مدل تاث ر تصادف ا داده‌ها ا پاسخها م خته

علی صادقی[۱] مجتبی گنجعلی[۲]

[۱] کارشناس ارشد مار اقتصاد اجتماع دانشگاه شه د بهشت
[۲] ع و ه ت علم دانشگاه شه د بهشت

چک ده درکار ردهای مختلف مار چند متغ ره در شاخه‌های مختلف علوم ا مدلها ی تاث ر تصادفی مواجها م که متغ ر پاسخ رداری از داده‌های م خته پ وسته و گسسته است در چن ن موارد شناسا ی همبستگی م ان متغ رهای پ وسته و گسسته اهم ت س اری دارد هنگامی که تمامی متغ رهای پاسخ پ وسته اشند می‌توان از تحل ل عاملی استفاده نمود و هنگامی که کل ه متغ رها گسسته اشند استفاده از روشهای چند متغ ره گسسته می‌تواند سودمند اشد در ا ن مقاله ا استفاده از روش تجز ه تک عاملی روشی ارا ه گرد ده است که می‌توان م خته‌ای از متغ رهای پ وسته و گسسته را ا ه دست وردن همبستگی‌ها در مدل اثرات تصادفی مورد ررسی قرار داد در ا ن مدل متغ رهای پ وسته از توز نرمال و متغ رهای گسسته از توز دود ی تبع ت می‌کنند شده‌اند ا استفاده از روش تجز ه تک عاملی علاوه ر دست وردن همبستگی م ان متغ رهای پاسخ تاث ر عوامل مح ی و ژنت کی را ن ز در ن ر می‌گ ر م در ا ن مقاله روشی مبتنی ر استفاده از تاث رهای تصادفی را رای مدل ندی پاسخهای م خته نرمال و دودو ی معرفی و نحوه تحل ل تاث ر رخی متغ رهای تب نی ر ا ن پاسخها را ن ز شرح می‌ده م

واژه‌ها کل د اثرات تصادفی تجز ه عاملی متغ ر پنهان پاسخهای م خته

مقدمه

در س اری از تحق قات ماری ا حالتی مواجه هست م که م خته‌ای از داده‌های پ وسته و گسسته ه عنوان متغ ر پاسخ مد ن ر است و ع ی از متغ رها دارای اثرات تصادفی هستند ا ن مدلها توس ل تل[1] و را ن[2] س ار مورد توجه قرار گرفته است در چن ن حالتی ر خلاف هنگامی که تمامی اثرات ثا ت هستند نمی‌توان از روشهای معمولی مورد استفاده در علم مار مدل مناس را ه داده‌ها رازش داد اگرستی[3] روشها ی که تاکنون رای رازش مدل‌های ا اثرات تصادفی رای داده‌های م خته مورد ررسی قرار گرفته‌اند ه دل ل استفاده از روشهای عددی در رازش مدل مناس از ت وری پ چ ده‌ای استفاده کرده‌اند و س اری از نها تنها در شرا خصوصی ه نت جه مورد ن ر می‌رسند در ا ن مقاله سعی گرد ده است که ا استفاده از ت وری ساده و ا استفاده از روشهای عددی ساده‌تر ا ن حث مورد توجه قرار گ رد

1) Little 2) Rubin 3) Agresti

مدل اثرات تصادف را پاسخها م خته

مانند مدل مورد استفاده سامل و ر ان[۴] ک مدل اثرات تصادفی را ا M رمد مشاهده شده که تعداد $M_۱$ تا از ا ن رمدها دارای توز دودو ی و $M_۲ = M - M_۱$ تا از رمدهای اقی مانده از توز نرمال پ روی می‌کنند در ن ر می‌گ ر م ذکر ا ن نکته روری است که متغ رهای تصادفی در مدل مورد حث مانند اثرات متغ رهای پنهان رفتار می‌کنند کر[۵] رای ا ن من ور مدل اثرات تصادفی را رای زمودنی i ام ه صورت ز ر در ن ر می‌گ ر م

$$
\begin{array}{rclr}
y^*_{im} & = & X'_{im}\beta + \varepsilon_{im} \quad m = \ ,\ldots,M_۱ & (\ ) \\
z_{im} & = & U'_{im}\beta + \varepsilon_{im} \quad m = M_۱ + \ ,\ldots,M &
\end{array}
$$

که در ن X_{im} و U_{im} ماتر س متغ رهای تب ی رای زمودنی i ام در رمد m ام است در ا ن مدل y_{im} مقدار مشاهده شده ک متغ ر پنهان y^*_{im} ا توز نرمال استاندارد است که فق کوچکتر ا زرگتر از عدد ودن نرا در دست دار م و اگر $y^*_{im} >$ نگاه $y_{im} =$ و در غ ر ا ن صورت $y_{im} =$ خواهد ود
ماتر س وار انس کوار انس مدل فوق دارای پارامترهای ز ادی حتی ا در ن رگرفتن مقاد ر کوچک M خواهد ود رای درک هتر ا ن مس له حالتی را در ن ر می‌گ ر م که $M_۱ =$ و $M_۲ =$ است در ا ن صورت مدل ه صورت ز ر ان می‌شود

$$
\begin{array}{rclr}
y^*_{im} & = & X'_{im}\beta + \varepsilon_{im} \quad m = \ ,\ ,\ , & (\ ) \\
z_{im} & = & U'_{im}\beta + \varepsilon_{im} \quad m = \ , &
\end{array}
$$

در ا ن صورت ساختار ماتر س وار انس کوار انس غ رساختاری ه صورت ز ر در خواهد مد

$$
\Sigma = \begin{pmatrix}
 & \rho_{۱۲} & \rho_{۱۳} & \rho_{۱۴} & \sigma_۱\rho_{۱۵} & \sigma_۲\rho_{۱۶} \\
\rho_{۱۲} & & \rho_{۲۳} & \rho_{۲۴} & \sigma_۱\rho_{۲۵} & \sigma_۲\rho_{۲۶} \\
\rho_{۱۳} & \rho_{۲۳} & & \rho_{۳۴} & \sigma_۱\rho_{۳۵} & \sigma_۲\rho_{۳۶} \\
\rho_{۱۴} & \rho_{۲۴} & \rho_{۳۴} & & \sigma_۱\rho_{۴۵} & \sigma_۲\rho_{۴۶} \\
\sigma_۱\rho_{۱۵} & \sigma_۱\rho_{۲۵} & \sigma_۱\rho_{۳۵} & \sigma_۱\rho_{۴۵} & \sigma_۱^۲ & \sigma_۱\sigma_۲\rho_{۵۶} \\
\sigma_۲\rho_{۱۶} & \sigma_۲\rho_{۲۶} & \sigma_۲\rho_{۳۶} & \sigma_۲\rho_{۴۶} & \sigma_۱\sigma_۲\rho_{۵۶} & \sigma_۲^۲
\end{pmatrix}
$$

ا وجود اثرات تصادفی در مدل نمی‌توان از روشهای معمول در مار ه من ور رازش مدل ا وجود م خته‌ای از رمدهای پ وسته و گسسته استفاده کرد رای دست وردن پارامترها از روشهای عددی استفاده می‌گردد اما کی از مشکلات اصلی استفاده از روشهای عددی ن است که هنگاهی که تعداد کمی داده در اخت ار دار م و مدل دارای تعداد ز ادی رمد است نمی‌توانند ه خو ی عمل کنند رای مثال هنگاهی که پاسخ گسسته و پاسخ پ وسته مانند مدل الا در

4) Sammel and Ryan 5) Baker

دست اشد احت اج ه رورد پارامتر در ماتر س کوار انس خواه م داشت و تعداد کم داده‌ها می‌تواند مشکل ساز گردد و دقت مورد ن ر در رورد پارامترها حاصل نخواهد شد اول ن ار سامل لو ز[۶] و ر ان ا ن مدل را مورد رسی قرار دادند اما روش مورد استفاده نها از ت وری عددی پ چ ده‌ای رای دست وردن پارامترهای مدل استفاده می‌کند و د گر اشکال ن حساس ت س ار ز اد مدل در را ر داده‌های کوچک است کی از راههای رخورد ا چن ن مشکلی کاهش تعداد پارامترهای مدل است که رای اول ن ار در ا ن مقاله رای چن ن داده‌ها ی مورد رسی قرار گرفته است رای ا ن من ور استفاده از روش تجز ه تک عاملی که اول ن ار توس هکمن ارا ه شده است می‌تواند س ار سودمند اشد

مدل اثرات تصادف ا استفاده از تجز ه تک عامل

همانگونه که ان شد روش تک عاملی می‌تواند در کاهش تعداد پارامترهای مدل مناس اشد رای تو ح ا ن روش ا تدا جملات خ ا ε_{im} را رای $m = \,,\ldots,M$ ه دو جز ا اثرات تصادفی b_{im} و v_{im} ه صورت ز ر تقس م می‌کن م

$$\varepsilon_{im} \;=\; b_{im}+v_{im} \qquad\qquad m = \;,\ldots,M$$

در ا ن صورت مدل ا کار ردن روش تجز ه تک عاملی که توس هکمن[۷] رای اثرات تصادفی مورد استفاده قرار گرفته است ه صورت ز ر در خواهد مد

$$\begin{aligned} y^*_{im} &= X'_{im}\beta_m + b_{im} + v_{im} \qquad m = \;,\ldots,M_1 \\ z_{im} &= U'_{im}\beta_m + b_{im} + v_{im} \qquad m = M_1 + \;,\ldots,M \qquad (\;) \end{aligned}$$

نگاه رای قا ل تشخ ص ودن مدل خواه م داشت

$$var(\varepsilon_{im}) \;=\; var(b_{im}+v_{im}) = \sigma^2_{b_m} + \sigma^2_{v_{im}} \qquad m = \;,\ldots,M$$

که در ن

$$var(\varepsilon_{im}) \;=\; var(b_{im}+v_{im}) = \sigma^2_{b_m} + \sigma^2_{v_{im}} = \qquad m = \;,\ldots,M_1$$

ا ن نو تعر ف رای اثرات تصادفی ک ساختار رای ماتر س وار انس کوار انس همانند Σ ا جاد می‌کند

می‌توان تعداد اثرات تصادفی را ا لحا محدود ت روی ساختارکوار انس توس تجز ه تک عاملی

6) Louis 7) Heckman

که توسط هکمن ان شده است کاهش داد به این معنی که فرض کنیم $b_{im} = \lambda_m b_i$ برای $m = \quad, \ldots, M$ که برای رمدهای دودویی $- \quad < \lambda_m < \quad$ و برای رمدهای نرمال $-\infty < \lambda_m < +\infty$ خواهد بود که در آن b_i دارای توزیع نرمال استاندارد است و λ_m پارامترهای مقیاس هستند. در این حالت ساختار همبستگی به صورت زیر است

$$cov(y_m, y_{m'}) = \lambda_m \lambda_{m'} \qquad m, m' = \quad, \ldots, M_۱ \qquad m \neq m'$$

و

$$cov(z_m, z_{m'}) = \lambda_m \lambda_{m'} \qquad m, m' = M_۱, \ldots, M_۲ \qquad m \neq m'$$

و همچنین

$$cov(z_{m'}, y_m) = \lambda_m \lambda_{m'} \qquad m = \quad, \ldots, M_۱ \, , \;\; m' = M_۱, \ldots, M_۲$$

در این صورت تعداد پارامترهای مدل و ماتریس واریانس کوواریانس به میزان قابل توجهی کاهش می‌یابد. برای مثال با $M = \quad$ و $M_۱ = \quad$ و $M_۲ = \quad$ ساختار ماتریس واریانس کوواریانس به صورت زیر در خواهد آمد

$$\Sigma = \begin{pmatrix} & \lambda_۱\lambda_۲ & \lambda_۱\lambda_۳ & \lambda_۱\lambda_۴ & \lambda_۱\lambda_۵ & \lambda_۱\lambda_۶ \\ \lambda_۱\lambda_۲ & & \lambda_۳\lambda_۲ & \lambda_۲\lambda_۴ & \lambda_۲\lambda_۵ & \lambda_۶\lambda_۲ \\ \lambda_۱\lambda_۳ & \lambda_۲\lambda_۳ & & \lambda_۳\lambda_۴ & \lambda_۳\lambda_۵ & \lambda_۳\lambda_۶ \\ \lambda_۱\lambda_۴ & \lambda_۲\lambda_۴ & \lambda_۳\lambda_۴ & & \lambda_۵\lambda_۴ & \lambda_۶\lambda_۴ \\ \lambda_۱\lambda_۵ & \lambda_۲\lambda_۵ & \lambda_۳\lambda_۵ & \lambda_۴\lambda_۵ & \sigma_۵^۲ + \lambda_۵^۲ & \lambda_۵\lambda_۶ \\ \lambda_۱\lambda_۶ & \lambda_۲\lambda_۶ & \lambda_۳\lambda_۶ & \lambda_۴\lambda_۶ & \lambda_۵\lambda_۶ & \sigma_۶^۲ + \lambda_۶^۲ \end{pmatrix}$$

با استفاده از روش تک عاملی برای $M = \quad$ تعداد پارامترهای مدل از پارامتر به پارامتر کاهش می‌یابد.

مثال کاربرد

لیتل و شلوچتر[8] تحقیق روانشناسی بر روی نمونه‌ای از خانوار انجام دادند. هدف از این تحقیق شناسایی تاثیر ناهنجاری‌های والدین بر روی ناهنجارهای فرزندان خانوار است. از آنجا که نمی‌توان اثر شدت ناهنجاری والدین بر روی فرزندان را مشاهده کرد از مدل اثرات تصادفی در برازش داده‌های فوق استفاده شده است. در مدل فوق از رمد که هر یک به نوعی نشان دهنده انعکاس دهنده ناهنجارهای فرزندان خانوار هستند استفاده شده‌اند. هدف از این تحقیق آن است که بدانیم شدت ناهنجارهای والدین چه تاثیری بر روی هر یک از رمدها که رمدهای مشاهده شده از ناهنجارهای فرزندان هستند دارد. در اینجا یک رمد دودویی و دو رمد باقی مانده دارای توزیع نرمال هستند استفاده شده است که عبارتند از

8) Little and Schluchter

جدول خصوص ات توص فی رمدها

متغ ر	Total انحراف مع ار	Total م انگ ن	H/M Risk انحراف مع ار	H/M Risk م انگ ن	Low Risk انحراف مع ار	Low Risk م انگ ن
y	۰٫۴۸۵	۰٫۶۵۳	۰٫۴۲۷	۰٫۷۷۷	۰٫۵۱۷	۰٫۳۷۵
$z_۱$	۱۵٫۳۶۲	۱۰۳٫۱۷۳	۱۱٫۸۱۵	۹۴٫۷۷۷	۱۵٫۹۷۷	۱۱۰٫۸۷۵
$z_۲$	۳۲٫۷۶۴	۱۱۵٫۲۱۷	۳۱٫۳۲۴	۹۶٫۵۵۵	۳۰٫۳۴۹	۱۳۶٫۷۵

متغ ر $y_۱$ که دارای توز دودو ی است و مقدا را می‌پذ رد اگر فرزند خانواده دارای علا م ناهنجاری در رفتار اشد و مقدار را می‌پذ رد اگر فرزند خانواده دارای علا م ناهنجاری در رفتار نباشد توجه ه ا ن نکته روری ه ن ر می‌رسد که ا ن رمد خود رمد مشاهده شده از ک متغ ر پنهان $y_۱^*$ ا توز نرمال است

متغ ر $z_۱$ که دارای توز نرمال است و نشان دهنده نمرات کس شده فرزند خانوار در زمون درک م ل است

متغ ر $z_۲$ که دارای توز نرمال وده و نشان دهنده نمرات کس شده فرزند خانوار در ک زمون قدرت محاوره است

در ا ن مدل از ک متغ ر کمکی X که ک متغ ر دودو ی است استفاده می‌کن م ا ن متغ ر کمکی مقدار را می‌پذ رد هنگامی که والد ن دارای ناهنجار های روانی اندکی اشند LR و را می‌پذ رد هنگامی که حداقل کی از والد ن دارای سا قه ماری روانی و ا ناهنجاری روانی در حد متوس ا ز اد H/MR اشند تحق قات افراد د گر ر روی داده‌های فوق نشان داده است که تاث ر دو گروه ا ناهنجاری متوس و ز اد کسان است ه هم ن دل ل ا ن دو گروه را در ا ن مقاله ا هم ترک کرده‌ا م گنجعلی[9]

تحل ل توص ف داده‌ها

جدول خصوص ات توص فی رمدها را ان می‌کند همانگونه که در جدول ا لاعات توص فی مر و ه رمدها مشخص است نمره درک م ل در ن کودکانی که والد ن نها در معرض کمتر ناهنجاری روانی LR هستند از کودکانی که والد ن نها در معرض ناهنجاری متوس و شد د روانی H/MR هستند شتر است همچن ن رای نمره زمون قدرت محاوره چنانکه در جدول ا لاعات توص فی رمدها مشخص است نمره کس شده توس کودکان در گروه ناهنجاری کمتر LR شتر از گروه H/MR است

ا ن دان معنی است که در صورت معنی‌داری نمره در ن کودکانی که والد ن نها در معرض ناهنجاری متوس و شد د روانی H/MR هستند نسبت ه کودکانی که در معرض کمتر ناهنجاری والد ن هستند LR کمتر است و م زان تاث ر ناهنجاری والد ن در ا ن دو گروه متفاوت خواهد ود

9) Gangali

جدول نتا ج دون در ; رگرفتن همبستگی‌ها

پارامتر	رورد	سـ ح معنی‌داری
β_{01}		
β_{11}		
β_{02}		
β_{12}		
β_{03}		
β_{13}		

تجز ه و تحل ل مدل اثرات تصادفی

مدل اثرات تصادفی م ا ق مدل ه صورت ز ر در خواهد مد

$$\begin{aligned} y^*_{i1} &= \beta_{0m} + \beta_{1m}X + \varepsilon_{im} \qquad m = \\ z_{im} &= \beta_{0m} + \beta_{1m}X + \varepsilon_{im} \qquad m = \ , \end{aligned} \qquad (\ )$$

در ا ن مدل هنوز همبستگی م ان متغ رها λ ها را در ; ر نگرفته‌ا م رورد پارامترهای مدل در جدول داده شده است

چنانچه در جدول مشخص است م ان ناهنجاری والد ن و ناهنجاری فرزندان را ه معنی داری وجود ندارد و شدت ناهنجاری والد ن تا ثری در ناهنجاری در ن فرزندان در سـ ح $\alpha = \ /$ ندارد اما نمره درک م ا و نمره زمون محاوره در ن کودکانی که والد ن نها دارای ناهنجاری کمتری LR هستند شتر از فرزندانی است که والد ن نها دارای ناهنجاری متوسـ ا ز اد H/MR هستند حال همبستگی م ان دو متغ ر پوسته را ; ز ا واردکردن اثرات تصادفی و استفاده از روش تک عاملی وارد مدل می‌کن م در ا ن صورت مدل ه صورت ز ر در خواهد مد

$$\begin{aligned} y^*_{i1} &= \beta_{0m} + \beta_{1m}X + v_{i1} \qquad m = \\ z_{im} &= \beta_{0m} + \beta_{1m}X + \lambda_m b_i + v_{im} \qquad m = \ , \end{aligned} \qquad (\ )$$

رورد پارامترهای مدل در جدول داده شده است

چنانچه در جدول ; ز ملاح ه می‌گردد همان نتا ج مدل قبل ; ز در ا ن مدل دست می د علاوه ر ن می‌توان مقدار همبستگی م ان رمدها را ; ز دست ورد حال حالت کلی را عنی ا وارد کردن اثر تصادفی رای رمد دودو ی ان می‌کن م در حق قت ا در ; رگرفتن اثر تصادفی همبستگی م ان متغ ر گسسته دودو ی و متغ رهای پ وسته را ; ز در ; ر می‌گ ر م در ا ن صورت مدل نها ی ه صورت ز ر در خواهد مد

جدول نتا ج تنها ا در ن رگرفتن همبستگی م ان متغ رهای پ وسته

پارامتر	رورد	س ح معنی داری
$\beta_{\circ 1}$		
β_{11}		
$\beta_{\circ 2}$		
β_{12}		
λ_2		
$\beta_{\circ 3}$		
β_{13}		
λ_3		
σ_{v_2}		
σ_{v_3}		

جدول نتا ج کلی مدل

پارامتر	رورد	س ح معنی داری
$\beta_{\circ 1}$		
β_{11}		
λ_1		
$\beta_{\circ 2}$		
β_{12}		
λ_2		
$\beta_{\circ 3}$		
β_{13}		
λ_3		
σ_{v_2}		
σ_{v_3}		

$$
\begin{aligned}
y_{i1}^* &= \beta_{\circ m} + \beta_{1m}X + \lambda_1 b_i + v_{i1} && m = && (\ ) \\
z_{im} &= \beta_{\circ m} + \beta_{1m}X + \lambda_m b_i + v_{im} && m = \ ,
\end{aligned}
$$

رورد پارامترهای ا ن مدل در جدول داده شده است
ا ن روردها ا استفاده از رنامه NAG و ز ر روال UCF E دست مده‌اند چنانچه از س ح معنی‌داری پارامترهای رورد شده ن ز مشخص است تمامی پارامترها در س ح $\alpha =$ معنی‌دار هستند و ا ن دان معنی است که شدت ناهنجاری والد ن ر روی ناهنجاری فرزندان تا ث ر دارد و هر چه والد ن دارای ناهنجاری شتری اشند خ ر وجود ناهنجاری در فرزندان ن ز شتر است همچن ن شدت ناهنجاری در والد ن در درک م ل و قدرت محاوره فرزندان تا ث ر گذار است ه ا ن صورت که هر چه والد ن دارای ناهنجاری روانی شتری اشند قدرت درک

م ا و محاوره فرزندان کاهش می ا د ا توجه ه رورد پارامترها ماتر س وار انس کوار انس ه صورت ز ر در می د

$$\Sigma = \begin{pmatrix} & / & / \\ / & / & / \\ / & / & / \end{pmatrix}$$

حال ا توجه ه ماتر س وار انس کوار انس الاکوار انس م ان رمدهای گسسته دودو ی و پ وسته نرمال ا وجود اثرات تصادفی در اخت ار است

نت جه‌گ ر و پ شنهادات

چنانکه در جدول مشاهده می‌گردد هنگامی که همبستگی م ان متغ رها را در مدل در ن ر نگرفته‌ا م م ان ناهنجاری والد ن و ناهنجاری فرزندان را ه معنی‌داری وجود ندارد اما هنگامی که در مدل نها ی همبستگی م ان متغ رها را وارد مدل می‌کن م مدل دق قتر عمل نموده و تاث ر ناهنجاری والد ن ر روی ناهنجاری فرزندان مشخص می‌گردد نتا ج حاصل شده از ا ن مدل دق قتر از مدل قبل می اشد

در مدل اثرات تصادفی رای داده‌های م خته ما تغ رات وار انس رمدها را درون دو گروه ا ناهنجاری کمتر والد ن و ناهنجاری متوس ا شتر والد ن ثا ت در ن رگرفت م پ شنهاد می‌گردد هنگاهی که تغ رات وار انس معنی‌دار است اثرات تغ رات وار انس را ن ز وارد مدل کن م ه ور حتم ا ن امر اعث رازش هتر مدل خواهد شد

مراج

[1] Agresti, A. (2002). Categorical data Analysis. New York: Wiley.

[2] Baker, F. B. (1992). Item response Theory: Parameter Estimation Techniques. New York: Dekker.

[3] Gangali. M. (2003). A model for Mixed Continuous and Discrete Response with possible of Missing Response . Journal of Sciences, Islamic Republic of Iran, 14(1) , 53-60.

[4] Heckman, J. (1981). Statistical models for discrete panel data in Manski, C.& Mc Fadden D., strutural analysis of discrete data with econometric applications. 114-195, cambridge, mass: MIT press

[5] Little, R. J. and Schelucher M. (1985). Maximum liklihood estimation for mixed continuous and categorical data with missing value. Biometrika, 72(3) , 497-512.

[6] Little, R. J. A. and Rubin, D. B. (2002). Statistical analysis with Missing Data. New York: Wiley.

[7] NAG, (1996). Numerical Algoritms Group Manual, Mark 16, Oxford, UK.

[8] Sammel, M. D.and Ryan, L. M.(1996). Latent variable models with fixed efects. Biometrics, 52, 220-243.

[9] Sammel, M. D.and Louise. M. and Ryan, L. M.(1997) Latent variable models for Mixed Discrete and Continuous Outcomes. J. R. Statist. Soc. B, 59, No 3, 667-678.