

## آشنایی با مدل سازی معادلات ساختاری تعمیم یافته و کاربرد آن در پژوهش های بهداشتی

**سمیه برمر:** دانشجوی دوره کارشناسی ارشد، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

**معصومه علی محمدیان:** مربی، پژوهشکده بیماری های کبد و گوارش، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

**سید علیرضا سجادی:** پزشک، پژوهشکده بیماری های کبد و گوارش، دانشگاه علوم پزشکی تهران، تهران، ایران

**حسین پوستچی:** دانشیار، پژوهشکده بیماری های کبد و گوارش، دانشگاه علوم پزشکی تهران، تهران، ایران

**سید مصطفی حسینی:** استاد، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

**مهدی یاسری:** استادیار، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران- نویسنده رابط:

m.yaseri@gmail.com

تاریخ پذیرش: ۱۳۹۶/۷/۲۷

تاریخ دریافت: ۱۳۹۶/۶/۱۰

### چکیده

**زمینه و هدف:** مدل معادلات ساختاری تعمیم یافته (Generalized Structure Equation Modeling) خانواده ای از تکنیک های آماری در تجزیه و تحلیل داده های چند متغیره، طبقه بندی شده و ترتیبی است که به اندازه گیری متغیرهای پنهان و روابط میان آنها می پردازد. هدف از این مطالعه توجه بر نوع ساختار داده ها و معرفی مدل GSEM به محققان علوم پزشکی و ارائه مثالی کاربردی در پژوهش های پزشکی است. روش کار: ابتدا مقدمه ای از مدل معادلات ساختاری (Structure Equation Modeling) و نقاط ضعف و قوت آن بیان و سپس به تعریف GSEM و انواع آن پرداخته شد. با ارائه مثالی در بررسی عوامل خطر وضعیت پرفشاری خون در میان افراد مبتلا به دیابت، روش استفاده از GSEM برای متغیرهای پاسخ برنولی توضیح داده شد. داده های این مطالعه نمونه ای تصادفی به حجم ۲۷۱۶ نفر از افراد مبتلا به دیابت در مطالعه کوهورت استان گلستان است.

**نتایج:** متغیرهای سن، شاخص توده بدنی، چاقی شکمی، محل سکونت، وضعیت اقتصادی-اجتماعی، میزان مصرف نمک اثر مستقیمی بر فشار خون بالا داشتند و متغیرهای نژاد، تحصیلات، میزان ویتامین دی و فعالیت بدنی اثر مستقیم و معکوسی بر فشارخون بالا داشتند ( $p < 0.05$ ). نتیجه گیری: این مدل برخلاف مدل SEM نیاز به فرض محدودکننده نرمال بودن داده ها ندارد و ابزاری قدرتمند در تجزیه و تحلیل داده های طبقه بندی شده است. با این حال استفاده از این روش دارای محدودیت هایی است. در حال حاضر در این روش امکان انجام آزمون نیکویی برازش و برآورد اثرات مستقیم و غیرمستقیم و روش اصلاح و تعدیل مدل به طور مستقیم وجود ندارد. واژگان کلیدی: مدل معادلات ساختاری تعمیم یافته، متغیرهای طبقه بندی شده، مدل های علیتی

### مقدمه

پژوهشگر این امکان را می دهد تا یک فرضیه را به یک مدل آزمایشی تبدیل کند. مدل معادلات ساختاری بر اساس تکنیک های آماری همچون همبستگی - رگرسیون و تجزیه و تحلیل واریانس بنا شده است. (Violato 2007). و می تواند تمام

مدل معادلات ساختاری خانواده ای از تکنیک های آماری است که برای تجزیه و تحلیل سیستماتیک داده های چند متغیره برای اندازه گیری ساختارهای نظری (متغیرهای پنهان) و روابط میان آنها استفاده می شود. این روش به

میان خطاها که در بسیاری از مدل‌های کلاسیک جزء محدودیت‌ها است در این مدل مجاز است و در نظر گرفته خواهد شد. علاوه بر این، این مدل‌ها توانایی آن را دارند تا به عنوان مدلی برای تجزیه و تحلیل مشاهدات وابسته مانند مطالعات طولی و یا مطالعات منحنی رشد مورد استفاده قرار گیرند (Beran 2010). با وجود اینکه استفاده از مدل معادلات ساختاری در آمار پزشکی برای حل یک شبکه‌ای از معادلات رگرسیونی وابسته در حال افزایش است، اما یک مانع بزرگ برای استفاده وسیع‌تر از این مدل‌ها وجود متغیرهای طبقه‌بندی‌شده و یا ترتیبی و یا هر نوع متغیر پیوسته غیر نرمال در این نوع مطالعات است. به عبارت دیگر در این مدل‌ها، ساختاری مانند مدل‌های تعمیم‌یافته خطی وجود ندارد (Kupek 2006). هدف از این مقاله معرفی مدل‌های معادلات ساختاری تعمیم‌یافته به عنوان یک ابزار مفید آماری در مطالعاتی است که متغیرهای مدنظر در آن شامل متغیرهای طبقه‌بندی شده و یا غیر نرمال هستند و ارائه مثالی برای متغیرهای پاسخ دودویی (Binary) است.

تعریف مدل معادلات ساختاری تعمیم‌یافته: این مدل‌ها بر اساس ادغام دو چارچوب مدل خطی تعمیم‌یافته (GLM) و نیز معادلات ساختاری توسعه‌یافته تشکیل شده‌اند و برای ساخت مدل‌هایی با متغیرهای پنهان استفاده می‌شوند (Baum et al. 2015).

تفاوت SEM و GSEM: مدل‌های خطی استاندارد و GSEM، مدل‌های خطی تعمیم‌یافته را برازش می‌دهند. در SEM، پاسخ‌ها پیوسته و مدل‌ها رگرسیون خطی می‌باشند. در GSEM، پاسخ‌ها پیوسته (استفاده از مدل‌های رگرسیون خطی)، باینری (استفاده از مدل‌های رگرسیون لجستیک و یا پروبیت)، رتبه‌ای (استفاده از مدل‌های پروبیت رتبه‌ای، لوجیت رتبه‌ای و Clog log رتبه‌ای)، شمارشی یا چندجمله‌ای (استفاده از مدل‌های پواسن و دو جمله‌ای) هستند و مدل‌های مورد استفاده در آنها: رگرسیون خطی، رگرسیون گاما، لوجیت، پروبیت، لوجیت رتبه‌ای، پروبیت رتبه‌ای، پواسن،

معادلات رگرسیونی را بصورت همزمان مدلسازی کند، بنابراین یک چارچوب انعطاف پذیر برای آزمون طیف وسیعی از روابط ممکن بین متغیرها در مدل، از جمله اثرات واسطه و متغیرهای مخدوشگر پنهان ارائه می‌دهد. علاوه بر این در یک سطح عمومی‌تر پارامترهای این مدل می‌توانند سهم هر یک از پیش‌بینی‌کننده‌ها را در ساختار کواریانس اندازه‌گیری کنند (Kupek 2006). مدل معادلات ساختاری بر اساس این فرضیه بنا شده که، متغیرهای اندازه‌گیری شده یک ساختار کواریانس مشخص تولید می‌کنند که دارای توزیع پیوسته نرمال چند متغیره است. این روش یک رویکرد جامع و انعطاف‌پذیر برای طراحی پژوهش و تجزیه و تحلیل داده‌ها و ارزیابی همزمان سازه‌های اندازه‌گیری و ساختار مسیرهای بین این سازه‌ها را فراهم می‌کند در واقع این روش، انعطاف پذیری کافی برای کار کردن با معادلات متعدد مرتبط را بطور همزمان فراهم و تصویر دقیقی از روابط علی بین ساختارهای کلیدی ارائه می‌دهد. استفاده از مدل معادلات ساختاری در سال‌های اخیر به ویژه در علوم اجتماعی، آموزش و پرورش، تجارت، پزشکی و علوم زیستی افزایش یافته است. (Zhu et al. 2006). با اینکه این روش در بسیاری از رشته‌ها مورد استفاده قرار می‌گیرد اما تاکنون استفاده از این روش و بحث در مورد آن در پژوهش‌های پزشکی و اپیدمیولوژی محدود بوده است (Amorim et al. 2010). اگر چه روش‌های مناسب بسیاری برای تجزیه و تحلیل داده‌های پزشکی وجود دارد، اما مدل معادلات ساختاری به‌طور ویژه می‌تواند مفید باشد زیرا امکان به تصویر کشیدن روابط مد نظر و آزمون کردن آن را برای پژوهشگر فراهم می‌کند (Stephenson 2006). قابلیت‌هایی که این مدل‌ها را از سایر مدل‌های مورد استفاده متمایز می‌کند توانایی در برازش رابطه بین متغیرهای مطالعه و در عین حال تقسیم بندی و جداسازی خطای اندازه‌گیری است. به عبارت دیگر این مدل‌ها توانایی آن را دارند تا خطای اندازه‌گیری را از سایر خطاهای موجود در مدل تشخیص داده و آن را متمایز کنند، همچنین همبستگی

به فرض نرمال شرطی نیاز ندارد) (Skron dal 2005; Hesketh 2006).

خلاصه‌ای از مدل‌هایی که توسط GSEM برازش داده می‌شوند:

مدل رگرسیون گاما: این مدل به عنوان رگرسیون لگ گاما شناخته شده است و زمانی که رخدادهای پیوسته و نامنفی باشند و محدوده‌ای از صفر تا بی‌نهایت داشته و اغلب با داده‌هایی که چولگی مثبت دارند مورد استفاده قرار می‌گیرد. استفاده از این مدل زمانی مناسب است که بتوان فرض کرد واریانس خطا با میانگین افزایش می‌یابد.

مدل‌های باینری: مدل‌هایی با رخدادهای باینری متغیرهای پاسخ صفر و یک دارند. این مدل‌ها شامل رگرسیون لجستیک و مدل پروبیت است که با فرمول زیر نشان داده می‌شود.

$$\text{Log } f(y|\mu) = y \log \mu + (1 - y) \log (1 - \mu)$$

که  $\mu$  احتمال موفقیت است.

مدل‌های شمارشی: مدل‌های شمارشی متغیرهای پاسخ شمارشی دارند مانند، مدل‌های پواسن و دوجمله‌ای منفی.

مدل‌های رتبه‌ای: مدل رتبه‌ای امکان برازش مدل برای پاسخ‌هایی را که از نوع رتبه‌ای (مانند سطح رضایت کم، متوسط و زیاد) است را دارا است. این رخدادهای معمولاً از ۱ تا  $k$  شماره‌گذاری شدند و شامل مدل‌های پروبیت رتبه‌ای، لوجیت رتبه‌ای و  $\text{Clog log}$  رتبه‌ای هستند. و با فرمول زیر نشان داده می‌شوند.

$$\Pr(Y = y|z) = \Pr(Y * < ky - z) - \Pr(Y * < ky - 1 - z)$$

مدل رگرسیون چندجمله‌ای: رگرسیون لجستیک چندجمله‌ای که به عنوان لوجیت چندجمله‌ای نیز شناخته می‌شود مشابه مدل‌های رتبه‌ای است و با پاسخ‌های چندگانه سر و کار دارد ولی پاسخ‌های لوجیت چندجمله‌ای نمی‌توانند رتبه‌گذاری شوند. همانند مدل‌های رتبه‌ای، رخدادهای معمولاً با یک متغیر شامل ۱ تا  $k$  شماره‌گذاری می‌شوند (Hesketh 2006; Skron dal 2004; Collett 2003; Cramer 2003).

دو جمله‌ای منفی و لوجیت چند جمله‌ای میباشند. SEM، مدل‌های یک سطحی را برازش می‌دهد و GSEM می‌تواند مدل‌هایی از داده‌های یک سطحی یا چند سطحی را برازش دهد. همچنین GSEM، می‌تواند مدل‌هایی با اثرات آمیخته (Mixed effects)، شامل اثرات تصادفی (Random effects)، اثرات آشیانه‌ای (Nested effects) و اثرات متقاطع (crossed effects) را برازش دهد. با این حال SEM، ویژگی‌هایی ارائه می‌دهد که در حال حاضر در نرم افزارهای مورد استفاده در GSEM ارائه نمی‌شود: از جمله سادگی تفسیر و اجرای آن، برآورد داده‌های گمشده تحت فرض نرمال دو متغیره، آزمون نیکویی برازش، شاخص تعدیل و اصلاح مدل، آزمون اثرات مستقیم و برآورد جمع اثرات مستقیم و غیرمستقیم و برازش مدل با استفاده از خلاصه آماری داده‌ها است (Hesketh 2004; Skron dal 2004).

روش‌های برآورد در GSEM: در اکثر نرم‌افزارهای متداول برای برازش مدل‌های GSEM، تنها دو روش ارائه شده است: روش حداکثر درستنمایی (Maximum Likelihood) و روش شبه درستنمایی (Quasi Maximum Likelihood). که روش حداکثر درستنمایی به صورت پیش فرض در مدل استفاده می‌شود. تابع حداکثر درستنمایی در SEM، فرض می‌کند که توزیع همه متغیرها نرمال چند متغیره است ولی تابع حداکثر درستنمایی در GSEM، فرض می‌کند که همه متغیرها توزیع نرمال شرطی دارند. تابع درستنمایی برای مدل، تحت فرض استقلال مشاهدات عمل می‌کند. این فرضیات روی متغیرهای پنهان و متغیرهای مشاهده شده برون‌زا نیاز خواهند بود (Brown 2006). در واقع روش حداکثر درستنمایی مشخص می‌کند که تا چه حد واریانس‌ها و کوواریانس‌های مشاهده شده بین شاخص‌ها می‌تواند با مدل مشخص شده توسط پژوهش بازسازی شوند. روش (QML) از حداکثر درستنمایی برای برازش مدل استفاده می‌کند ولی برای برآورد خطای استاندارد

اندازه‌گیری و مدل‌های ساختاری است. مدل اندازه‌گیری، روابط میان متغیرهای اندازه‌گیری شده و پنهان را مشخص می‌کند و مدل ساختاری روابط میان متغیرهای پنهان و رگرسیون متغیرهای پنهان روی متغیرهای مشاهده شده را نشان می‌دهد. هنگامی که متغیرها طبقه‌بندی شده هستند نیاز است که مدل اندازه‌گیری معمولی برای متغیرهای مشاهده شده طبقه‌بندی شده تعدیل شود با این حال این متغیرها در مدل ساختاری می‌تواند مشابه موارد پیوسته باقی بماند (Muthen 1984).

مراحل انجام مدل‌سازی معادلات ساختاری تعمیم‌یافته به شرح زیر هست:

فرآیندهای تجزیه و تحلیل مدل‌سازی ساختاری شامل گام‌هایی است که به محقق توصیه می‌شود که حتماً به صورت متوالی این گام‌ها را انجام دهد. این گام‌ها عبارت‌اند از: تدوین مدل، تشخیص مدل، برآورد مدل، آزمون مدل و اصلاح مدل (Kline 2012; Hoyle 2011; Schumacker 2005). برخی از این گام‌ها در حال حاضر در نرم افزارهای متداول مدل معادلات ساختاری تعمیم‌یافته در نظر گرفته نشده است.

مرحله اول: تدوین مدل

مدل معادلات ساختاری در ساده‌ترین سطح، یک گزاره آماری درباره روابط میان متغیرهاست. تدوین مدل یکی از مهم‌ترین مراحل موجود در این مدل‌هاست و شامل فرمول‌بندی گزاره‌هایی درباره پارامترها نیز هست. پارامترها ضرایب عددی هستند که رابطه بین سازه‌ها را توصیف می‌کنند. با تعیین پارامترها اینکه روابط دارای یک جهت یا چند جهت هستند نیز تعیین می‌شود. در تدوین مدل پژوهشگر باید بر اساس نظریه‌ها و یافته‌های تجربی قبلی فرضیاتی در مورد روابط بین متغیرها و سازه‌ها را ارائه دهد و با ترسیم نمودار مسیر متغیرهای درون‌زا و برون‌زا و روابط علیتی بین این متغیرها را مشخص کند. بعد از مشخص کردن متغیرهای پنهان درون‌زا و برون‌زا در نمودار مسیر لازم است تا برای متغیرهای پنهان، شاخص‌هایی (متغیرهای مشاهده شده) مناسبی انتخاب و مرتبط شوند. بهتر است که تعداد بیشتری شاخص برای اندازه‌گیری متغیرهای

انواع متغیرها و دیاگرام مسیر در مدل معادلات ساختاری تعمیم‌یافته: متغیرها در مدل معادلات ساختاری تعمیم‌یافته نیز مانند مدل معادلات ساختاری از جنبه‌های متفاوتی در مدل می‌توانند به صورت متغیرهای پنهان، متغیرهای اندازه‌گیری شده و متغیرهای نشانگر طبقه‌بندی شوند. به عبارت دیگر متغیرهایی که به طور مستقیم قابل اندازه‌گیری نیستند متغیرهای پنهان یا سازه‌ها نامیده می‌شوند که به مفاهیم نظری که قابل مشاهده نیستند اشاره می‌کنند و شکل‌گیری روابط علیتی برآورد شده توسط مدل را ممکن می‌سازند این متغیرها به طور تقریبی به وسیله مجموعه‌ای از متغیرهای مشاهده شده نشان داده می‌شوند (Hair 2005). یک روش برای طبقه‌بندی متغیرها در این مدل‌ها توجه به تأثیری است که متغیر مدنظر بر بقیه متغیرها دارد. که به این ترتیب می‌توان متغیرها را به صورت برون‌زا و درون‌زا طبقه بندی کرد. متغیرهای برون‌زا متغیرهای مستقلی هستند که تحت تأثیر دیگر متغیرها در مدل قرار نمی‌گیرند. متغیرهای درون‌زا متغیرهای وابسته‌ای هستند که تحت تأثیر متغیرهای بیرونی در مدل قرار می‌گیرند (Kline 2005; Kaplan 2000). از آنجائی که مدل‌های معادلات ساختاری شامل مدل‌های پیچیده‌ای هستند بسیاری از پژوهشگران دریافتند که فرم نموداری برای به تصویر کشیدن این مدل‌ها مناسب است. فرم نموداری که دیاگرام مسیر نامیده می‌شود به محقق امکان می‌دهد تا به سرعت روابط وابسته در مدل نظری را تجسم کند (Hair et al. 2005). دیاگرام مسیر توسط اشکال هندسی و بردارها انواع متغیرها (مشاهده شده و پنهان) و روابط میان آن‌ها را نشان می‌دهند. در نمودار مسیر متغیرهای مشاهده شده با مربع یا مستطیل و متغیرهای پنهان با بیضی یا دایره و روابط یک‌طرفه میان متغیرها با یک فلش یک‌طرفه نشان داده می‌شود (Hoyle 2012).

مؤلفه‌های مدل معادلات ساختاری تعمیم‌یافته:

مدل معادلات ساختاری تعمیم‌یافته نیز مانند معادلات ساختاری از دو مؤلفه تشکیل شده است که ترکیبی از مدل‌های

می‌کند تا چه اندازه برآوردها مقدار ماتریس کوواریانس نمونه را پیش‌بینی می‌کند. و هر چه این مقدار به صفر نزدیک‌تر باشد برآزش بهتر است (Bentler 1990).

مرحله چهارم آزمون نیکویی برآزش:

بعد از اینکه برآورد پارامترها برای یک مدل تدوین شده و مشخص شده به دست آمدند پژوهشگر باید تعیین کند که داده‌ها تا چه حد با مدل برآزش دارند؟ در مدل معادلات ساختاری به آزمون فرض صفر  $\sum = \sum \theta$  پرداخته می‌شود. به این معنی که ماتریس کوواریانس مدل با ماتریس کوواریانس داده‌های مشاهده شده متناسب باشد. در مدل‌های معادلات ساختاری آزمون و شاخص‌های برآزش زیادی وجود دارد که هنوز هم به‌طور فزاینده‌ای در حال گسترش است. مهم‌ترین شاخص برآزش مدل آزمون کای دو است ولی به دلیل اینکه تحت شرایط خاصی مانند یک بازه از حجم نمونه عمل می‌کند، شاخص‌های ثانویه‌ای ارائه گردیده است که مهم‌ترین آن‌ها (RMSEA) و CFI (comparative fit index) و Root Mean Square Error of Approximation (Hoyle 2011; Bentler 1980; Iacobucci 2010). در حال حاضر در نرم‌افزارهای متداول برای مدل معادلات ساختاری تعمیم‌یافته امکان آزمون نیکویی برآزش در نظر گرفته نشده است.

مرحله پنجم اصلاح مدل:

اگر برآزش یک مدل نظری (مدل اولیه) به اندازه‌ای که انتظار داشتیم قوی نبود پژوهشگر می‌تواند از طریق شاخص‌هایی که در نرم‌افزارهای مربوطه وجود دارد برای بهبود و اصلاح مدل استفاده کند و مدل خود را بهبود ببخشد. هدف از اصلاح مدل برآورد کردن پارامترهایی است که به لحاظ آماری معنا دار و به لحاظ نظری دارای معنا و مفهوم بوده و دارای برآزش بهتری نسبت به مدل اولیه باشد. برای به دست آوردن نتایج برآزش بهبود یافته مرحله‌ای که در بالا ذکر شد تا زمانی که مدل مناسب به دست بیاید تکرار می‌شود. در حال حاضر در نرم‌افزارهای متداول برای مدل معادلات ساختاری

پنهان استفاده شود که این کاربر اساس تعاریف مفهومی و عملیاتی صورت می‌گیرد (Kline 2005).

مرحله دوم: شناسایی مدل. در مدل معادلات ساختاری بعد از تدوین مدل بسیار با اهمیت است که پژوهشگر قابلیت تشخیص (و یا قابلیت شناسایی) مدل را ارزیابی کند. تشخیص یک مدل مستلزم مطالعه شرایطی برای به دست آوردن یک راه حل منحصر به فرد برای پارامترهای مشخص شده در مدل می‌باشد. در واقع در این مرحله تعیین می‌شود که آیا این مدل با توجه به تعداد نمونه‌ها و روابطی که تعریف شده است می‌تواند برآزش داده شود یا خیر. برای هر پارامتر در مدل مورد استفاده، نیاز است تا یک مقدار از واریانس و یا کوواریانس در ماتریس کوواریانس وجود داشته باشد. در غیر این صورت در اصطلاح مدل غیر قابل شناسایی خواهد بود و یک راه حل واحد برای آن وجود نخواهد داشت. این مشکل همچنین زمانی رخ می‌دهد که متغیرها هم خطی (multicollinearity) بسیاری داشته باشند. در مدل‌های غیر قابل شناسایی تعداد راه حل‌ها نامحدود است و در نتیجه یک جواب منحصر به فرد وجود ندارد (Hoyle 2011).

مرحله سوم برآورد مدل:

هنگامی که مدل تدوین شد و شناسایی مدل، مورد ارزیابی قرار گرفت نوبت به برآورد پارامترها از روی مجموعه‌ای از داده‌های مشاهده شده است. این مرحله شامل مجموعه‌ای از فرآیندهای تکراری است که در هر تکرار ماتریس کوواریانس ضمنی (Implied) ساخته می‌شود و با ماتریس کوواریانس داده‌های مشاهده شده مقایسه می‌گردد. مقایسه این دو ماتریس منجر به تولید یک ماتریس باقیمانده (Residual) می‌شود و این تکرارها تا جایی ادامه می‌یابد که ماتریس باقیمانده به حداقل ممکن برسد (Hoyle 2011). روش برآورد در مدل معادلات ساختاری تعمیم‌یافته روش حداکثر درست‌نمایی است که به عنوان پیش‌فرض در نرم‌افزار مربوطه قرار دارد. روش ML یک فرآیند تکرارشونده است که تعیین

تعمیم یافته است و برای تجزیه و تحلیل داده‌ها از نرم‌افزار STATA ورژن ۱۴ استفاده شده است.

متغیرها و اثرات ارزیابی شده برای مدل معادلات ساختاری تعمیم یافته: در این مطالعه افراد بالاتر از ۴۰ سال مورد بررسی قرار گرفتند، متغیرهای ارزیابی شده شامل سن با چهار رده (۴۹-۴۰ سال، ۵۹-۵۰ سال، ۶۹-۶۰ سال، ۷۰+ سال) جنسیت، نژاد/ قومیت با دو رده (ترکمن، سایر)، تعداد سال‌های تحصیلات شامل (بی‌سواد، کمتر یا مساوی ۵ سال، ۶ تا ۱۲ سال، تحصیلات دانشگاهی)، وضعیت ازدواج (مجرد، متأهل)، محل سکونت (شهر/ روستا) و اندازه‌گیری شاخص توده بدنی شامل چهار طبقه (کم‌وزن، نرمال، اضافه‌وزن، چاق) و WHR (نسبت دور کمر به باسن) شامل دو رده (نرمال / غیر نرمال)، مصرف چای سیاه شامل (کمتر از ۱۰۳۵ میلی‌گرم در روز و بیشتر از ۱۰۳۶ میلی‌گرم در روز) و چای سبز (می‌نوشد/ نمی‌نوشد)، میزان ویتامین دی (کمتر از ۱۰ نانوگرم بر میلی‌لیتر، ۳۰-۱۰ نانوگرم بر میلی‌لیتر، بیشتر از ۳۰ نانوگرم بر میلی‌لیتر)، میزان مصرف نمک (کمتر از ۵ گرم و بیشتر از ۵ گرم) و فعالیت بدنی مبتنی بر شغل (دارد/ ندارد) و وضعیت اقتصادی اجتماعی که بر اساس مالکیت بعضی از لوازم شامل یازده سؤال (داشتن اتومبیل شخصی-تلویزیون رنگی-سیستم صوتی تلویزیون-کولر-ماشین ظرف‌شویی-یخچال فریزر-کامپیوتر و اندازه منزل) که به عنوان ساختار پنهان در مدل در نظر گرفته می‌شود. فشار خون در هر بازو در وضعیت نشسته دو بار اندازه‌گیری شد و دو دقیقه استراحت بین هر اندازه‌گیری وجود داشت. فشار خون شخص بر حسب متوسط دو بار اندازه‌گیری در هر بازو محاسبه شده است و به صورت متغیر دودویی (دارد/ ندارد) طبقه‌بندی شد.

روش اجرا: مدل مفهومی که روابط میان مفاهیم ارائه شده در این مطالعه را تعیین می‌کند، در شکل ۱ نشان داده شده است. ما قبلاً ۲۸ مسیر را پیش‌بینی کردیم که به‌طور مستقیم و غیرمستقیم بر فشارخون بالا تأثیر دارند. این ۲۸ مسیر از ۱۲ متغیر شامل (سن، جنسیت، نژاد، تحصیلات، وضعیت ازدواج، شاخص توده

تعمیم یافته امکان آزمون نیکویی برازش و اصلاح مدل در نظر گرفته نشده است. (Brown 2006).

تفسیر مدل: پس از اینکه آزمون‌های نیکویی برازش، نشان‌دهنده تناسب داده‌ها و مدل می‌باشند، معنی‌داری پارامترهای مدل، مورد ارزیابی قرار می‌گیرد. برای تفسیر پارامترهای برآورد شده در مدل معادلات ساختاری تعمیم یافته ممکن است نیاز باشد تا تبدیلی بر روی ضرایب اعمال گردد. به‌طور مثال از exp ضرایب استفاده شود.

مثالی از مدل معادلات ساختاری تعمیم یافته با متغیر پاسخ برنولی و مراحل انجام آن:

داده‌ها: در این مثال از بخشی از داده‌های مقطعی مطالعه کوهورت گلستان که در طی سال‌های (۲۰۰۸-۲۰۰۴) انجام شده، استفاده گردید. در این مطالعه از پرسشنامه و مصاحبه چهره به چهره و آزمایش‌ها بالینی برای دریافت اطلاعات استفاده شده است. جزئیات روش اجرای این مطالعه کوهورت در مقاله دیگر توضیح داده شده است (Pourshams et al. 2010). در این مطالعه از بین ۵۰۰۴۵ فرد شرکت‌کننده در مطالعه کوهورت گلستان، ۲۷۱۶ نفر از افرادی که مبتلا به دیابت نوع دو بودند بطور تصادفی انتخاب شده و مورد آنالیز قرار گرفتند. با توجه به معیارهای (JNC7) افرادی با داشتن فشارخون سیستولیک کمتر از ۱۲۰ میلی‌متر جیوه و فشارخون دیاستولیک کمتر از ۸۰ میلی‌متر جیوه در صورت عدم مصرف داروهای ضد فشارخون به عنوان افرادی با فشار خون نرمال در نظر گرفته شده‌اند افرادی با فشارخون سیستولیک بزرگتر و یا مساوی ۱۴۰ میلی‌متر جیوه یا فشارخون دیاستولیک بزرگتر و یا مساوی ۹۰ میلی‌متر جیوه به عنوان افرادی با فشارخون بالا طبقه‌بندی شدند (Ferguson et al. 2008). همچنین افرادی که داروهای کاهش دهنده فشارخون مصرف می‌کردند، شناسایی و به عنوان فشارخون بالا ثبت شده‌اند. هدف از این بخش از مطالعه بررسی عوامل خطر وضعیت فشارخون در بین افراد مبتلا به دیابت با به کارگیری مدل معادلات ساختاری

مستقیم WHR به این صورت بود که شانس ابتلا به پرفشاری خون با داشتن چاقی شکمی  $1/30$  برابر افرادی که چاقی شکمی نداشتند بود این افزایش از لحاظ آماری معنی دار و مثبت بود. فعالیت بدنی اثر مستقیم و معکوسی بر پرفشاری خون داشت بطوریکه شانس ابتلا به پرفشاری خون در افرادی که فعالیت بدنی داشتند  $0/85$  برابر افرادی که فعالیت بدنی نداشتند بود. علاوه بر این فعالیت بدنی از طریق WHR نیز بر فشارخون بالا تأثیر داشت که این اثر نیز معنی دار و مثبت بود. نژاد اثر مستقیم و معکوسی بر فشارخون بالا داشت بطوریکه شانس ابتلا به فشارخون بالا در افراد ترکمن  $0/93$  برابر سایر افراد محل سکونت اثر مستقیمی بر پرفشاری خون داشت بطوریکه شانس ابتلا به پرفشاری خون در افراد روستایی  $1/10$  برابر افراد ساکن شهر بود. داشتن میزان کافی از ویتامین دی، اثر مستقیم و معکوسی بر پرفشاری خون داشت. بطوریکه شانس ابتلا به پرفشاری خون در افرادی که میزان کافی از ویتامین دی داشتند  $0/80$  و  $0/71$  برابر افرادی که ویتامین دی کمتری داشتند بود. تحصیلات اثر مستقیم و معکوسی بر پرفشاری خون داشت. به طوری که شانس ابتلا به فشارخون بالا در افرادی با تحصیلات ۶-۱۲ سال  $0/82$  برابر افراد بیسواد بود. وضعیت اقتصادی- اجتماعی اثر مستقیمی بر فشارخون بالا داشت بطوریکه شانس ابتلا به فشارخون بالا با کاهش وضعیت اقتصادی- اجتماعی  $1/18$  برابر می شد. مصرف نمک اثر مستقیمی بر فشارخون بالا داشت بطوریکه شانس ابتلا به فشارخون بالا که بیشتر از ۵ گرم در روز نمک مصرف می کردند  $1/04$  برابر افرادی که کمترین میزان نمک مصرف می کردند بود. پرفشاری خون با متغیرهای جنسیت ( $p=0/194$ ) و وضعیت ازدواج ( $p=0/127$ ) ارتباط معنی دار آماری نشان نداد. در این مثال فقط به بیان روابط مسیرهای اصلی منتهی به فشارخون پرداخته شده است. مسیرهای فرعی که شامل متغیرهای جنسیت، سن، فعالیت بدنی، وضعیت اقتصادی- اجتماعی هستند و از طریق متغیرهای واسطه BMI و WHR بر فشارخون بالا تأثیر دارند در جدول

بدنی (BMI)، نسبت دور کمر به باسن (WHR)، میزان مصرف نمک، میزان ویتامین دی، فعالیت بدنی، محل سکونت که متغیرهای قابل مشاهده هستند و متغیر وضعیت اقتصادی- اجتماعی SES که به صورت متغیر پنهان در نظر گرفته شده اند و متغیرهایی که برای ساخت متغیر پنهان وضعیت اقتصادی- اجتماعی در نظر گرفته شده است خارج می شوند و به متغیر وضعیت فشارخون مرتبط می شوند. در این مدل وضعیت فشارخون که به صورت یک متغیر دودویی است از طریق تابع ربط برنولی با لینک لوجیت (Bernoulli/logit) به سایر متغیرهای این مدل متصل شده است. از طرف دیگر شاخص توده بدنی با استفاده از تابع ربط رتبه ای با لینک لوجیت (Ordinal/logit) و نسبت دور کمر به باسن نیز با استفاده از تابع ربط برنولی با لینک لوجیت، که این متغیرها نقش متغیرهای واسطه (میانی) را بر عهده دارند به سایر متغیرها در مدل متصل اند.

## نتایج

در این داده ها از ۲۷۱۶ فرد دیابتی مورد بررسی، ۲۱۱۷ نفر ( $78\%$ ) فشارخون داشتند. برآورد ضرایب مسیرها در مدل معادلات ساختاری تعمیم یافته در نمودار مفهومی در شکل یک نشان داده شده است. در جدول روابط میان متغیرهایی که به فشارخون منتهی شده اند نشان داده شده است. نسبت شانس ابتلا به فشارخون بالا بر اساس هر متغیر به همراه سطح معناداری و فاصله اطمینان آنها در جدول شماره یک درج شده است. بر اساس نتایج جدول ۱، شاخص توده بدنی اثر مستقیمی بر پرفشاری خون داشت به طوری که شانس ابتلا به پرفشاری خون با هر یک طبقه افزایش در شاخص توده بدنی  $1/66$  برابر افرادی با شاخص توده بدنی نرمال بود. سن اثر مستقیمی بر پرفشاری خون داشت بطوریکه با افزایش هر دهه سن شانس ابتلا به پرفشاری خون  $1/75$  برابر افرادی با گروه سنی ۴۹-۴۰ بود. داشتن چاقی شکمی (WHR) اثر مستقیمی بر پرفشاری خون داشت. اثر

ندارد. بنابراین مدل معادلات ساختاری تعمیم یافته یک ابزار قدرتمند در تجزیه و تحلیل داده‌های طبقه‌بندی شده است. یکی از محدودیت‌های مدل معادلات ساختاری تعمیم یافته نسبت به مدل معادلات ساختاری کلاسیک این است که در این روش در حال حاضر در نرم افزارهای مورد نظر امکان انجام آزمون نیکویی برازش (Goodness of fit) و روش اصلاح و تعدیل مدل (Model Modification) به طور مستقیم وجود ندارد. همچنین هنوز در این مدل ها امکان برآورد اثرات مستقیم و غیرمستقیم در نظر گرفته نشده است.

### نتیجه گیری

در این مطالعه متغیرهای سن، شاخص توده بدنی، چاقی شکمی، محل سکونت، وضعیت اقتصادی- اجتماعی، میزان مصرف نمک اثر مستقیمی بر فشار خون بالا داشتند و متغیرهای نژاد، تحصیلات، میزان ویتامین دی و فعالیت بدنی اثر مستقیم و معکوسی بر فشارخون بالا داشتند ( $p < 0/05$ ). در این مطالعه متغیرهای جنسیت و وضعیت ازدواج با فشارخون بالا ارتباط معنی داری نداشتند. با وجود برخی از محدودیت ها، در GSEM امکان برآورد همزمان متغیرهای مورد بررسی وجود دارد و این مدل روشی برای آزمون‌های پیچیده از انواع متغیرهای طبقه‌بندی شده ارائه می‌دهد و تدوین مدل‌های نظری یکپارچه و پژوهش‌های پیشرفته را ممکن می‌سازد که در مدل‌های معادلات ساختاری کلاسیک امکان پذیر نیست.

بیان نشده اند اما این اثرات نیز از لحاظ آماری در سطح پنج درصد معنی دار می باشند.

### بحث

با توسعه مدل‌های معادلات ساختاری، امروزه محققان پزشکی باید ابزاری توانمند را برای تجزیه و تحلیل‌های پیچیده از مدل‌های علیتی بکار بگیرند. این مدل‌ها برتر از روش‌های همبستگی مانند رگرسیون در تجزیه و تحلیل همزمان متغیرهای چندگانه هستند و عوامل پنهان، خطای اندازه‌گیری را کاهش می‌دهند. در مدل معادلات ساختاری توانایی رسیدگی به رخدادهایی با توزیع‌های پارامتریک همچون پواسون، گاما و شبه پارامتری همچون لگ توابع خطر به‌طور سنتی وجود ندارد. همچنین تبدیلات جایگزین مانند لجستیک یا Clog-log در آن‌ها در دسترس نیست. ادغام توانایی‌های اصلی SEM‌های اولیه با مزایای هر دو معادلات همزمان و مدل‌های اندازه‌گیری با مدل‌سازی تعمیم یافته خطی رخ داده است. مدل معادلات ساختاری تعمیم یافته برآورد مدل‌های SEM با رخدادهایی از توزیع‌های متفاوت و برآورد پارامتر با رخدادهای متفاوت (معادلات) را ممکن می‌سازد (Christ 2014). در این مقاله در مورد مدل‌های معادلات ساختاری تعمیم یافته برای متغیرهای پاسخ غیر پیوسته بحث شد. این مدل برخلاف مدل معادلات ساختاری نیاز به فرض محدودکننده نرمال بودن داده‌ها که در اکثر متغیرهای مورد استفاده در تحقیقات بهداشتی برقرار نیست،

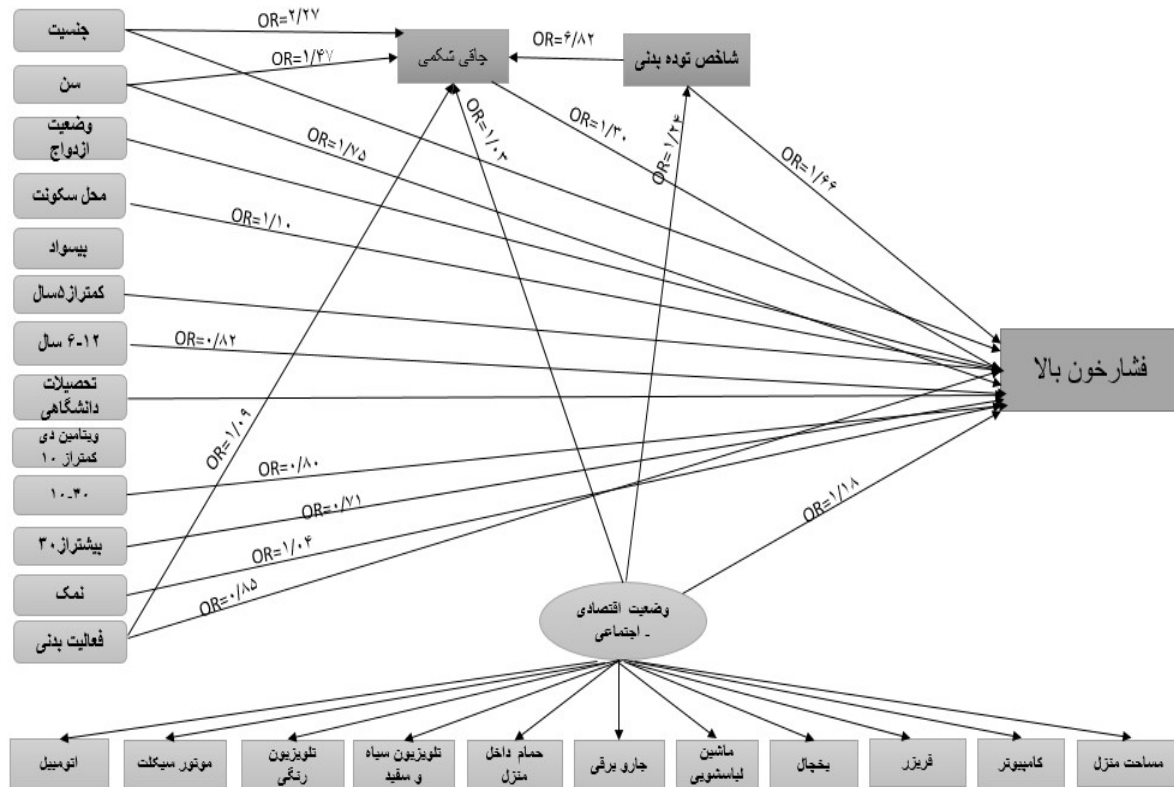
### تشکر و قدردانی

از کلیه دست اندرکاران، مسئولان و همکاران طرح کوهورت گلستان جهت گردآوری و ارائه داده ها و نیز مشاوره های علمی تشکر و قدردانی می گردد.

جدول ۱ - نتایج مدل معادلات ساختاری تعمیم یافته برای مسیرهای اصلی مورد بررسی

p-value	آماره Z	فاصله اطمینان ۹۵٪		نسبت شانسی*	رده مورد بررسی	متغیر
		حد پایین	حد بالا			
۰/۱۹۴	۱/۳۰	۰/۹۹	۱/۰۸	۱/۰۳	مرد زن	جنسیت
<۰/۰۰۱	۴۱/۳۸	۱/۷۵	۱/۸۰	۱/۷۵	۴۹-۵۰ سال ۵۹-۵۰ سال ۶۹-۶۰ سال ۷۵+ سال	سن افزایش هر دهه سن
۰/۰۰۱	۳/۸۵	۱/۰۵	۱/۱۷	۱/۱۰	شهری روستایی	محل سکونت
۰/۰۶۰	-۲/۳۵	۰/۸۷	۱/۰۱	۰/۹۵	گروه پایه ترکمن سایر بی سواد برابری کمتر از ۵ سال	نژاد میزان تحصیلات
<۰/۰۰۱	-۲/۵۷	۰/۹۰	۰/۷۵	۰/۸۲	۱۲-۶ سال	وضعیت ازدواج
۰/۰۷۲	-۳/۲۱	۰/۷۲	۱/۰۱	۰/۸۶	تحصیلات دانشگاهی	شاخص توده بدنی
۰/۰۱۷	۲/۳۹	۱/۰۳	۱/۳۱	۱/۱۸	متغیر پنهان	وضعیت اقتصادی - اجتماعی
۰/۱۲۷	۱/۵۲	۰/۹۹	۱/۱۲	۱/۰۵	گروه پایه غیرمتاهل نرمال لاغر	وضعیت ازدواج شاخص توده بدنی
<۰/۰۰۱	۳۶/۱۸	۱/۶۱	۱/۷۱	۱/۶۶	گروه پایه ندارد دارد کمتر از ۱۰ ۳۰-۱۰ بیشتر از ۳۰	با افزایش هر طبقه شامل چاق نسبت دور کمر به باسن (چاقی شکمی) میزان ویتامین دی نانوگرم/میلی لیتر
<۰/۰۰۱	-۹/۸۵	۰/۸۴	۰/۷۵	۰/۸۰	گروه پایه ندارد دارد	فعالیت بدنی
<۰/۰۰۱	-۹/۷۲	۰/۶۷	۰/۷۶	۰/۷۱	گروه پایه ندارد دارد	میزان مصرف نمک در روز
۰/۰۳۸	۲/۰۷	۱/۰۱	۱/۰۸	۱/۰۴	گروه پایه کمتر از ۵ گرم بیشتر از ۵ گرم	

(ضرایب رگرسیونی) \* Exp



شکل ۱ - مدل مفهومی به همراه EXP ضرایب مسیر

\* در این شکل فقط مسیرهایی که معنادار شده اند نشان داده شده است. همچنین از سطوح پایه مسیری قرار داده نشده است (ویتامین دی برحسب نانوگرم بر میلی لیتر، میزان نمک برحسب گرم در هر روز)

## References

- Baum, Ch., Lööf, H. and Nabavi, P.S.A., 2015. A New Approach to Estimation of the R&D - Innovation - Productivity Relationship. *Economics of Innovation and New Technology*, 26(1-2), pp. 121-133.
- Bentler, PM., 1990. comparative fit indexes in structural models. *Psychological Bulletin*, 107, pp. 238-246.
- Bentler PM, B.D.G., 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, pp. 588-666.
- Beran T, V.C., 2010. Structural equation modeling in medical research: a primer. *BMC Research Notes*, 3, P. 267.
- Brown, TA., 2006. *Confirmatory Factor Analysis for Applied Research Second Edition*. New York: THE GUILFORD PRESS.
- Christ, Sh., J. Lee, D. and Lam B, Z.D., 2014. Structural Equation Modeling: A Framework for ocular and other Medical Sciences Research. *Informa Healthcare USA*, 21(1), pp. 1-134.
- Collett, D., 2003. *Modeling binary data*. 2nd, edn. london: chapman&Hall. BOOK.
- Cramer, J.S., 2003. *Logit models from economics and other fields*. Cambridge

- University press.
- Kupek, 2006. Beyond logistic regression: structural equations modelling for binary variables and its application to investigating unobserved confounders. *BMC Medical Research Methodology*, 6, P.13.
- Hair, J. and Anderson R, T.B.W., 2005. Multivariate data analysis.
- Hoyle, RH., 2012. *Handbook of Structural Equation Modeling*. 1 st ed. New York: The Guilford press.
- Hoyle, RH., 2011. *Structural Equation modeling: concept, issue and Applications*. Guilford press.
- Iacobucci, D., 2010. Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*.
- Kaplan, D., 2000. Structural equation modeling: foundations and extensions. Thousand Oaks: Sage Publications. *Advanced Quantitative Techniques in the Social Sciences Series*, 10.
- Kline, RB., 2005. *Principles and practice of structural equation modeling* New York: The Gilford Press.
- Amorim, L., Fiaccone, R., Santos, C., Santos, T., de Moraes Oliveira N., S Barbosa, D., Santos, L.S. and S Matos, M.B., 2010. Structural equation modeling in epidemiology. 26(12), pp. 2251–2262.
- Muthen, BO., 1984. A general structural equation model with dichotomous, ordered categorical and continuous latent indicators. *Psychometrika*, 49, pp.115–132.
- Pourshams, A., Khademi, H., Fazeltabar Malekshah, A., Islami, F., Nouraei, M., Sadjadi, AR., Jafari, E., Rakhshani, N., Salahi, R., Semnani, Sh., Kamangar. F., Abnet, Ch., Ponder, B., Day, N., M Dawsey, S. and Boffetta, PM.R., 2010. Cohort Profile: The Golestan Cohort Study—a prospective study of oesophageal cancer in northern Iran. *Int J Epidemiol*, 39(1), pp. 52–59.
- Hesketh, S.R., Skrondal A, P.A., 2004. Generalized multilevel structural equation modeling. *Psychometrika*, 69, pp.167–190.
- Hesketh, S.R. and Skrondal A, P.A., 2005. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128, pp. 301–323.
- Hesketh S.R., 2006. Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society*, 169, pp. 805–827.
- Schumacker RE, L.R., 2012. *A Beginners Guide to Structural Equation Modeling: Third Edition*.
- Skrondal A, R.H.S., 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. BOOK.
- Stephenson M, R.L.H., 2006. On the Use of Structural Equation Modeling in Health Communication Research we focus on SEM as an analytical tool that. *HEALTH COMMUNICATION*, 20(2), pp.159–67.
- Ferguson, L., Trevor, S., Novie, OM., Younger., L., Marshall, K., Tulloch-Reid1, Marilyn. B., Lawrence, W., Elizabeth, M., Ward2, D.E.A. and R.J.W., 2008. Prevalence of prehypertension and its relationship to risk factors for cardiovascular disease in Jamaica: Analysis from a cross-sectional survey. *BMC Cardiovascular Disorders*.
- Violato C, G.H.K., 2007. How to Use Structural Equation Modeling in Medical Education Research: A Brief Guide. *Teaching and Learning in Medicine: An International Journal*, (19:4), pp. 362–371.
- Zhu, B., Walter, SD., Rosenbaum, PL. and Russell D J, R.P., 2006. Structural equation and log-linear modeling: a comparison of methods in the analysis of a study on caregivers' health. *BMC Medical Research Methodology*, 6, P. 49.

## Generalized Structural Equation Modeling (GSEM) and its Application in Health Researches

**Barmar, S.,** MSc. Student, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

**Alimohammadian, M.,** MSc. Instructor, Digestive Diseases Research Center, Digestive Diseases Research Institute, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

**Sadjadi, SA.,** MD. Digestive Diseases Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran

**Poustchi, H.,** Ph.D. Associate Professor, Digestive Diseases Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran

**Hosseini, SM.,** Ph.D. Professor, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Islamic Republic of Iran, Tehran, Iran

**Yaseri, M.,** Ph.D. Assistant Professor, Department of Epidemiology and Biostatistics, School of Public Health, University of Medical Sciences, Tehran, Iran-Corresponding Author:m.yaseri@gmail.com

Received: Sep 1, 2017

Accepted: Oct 19, 2017

### ABSTRACT

**Background and Aims:** Generalized Structural Equation Modeling (GSEM) is a family of statistical techniques utilized in the analysis of multivariate, categorical and ordinal data in order to measure latent variables and their connection with each other. The aim of this study is to consider the structure of data, and introducing GSEM to medical science researchers and presenting a practical example of in medical science researches.

**Materials and Methods:** An introduction to Structural Equation Modeling (SEM), along with its advantages and disadvantages was presented, and also GSEM and its all kind of forms was specified. An example to study hypertension risk factors in patients suffering from diabetes was carried out, which was a demonstration of using GSEM method for binary response variables. The data includes a random sample of 2716 people from Golestan province cohort studies.

**Results:** Age, body mass index, abdominal obesity, residence place, socioeconomic status, salt intake had direct effect on hypertension. Race, education, vitamin D and physical activity had direct and reverse effect on hypertension (p.value<0.05).

**Discussion:** Unlike SEM, the limitative hypothesis that our data should have a normal distribution do not needed in this model, also GSEM is powerful tool in the analysis of categorized data. Nevertheless this method cannot perform goodness of fit test, and adjustment and modification method of the model directly, and that they are some limitation in using this method.

**Keywords:** Generalized Structural Equation Modeling, Categorical Variables, Causal Modeling