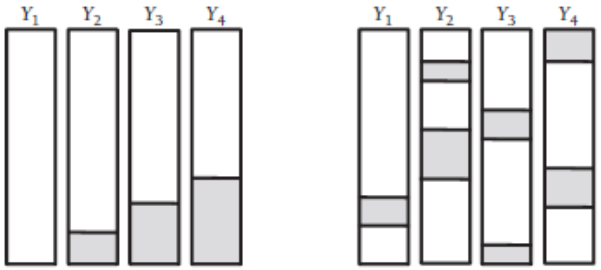
(ب) الگوی بی پاسخی واحد

(الف) الگوی تک متغیره



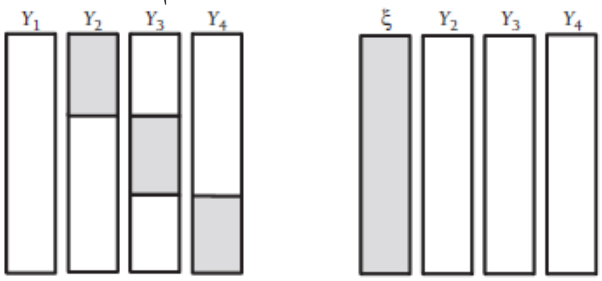
(د) الگوی عمومی (کلی)

(ج) الگوی یکنواخت



(ی) الگوی متغیر پنهان

(ه) الگوی گم‌شدگی با برنامه



شکل ۱- در شکل شش طرح الگوهای گم‌شدگی داده نشان داده شده است. نقاط سایه زده شده نشان دهنده مکان مقادیر گم‌شده در مجموعه داده‌ها با چهار متغیر هستند.

سازوکار گم‌شدگی

سه نوع سازوکار گم‌شدگی داده داریم:

برای توضیح انواع سازوکار گم‌شدگی، مجموعه داده‌ها را به صورت دو مولفه داده‌های کامل و داده‌های گم‌شده (X_{obs}, X_{mis}) در نظر می‌گیریم، همان طور که از نوع نامگذاری مشخص است، X_{obs} متغیری است که تمام مقادیر آن به طور کامل مشاهده شده‌اند و X_{mis} متغیری است که برخی از مقادیر آن گم شده‌اند.

۱. **گم‌شدن کاملاً تصادفی[10] (MCAR):** در این حالت گم‌شدگی در متغیر، X_{mis}، به هیچ یک از متغیرهای دیگر و خود متغیر X_{mis}، بستگی ندارد. بنابراین در این حالت تجزیه و تحلیل براساس داده‌های مشاهده شده، برآوردهای ناریبی از پارامترها (و نیز خطای معیار آن‌ها) نتیجه خواهد داد. گم‌شدگی کاملاً تصادفی تنها سازوکار گم‌شدگی است که تشخیص آن آزمون پذیر می‌باشد (۳). کارشناسان تعدادی از روش‌ها را برای آزمون سازوکار گم‌شدگی کاملاً تصادفی ارائه داده‌اند (۱۲-۵) به عنوان مثال مقایسه‌های آزمون تی تک متغیره[11] (۷) و آزمون گم‌شدگی کاملاً تصادفی لیتل[12] (۹،۱۳).

1. Univariate pattern
2. Unit Non Response Pattern
3. Survey Research
4. Monotone Pattern
5. Longitudinal Study
6. Attrition
7. Planned Missing Data Pattern
8. Latent Variable Pattern
9. Structural Equation models
10. Missing Completely At Random:
11. Univariate T-Test
12. Little's MCAR Test

Archive of SID

www.SID.ir

مقایسه‌های آزمون تی: در این روش مواردی را که مقادیر گم‌شده و مشاهده شده دارند را برای یک متغیر خاص جدا می‌کنیم و از آزمون تی برای مقایسه میانگین‌های این دو گروه بر روی سایر متغیرها استفاده می‌کنیم. سازوکار گم‌شدگی کاملاً تصادفی بیان می‌کند که میانگین موارد مشاهده شده و گم‌شده باید برابر باشند. قبول فرض صفر این آزمون منجر به پذیرش سازوکار کاملاً تصادفی داده‌ها می‌شود، و رد فرض صفر، معنی دار شدن آماره تی یعنی تفاوت زیاد میانگین‌ها، بیان می‌کند که مجموعه داده‌ها ممکن است دارای سازوکار گم‌شدگی تصادفی یا غیر قابل اغماض باشند (۷).

آزمون گم‌شدگی کاملاً تصادفی لیتل: لیتل (۹) یک آزمون تی که برای چند متغیره بسط داده شده است را پیشنهاد کرده که به طور همزمان تفاوت میانگین‌ها را برای تمامی متغیرهای مجموعه داده‌ها ارزیابی می‌کند. مانند آزمون تی یک متغیره، آزمون لیتل نیز تفاوت میانگین‌ها را در بین زیر گروه‌ها مقایسه می‌کند. فرض صفر در این آزمون وجود سازوکار گم‌شدگی کاملاً تصادفی داده‌ها می‌باشد. این آزمون در نرم افزارهای آماری مانند SPSS موجود می‌باشد (۱۳).

۲. **گم‌شدن تصادفی**[1] **(MAR)**: دراین حالت گم‌شدگی در متغیر، X_{mis} بستگی به متغیری که برای همه افراد شرکت‌کننده در مطالعه به طور کامل مشاهده شده است، X_{obs} دارد ولی به خود متغیر، X_{mis} بستگی ندارد. این سازوکار آزمون پذیر نمی‌باشد (۳، ۱۴).

برای تشخیص سازوکار گم‌شدگی تصادفی فلایس و همکاران[2] در کتاب خود (۲) روشی را ارائه کرده‌اند. این روش عمومی نبوده و زمانی کاربرد دارد که مجموعه داده‌ها به گونه‌ای باشد که بتوان از رگرسیون لوجستیک استفاده کرد و همچنین متغیری که دارای مقادیر گم‌شده، X_{mis}، می‌باشد متغیری رسته‌ای و دو حالتی باشد. متغیرهای X_{mis} و X_{obs} متغیر مستقل می‌باشند و متغیر Y را به عنوان متغیر وابسته در نظر می‌گیریم.

شیوه کار به صورت الگوریتم زیر می‌باشد:

۱. متغیر نشانگر گم‌شدگی دوحالتی Δ_i را به صورت زیر تعریف می‌کنیم:

$$\Delta_i = \begin{cases} 1 & \text{اگر } X_{mis(i)} \text{ مشاهده شده باشد} \\ \text{صفر} & \text{اگر } X_{mis(i)} \text{ گم‌شده باشد} \end{cases}$$

۲. مجموعه داده‌های بسط داده شده را به وسیله تکثیر کردن هر مقدار گم‌شده با جایگذاری $X_{mis(i)}=0$ و $X_{mis(i)}=1$ درست می‌کنیم. مجموعه داده بسط داده شده جدید شامل n' واحد می‌باشد. برای داده‌های جدید سایر متغیرها، $X_{obs(i)}$ و Y_i، را به همان صورت که مشاهده شده‌اند تکرار می‌کنیم. در این مجموعه داده مقدار n' از فرمول زیر به دست می‌آید:

$$n' = \sum_{i=1}^{n} \Delta_i + 2\sum_{i=1}^{n}(1-\Delta_i)$$

۳. مقادیر اولیه‌ای را برای αها در نظر می‌گیریم، با استفاده از این مقادیر اولیه برای هر واحد وزن‌های، w_i , را برای $i=1,2,...n'$ باید محاسبه کنیم. برای داده‌های مشاهده شده، $w_i=1$ می‌باشد. برای مقادیر تکثیر شده w_iها به صورت زیر محاسبه می‌شوند:

$$X_{mis(i)}=1 \Rightarrow w_i = P(X_{mis(i)} = 1 \mid Y_i, X_{obs(i)}, \Delta_i=0)$$

$$X_{mis(i)}=0 \Rightarrow w_i = P(X_{mis(i)} = 0 \mid Y_i, X_{obs(i)}, \Delta_i=0)$$

۴. مدل رگرسیون لوجستیک وزنی را برای حالتی که در آن Δ متغیر وابسته و X_{obs} و X_{mis} و Y متغیر مستقل باشند (نام پارامترهای این مدل را αمی‌گذاریم). سپس با استفاده از مجموعه بسط داده شده با n' واحد و وزن‌های W_i برآوردهای به روز شده αها را به دست می‌آوریم.

۵. با استفاده از برآوردهای جدید α، مجدداً وزن‌های واحدها را که در مرحله ۳ بیان شد محاسبه کرده و به مرحله ۴ رفته و برآوردهای جدید را با استفاده از وزن‌های جدید محاسبه می‌کنیم. این مراحل را به قدری تکرار می‌کنیم تا به هم‌گرایی برسیم.

پس از انجام عملیات بالا نتایج به دست آمده را بررسی کرده و در صورتی که برآورد پارامتر α_x (پارامتر ضریب متغیر X_{mis}) تقریباً مقداری نزدیک صفر با واریانسی بسیار کوچک باشد می‌توان گفت که مجموعه داده دارای سازوکار گم‌شدگی تصادفی می‌باشد.

۳. **گم‌شدن غیر قابل اغماض**[3] **(NI)**: در این حالت گم‌شدگی در متغیر کمکی، Y_{mis}، بستگی به خود Y_{mis} دارد، حتی بعد از اینکه روی سایر متغیرها مشروط شده باشد. برآوردهای هیچ یک از پارامترها از روی داده‌های با مشاهدات کامل، سازگار نخواهند بود و هر گونه استنباطی مستلزم در نظر گرفتن مفروضاتی درباره ارتباط بین متغیرهای Y_{obs} و Y_{mis} و نیز چرایی گم‌شدگی داده‌ها می‌باشد. ذکر این نکته حائز اهمیت است که هیچ تجزیه و تحلیلی از داده‌های دارای گم‌شدگی بدون اغماض، بدون انجام آزمون حساسیت کامل نیست (۱۵، ۱۶).

1. Missing Completely At Random 2. Feliess et al

3. Missing Completely At Random:

Archive of SID

www.SID.ir

۴.گم‌شدگی به علت ذات طرح[1] (MBND): گم‌شدگی به علت ذات طرح نوع دیگری از سازوکار گم‌شدگی داده است که مقادیر به علت این که آن‌ها را به طور طبیعی و معمول نمی‌توان اندازه گیری کرد گم‌شده‌اند. این یک نوع سازوکار گم‌شدگی جدیدی است که شامل مشکلاتی در اندازه گیری است (۱۷).

پس از تشخیص صحیح نوع سازوکار گم‌شدگی می‌توان روش مناسبی را انتخاب نمود و به بررسی و تحلیل داده‌های‌مان بپردازیم.

نتیجه‌گیری

بنابراین درصورت مواجه شدن باداده های گم‌شده اولین قدم بازنگری ومشاهده مجدد واحدهای مورد مطالعه و تکمیل مقادیر گم‌شده است. در مرحله بعد می‌بایست با محاسبه احتمالات گم‌شدگی در سطوح مختلف متغیرها، نسبت به تشخیص نوع سازوکار گم‌شدگی اطمینان حاصل نمود، و برای اطمینان بیشتر از آزمون های موجود تشخیص گم‌شدگی نیز استفاده کرد. پس از تشخیص صحیح نوع سازوکار گم‌شدگی می‌توان روش مناسبی را انتخاب نمود و به بررسی و تحلیل داده‌های‌مان بپردازیم.

تقدیر و تشکر

با تشکر از جناب آقای مایونقی چو پیک (Myunghee Cho Paik) که برنامه الگوریتم تشخیص سازوکار گم‌شدگی تصادفی را در اختیار نویسندگان قرار دادند.

Archive of SID

1. Missing by natural design

www.SID.ir

منابع

1. Satten GA, Carroll RJ. Conditional and Unconditional Categorical Regression Models with Missing Covariates. Biometrics. 2000;56:384-388.
2. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. 3rd ed. New York: John Wiley & Sons; 2002.
3. Enders CK. Applied Missing Data Analysis. New York and London: Guilford Press; 2010.
4. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. New York: John Wiley & Sons; 2000.
5. Chen HY, Little RA. Test of missing completely at random for generalised estimating equations with missing data. Biometrika. 1999; 86:1–13.
6. Diggle PJ. Testing for random dropouts in repeated measurement data. Biometrics. 1989; 45:1255–1258.
7. Dixon WJ. BMDP statistical software. Los Angeles: University of California Press; 1988.
8. Kim KH, Bentler PM. Tests of homogeneity of means and covariance matrices for multivariate incomplete data. Psychometrika. 2002; 67:609–624.
9. Little RJA. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc. 1988; 83:1198–1202.
10. Muthén B, Kaplan D, Hollis M. On structural equation modeling with data that are not missing completely at random. Psychometrika. 1987; 52:431–462.
11. Park T, Lee S-Y. A test of missing completely at random for longitudinal data with missing observations. Statistics in Medicine. 1997; 16:1859–1871.
12. Thoemmes F, Enders C K. A structural equation model for testing whether data are missing completely at random. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL 2007 April.
13. SPSS Missing Value Analysis™ 17. United States of America: SPSS Inc 2007. http://www.spss.com
14. Longford NT. Missing data and small area estimation. Springer; 2005.
15. Karimlou M, Jandaghi GR, Mohammad K, Wolfe R, Azam K. A Comparison of Parameter Estimates in Standard Logistic Regressin Using WinBUGS MCMC and MLE Methods in R for Different Sample Size. Far East J Theo Stat. 2006.
16. Ibrahim JG, Chen MH, Lipsitz SR. Bayesian methods for generalized linear models with covariates missing at random. Canadian Journal of Statistics. 2008.
17. Marwala T. Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques. South Africa: University of Witwatersrand IGI Global; 2009.

Archive of SID

www.SID.ir