

توزیع‌های چوله و کاربرد آن در تحلیل داده پزشکی کارخانه فولاد مبارکه اصفهان

محمد غلامی فشارکی^۱، انوشیروان کاظم نژاد^۲، فرید زایری^۳، محسن روضاتی^۴

چکیده:

امروزه استفاده از توزیع‌های منعطف مانند توزیع‌های چوله که توانایی بیشتری را در بیان رفتار داده‌های مشاهده شده دارند افزایش چشمگیری یافته است. این توزیع‌ها، بیشتر در علوم پزشکی و علوم رفتاری که مقادیر واقعی متغیرهای تصادفی دارای توزیعی غیر متقارن است مورد استفاده قرار می‌گیرد. با توجه به کاربرد زیاد توزیع‌های چوله، در این مقاله بعد از مروری مختصر در مورد توزیع‌های چوله معروف، توزیع‌های نرمال، تی، نرمال و تی چوله به داده‌های استخراج شده از داده‌های پزشکی شرکت فولاد مبارکه برازش داده شده و سپس با استفاده از شاخص AIC بهترین مدل، از بین مدل‌های موجود مورد انتخاب قرار می‌گیرد.

واژه‌های کلیدی: روش برشی پنهان، توزیع‌های نرمال چوله-مستقل، نرمال چوله، تی چوله.

۱ مقدمه

و هسو (۲۰۰۵)؛ جارا، کیونتانا و سن مارتین (۲۰۰۸)، پنهان شدن فرآیند تولید داده‌ها و کاهش اطلاعات (جارا و همکاران (۲۰۰۸)) و یکتا نبودن تبدیل در مطالعات مختلف می‌گردد (برای مثال عدم محاسبه تبدیل لگاریتمی در داده‌هایی با دامنه منفی) (بندیپدهایی، لاجوس، آبان‌تولی و قوش (۲۰۱۰)). استفاده از روش توزیع‌های غیر نرمال نیز علی‌رغم اینکه مشکلات قبلی را تا حدود زیادی مرتفع می‌کند اما محققان را از دسترسی به خصوصیات مناسب توزیع‌های بیضوی مانند نرمال دور می‌سازد. روش دیگر در حل این معضل به کار بردن توزیع‌هایی است که دارای خصوصیات بیضوی بوده و از لحاظ نظری قادر به تبیین تغییرات مشاهدات بدون استفاده از فرض نرمالیتی هستند. به این ترتیب استفاده از خانواده‌هایی از توزیع‌های نامتقارن که قابلیت

یکی از فرض‌های اساسی در بسیاری از مدل‌سازی‌های آماری فرضیه نرمال بودن توزیع مشاهدات است. از طرفی این فرضیه در بسیاری از حالات برقرار نیست، از این رو محققان جهت بررسی‌های خود به روش‌هایی چون تبدیل داده‌ها (به ویژه استفاده از تبدیل کاکس-باکس)، استفاده از روش‌های ناپارامتری و استفاده از توزیع‌های غیر نرمال دست می‌زنند. استفاده از روش‌های ناپارامتری علی‌رغم کاهش اطلاعات از لحاظ اجرایی دارای محدودیت‌های زیادی برای محققان بوده و از طرف دیگر استفاده از روش تبدیل متغیرها نیز باعث مشکلاتی چون تفسیر نامناسب پارامترهای تحت بررسی، احراز نشدن فرض نرمالیتی توأم متغیرها در حالات چند متغیره (تسونگ و لی (۲۰۰۳)؛ لی، لین، لی

^۱ گروه آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس، mohammad.gholami@modares.ac.ir.

^۲ گروه آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس، Kazem.an@modares.ac.ir.

^۳ گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی.

^۴ پزشک واحد پیگیری بیماری‌ها، بخش طب صنعتی، شرکت فولاد مبارکه اصفهان، مبارکه.

در این معادله در صورتی که $h(x) = g(x)$ انتخاب گردد، خانواده توزیع های چوله-مقارن^۶ (گپتا، چانگ و هوانگ، ۲۰۰۲) به دست می آید. که از آن جمله می توان به توزیع نرمال چوله، تی چوله، یکنواخت چوله، لاپلاس چوله و لجستیک چوله اشاره نمود. از این میان، توزیع نرمال چوله از جمله اولین، همانند و خوش تعریف ترین توزیع ها نسبت به توزیع نرمال است که خصوصیات آن بصورت وسیع در حالت تک متغیره توسط آزالینی (۱۹۸۵) و در حالت چند متغیره توسط آزالینی و دال ول (۱۹۹۶) مورد بررسی قرار گرفته است. اما آزالینی نشان داده است که میزان پارامتر چولگی این توزیع در بازه $(-۰/۹۹۵, ۰/۹۹۵)$ تغییر می کند و این بدان معنی است که از توزیع نرمال چوله تنها در مواقعی می توان استفاده نمود که چولگی داده ها زیاد نباشد. این موضوع را از دیدگاه دیگری نیز می توان توضیح داد. در صورتی که پارامتر چولگی λ از مقدار ۵ بیشتر شود، توزیع نرمال چوله به سرعت به توزیع نیم نرمال^۷ میل نموده و عملاً برای مقادیر بالای چولگی (از مقدار ۵ به بالا) امکان تشخیص توزیع نرمال چوله بسیار دشوار می گردد (اسپینوسا، ۲۰۰۸). همچنین از معایب دیگر توزیع نرمال چوله می توان به برآورد بی نهایت پارامتر چولگی این توزیع به خاطر صعودی شدن تابع درستیابی نسبت به پارامتر اشاره نمود (جنتون، ۲۰۰۴؛ دی، ۲۰۱۰). البته محققان مختلف به ارائه راهکارهایی جهت حل این مشکل پرداخته اند (پوسی، ۲۰۰۰).

در صورتی که $g(x) = \varphi(x)$ (که $\varphi(\cdot)$ تابع چگالی نرمال استاندارد است) و تابع $H(\cdot)$ دلخواه انتخاب شود خانواده توزیع های چوله نرمال-مقارن^۸ (ناداریج و کوتز، ۲۰۰۳) به دست می آید در صورتی که $H(x) = \Phi(x)$ (که $\Phi(\cdot)$ تابع توزیع نرمال استاندارد است) و $g(\cdot)$ دلخواه انتخاب گردد، خانواده توزیع های مقارن-نرمال^۹ (گمز، ونگاس و بولفارین، ۲۰۰۷) حاصل می شود. گپتا و همکاران (۲۰۰۲) با انتخاب توزیع های لاپلاس، یکنواخت و لجستیک نشان دادند که خانواده توزیع های چوله نرمال-مقارن برخلاف خانواده توزیع های نرمال چوله توانایی پوشش میزان چولگی بیشتری را دارا هستند، اما یکی از مشکلات در استفاده از آنها سنگینی محاسبات توزیع های این خانواده ها

تبیین چولگی و کشیدگی داده ها را دارا باشند، برای مدل بندی داده ها و تحلیل های آماری نیرومند نسبت به تخطی از فرض نرمالیتی می تواند مفید و کارساز باشد. از این رو امروزه استفاده از تحلیل های رگرسیونی و مدل های خطی با استفاده از توزیع ی بیضوی چوله به طور قابل ملاحظه ای مورد توجه پژوهشگران قرار گرفته است (آزالینی و کاپیتانیو، ۱۹۹۹)؛ فرناندز و استیل (۱۹۹۸)؛ ساهو، دی و برانکو (۲۰۰۳). از این رو با توجه به ضرورت استفاده از توزیع های چوله در مدلسازی و علی الخصوص تحلیل داده های پزشکی، در این مقاله بر آن شدیم تا بصورت مختصر به مروری در مورد توزیع های چوله پرداخته (بخش ۲) و سپس با استفاده از یکسری داده پزشکی با درجات مختلف چولگی و کشیدگی (بخش ۳) به برازش ۴ مدل نرمال، تی، نرمال چوله و بالاخره تی چوله (بخش ۴) بپردازیم و در نهایت با استفاده از نتایج حاصل، به بینشی نو در استفاده از توزیع های چوله دست یابیم.

۲ مروری بر فرآیندهای تولید توزیع های چوله

جهت ایجاد توزیع های چوله روش های متعددی وجود دارد. یکی از متداول ترین این روش ها استفاده از روش برشی پنهان^۵ (معادله ۱) است. (ابطحی، توحیدی و بهبودیان، ۲۰۰۹) آزالینی و کاپیتانیو (۲۰۰۳) و آزالینی (۲۰۰۵) ثابت کردند در صورتی که $w(x)$ یک تابع فرد و $H(\cdot)$ یک تابع توزیع با تابع چگالی مقارن حول صفر $h(\cdot)$ و $g(x)$ یک تابع چگالی مقارن باشد، آنگاه تابع چگالی $f_x(X)$ که به فرم معادله (۱) تعریف می شود، یک تابع چگالی نامقارن خواهد بود.

$$f_x(X) = 2g(x)H(w(x)), \quad -\infty < x < \infty. \quad (1)$$

در صورتیکه $w(x) = \lambda x$ انتخاب شود، معادله (۱) به فرم معادله (۲) در می آید.

$$f_x(X) = 2g(x)H(\lambda x), \quad -\infty < x < \infty. \quad (2)$$

⁵Hidden Truncation Method

⁶Skew Symmetric

⁷Half Normal

⁸Skew Normal Symmetric

⁹Skew Symmetric Normal

نموده‌اند. همچنین خردمندی و همکاران (۲۰۱۰) با استفاده از تبدیل توزیع نرمال به تی با v درجه آزادی در معادلات (۳-۵) به جای تابع چگالی نرمال به ارائه توزیع‌های تعمیم یافته جدید دست یافتند (معادلات (۹-۶)).

$$f_{\lambda_1, \lambda_2, v}(x) = \chi^2_{t_v}(x) \Phi\left(\frac{\lambda_1 x}{\sqrt{1 + \lambda_2 x^2}}\right), \quad (6)$$

$$-\infty < x < \infty, \quad \lambda_2 \geq 0,$$

$$f_{\lambda, \gamma, v}(x) = \chi^2_{t_v}(x) \Phi(\lambda x + \gamma x^3), \quad (7)$$

$$-\infty < x < \infty,$$

$$f_{\lambda, n}(x) = \chi^2_{c_n}(\lambda) \varphi(x) \Phi^n(\lambda x), \quad (8)$$

$$c_n(\lambda) = \left(\int_{-\infty}^{+\infty} \chi^2_{\varphi(x)} \Phi^n(\lambda x) dx \right)^{-1},$$

$$f_{\lambda_1, \lambda_2}(x) = \chi^2_{c_{n, \lambda_1, \lambda_2}}(\lambda) \times \varphi(x) \Phi^n\left(\frac{\lambda_1 x}{\sqrt{1 + \lambda_2 x^2}}\right), \quad (9)$$

$$-\infty < x, \lambda_1 < \infty, \quad \lambda_2 \geq 0,$$

از دیگر تحقیقات صورت گرفته در مورد توزیع بالا کریشان می توان به موارد زیر اشاره کرد. یادگاری و همکاران (۲۰۰۸) و نیز آرلانو ول و همکاران (۲۰۰۴)، به تعمیم دیگری از این توزیع دست یافتند (معادله (۱۰)). حسنعلی پور و شرفی (۲۰۱۰) به بررسی برخی خصوصیات این توزیع پرداختند. نکوخو و علامت ساز (۲۰۱۱) با در نظر گرفتن توزیع لاپلاس در عوض تابع چگالی نرمال (معادله (۱۱)) در توزیع تعمیم یافته آرلوند ول و همکاران (۲۰۰۴) به بررسی برخی از خصوصیات این توزیع پرداخته و اصطلاحاً این توزیع را توزیع تعمیم یافته متقارن چوله لاپلاس نرمال نامگذاری کردند.

$$f_{\lambda_1, \lambda_2}(x) = \chi^2_{c_{n, \lambda_1, \lambda_2}}(\lambda) \times \varphi(x) \Phi^n\left(\frac{\lambda_1 x}{\sqrt{1 + \lambda_2 x^2}}\right), \quad (10)$$

$$-\infty < x, \lambda_1 < \infty, \quad \lambda_2 \geq 0,$$

که در آن

$$c_{n, \lambda_1, \lambda_2}(\lambda) = \left(\int_{-\infty}^{+\infty} \chi^2_{\varphi(x)} \Phi^n\left(\frac{\lambda_1 x}{\sqrt{1 + \lambda_2 x^2}}\right) dx \right)^{-1},$$

است. همچنین در این راستا و با توجه به مشکلات خانواده توزیع نرمال و همچنین تك مدی بودن خانواده توزیع نرمال چوله و لزوم وجود خانواده‌هایی با پارامتر چولگی و کشیدگی وسیعتر و همچنین ارائه توزیع‌های دومدی، محققان با تغییر $w(x)$ به ارائه توزیع نرمال تعمیم یافته^{۱۰} پرداختند که در این زمینه می‌توان به کارهای ما وجنتن (۲۰۰۴) با در نظر گرفتن معادله (۱) و با پارامترهای زیر

$$w(x) = \lambda x + \gamma x^3,$$

$$g(x) = h(x) = \phi(x),$$

(معادله (۳))، بالا کریشان (۲۰۰۲) با در نظر گرفتن پارامترهای

$$w(x) = \lambda x,$$

$$g(x) = \varphi(x),$$

$$H(x) = c_n(\lambda) \Phi^n(\lambda x),$$

(معادله (۴))

$$f_{\lambda, \gamma}(x) = \chi^2_{\varphi(x)} \Phi(\lambda x + \gamma x^3), \quad (3)$$

$$-\infty < x < \infty,$$

$$f_{\lambda, n}(x) = \chi^2_{c_n}(\lambda) \varphi(x) \Phi^n(\lambda x), \quad (4)$$

$$c_n(\lambda) = \left(\int_{-\infty}^{+\infty} \chi^2_{\varphi(x)} \Phi^n(\lambda x) dx \right)^{-1},$$

آرلانو ول و همکاران (۲۰۰۴) با در نظر گرفتن معادله (۱) و با پارامترهای $w(x) = \frac{\lambda_1 x}{\sqrt{1 + \lambda_2 x^2}}$ و $g(x) = h(x) = \phi(x)$ (معادله (۵))

$$f_{\lambda_1, \lambda_2}(x) = \chi^2_{\varphi(x)} \Phi\left(\frac{\lambda_1 x}{\sqrt{1 + \lambda_2 x^2}}\right), \quad (5)$$

$$-\infty < x < \infty, \quad \lambda_2 \geq 0,$$

اشاره نمود.

در توزیع تعمیم یافته ول و همکاران در حالتی که $\lambda_2 = \lambda_1^2$ آنگاه توزیع حاصله توزیع نرمال چوله خمیده^{۱۱} نامیده می‌شود. یکی از مزایای استفاده از این تبدیل سادگی بدست آوردن توزیع مزدوج در محاسبه برآوردهای بیزی است [۳]. شرفی و بهبودیان (۲۰۰۶) در مورد خصوصیات توزیع بالا کریشان در حالتی که n برابر با ۲ است بحث

¹⁰Skew-Generalized Normal (SGN)

¹¹Skew-Curved Normal

توزیع تی چوله به توزیع کوشی چوله و در صورتی که v به بینهایت میل کند آنگاه این توزیع به توزیع نرمال چوله میل می‌کند که کاربرد این توزیع‌ها در مدل‌سازی در منابع مختلف بیان شده است (آزالینی، ۲۰۰۸؛ لین و لی، ۲۰۰۸). همچنین در صورتی که توزیع U را توزیع بتا با پارامترهای $Beta(v, 1)$ قرار دهیم، توزیع X به توزیع چوله بریده^{۱۵} تبدیل می‌شود. کاربرد این توزیع‌ها در مدل‌سازی در مقاله ونگ و جنتن (۲۰۰۶) بیان شده است. همچنین در صورتی که U یک توزیع گسسته که دو مقدار γ و 1 را با احتمال v و $1-v$ اختیار می‌کند انتخاب شوند، توزیع X به توزیع نرمال آلوده چوله^{۱۶} (SCN) به فرم معادله (۱۶) تبدیل خواهد شد. در این فرمول هرگاه $\gamma = 1$ آنگاه SCN به توزیع نرمال چوله تبدیل می‌شود.

$$f(x) = v \varphi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{\sqrt{\gamma}(x-\mu)}{\sigma}\right) + (1-v) \varphi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right). \quad (16)$$

همچنین در راستای دو مدی کردن چگالی توزیع‌های چوله اولیویرا و همکاران (۲۰۰۹) با اضافه کردن $\kappa = \int_{-\infty}^{+\infty} x^2 f_0(x) dy < \infty$ خانواده ارائه شده در معادله (۱) را به یک خانواده دومی تبدیل نموده‌اند. نحوه تعریف این معادله در معادله (۱۷) و بصورت نمونه یک خانواده توزیع چوله دو مدی تعریف می‌شود. در این معادله هر چه مقدار α بیشتر باشد دومی بودن تابع چگالی بیشتر و هر چه مقدار آن به صفر نزدیکتر باشد، میزان دومی بودن تابع چگالی کمتر می‌شود.

$$f(x) = v \left(\frac{1+\alpha x^2}{1+\alpha \kappa} \right) f_0(x) H(w(x)), \quad (17)$$

$$-\infty < x < \infty, \alpha \geq 0,$$

همچنین باید گفت علاوه بر دو روش بیان شده قبلی محققان از راه کارهای دیگری نیز جهت تعریف توزیع‌های چوله بهره می‌جویند که از آن جمله می‌توان به روش آماره ترتیبی^{۱۷} اشاره نمود. در این روش توزیع چوله به شکل معادله (۱۸) و با دو پارامتر نامعلوم a و b تعریف می‌شود. خصوصیات این توزیع بصورت مفصل بوسیله گپتا و ناداریچ (۲۰۰۴) و ایوجن و همکاران (۲۰۰۲) مورد بررسی قرار گرفته است. اصطلاحاً به

$$f_{\lambda_1, \lambda_2}(x) = e^{-|x|} \Phi\left(\frac{\lambda_1 x}{\sqrt{1+\lambda_2 x^2}}\right), \quad (11)$$

$$-\infty < x, \lambda_1 < \infty, \lambda_2 \geq 0.$$

روش دیگر در ساخت توزیع‌های تی، کوشی چوله استفاده از راه کاری است که اصطلاحاً به آن کلاس توزیع‌های نرمال چوله-مستقل^{۱۲} (SNI) می‌گویند (معادله (۱۲))

$$X = \mu + \frac{Z}{\sqrt{U}}. \quad (12)$$

در عبارت بالا μ پارامتر موقعیت بوده، U یک متغیر تصادفی مثبت با توزیع تجمعی $H(u|v)$ و تابع چگالی $h(u|v)$ است و v یک بردار که مشخص کننده توزیع متغیر تصادفی U می‌باشد در نظر گرفته می‌شود. Z نیز یک توزیع نرمال چوله با میانگین صفر، واریانس σ^2 و پارامتر چولگی λ است. $(ZSN(0, \sigma^2, \lambda))$ (آرلانوول، بولفارین و لاجوس، ۲۰۰۵؛ لاجوس، دی و کانچوک، ۲۰۰۹) با دانستن U توزیع X بصورت معادله (۱۳) و توزیع حاشیه‌ای به شکل معادله (۱۴) تبدیل می‌شود.

$$X|u \sim SN(\mu, u^{-1}\sigma^2, \lambda), \quad (13)$$

$$f(x) = v \int_0^\infty \Phi\left(\frac{\sqrt{u}\lambda(x-\mu)}{\sigma}\right) \times \phi\left(\frac{\sqrt{u}(x-\mu)}{\sigma}\right) dH(u|v). \quad (14)$$

در این معادله در صورتی که $\lambda = 0$ باشد آنگاه کلاس X به کلاس توزیع نرمال-مستقل^{۱۳} (NI) با تابع چگالی به شکل معادله (۱۵) تبدیل خواهد شد.

$$f(x) = \int_0^\infty \phi\left(\frac{\sqrt{u}(x-\mu)}{\sigma}\right) dH(u|v). \quad (15)$$

تابع چگالی کلاس NI در واقع یک کلاس از توزیع‌های نرمال آمیخته^{۱۴} است. همچنین در کلاس SNI در صورتی که به جای توزیع U ، توزیع کای دو با v قرار دهیم آنگاه توزیع X به توزیع تی چوله با v درجه آزادی که توسط آزالینی و کاپتانو (۲۰۰۳) و گوپتا (۲۰۰۳) مورد تحقیق قرار گرفته است تبدیل می‌شود. همچنین در صورتی که v برابر ۱ باشد،

¹²Skew-Normal/Independent Distributions

¹³Normal/Independent (NI)

¹⁴Mixture Normal

¹⁵Skew-Slash

¹⁶Skew-Contaminated Normal(SCN)

¹⁷Order Statistics Method

و

$$f_{\lambda}(x) = \frac{1}{\pi + 2 \tan^{-1}(\lambda)} \varphi(x) \Phi(\lambda |x|), \quad (22)$$

$$-\infty < x, \lambda < +\infty,$$

۳ مثال کاربردی

داده‌های مورد بررسی در این مقاله براساس اطلاعات جمع‌آوری شده از افراد داوطلب که طی یک مطالعه مقطعی در سال ۱۳۸۹ در مجتمع فولاد مبارکه اصفهان بصورت تصادفی انتخاب شده‌اند جمع‌آوری شده است (داده‌ها در ضمیمه ارائه شده است). در این مطالعه از میان ۶۴۰ نفر افراد مبتلا به فشار خون بالا با استفاده روش نمونه‌گیری تصادفی ساده ۱۶۶ نفر انتخاب و سپس این افراد فرم پرسشنامه استرس شغلی و اطلاعات دموگرافیک را تکمیل نمودند. در این مطالعه جهت اندازه‌گیری استرس شغلی از فرم استاندارد پرسشنامه GHQ12 (پرسشنامه سلامت عمومی) که جهت بررسی سلامت عمومی و مشکلات روانی در جامعه طراحی گردیده است استفاده شده است. منتظری و همکاران (۲۰۰۳) روایی و پایایی نسخه فارسی این پرسشنامه را مورد تأیید قرار داده‌اند. در این مطالعه اطلاعات جمعیت شناختی مانند سن و سابقه کاری با استفاده پرونده پزشکی جمع‌آوری شده و برای اندازه‌گیری فشار خون از دستگاه فشار خون جیوه‌ای کالیبره شده از دو دست و پس از ۵ دقیقه استراحت استفاده شده است. در این مطالعه همچنین متغیرهای دور کمر و شاخص چاقی^{۲۱} (BMI) نیز توسط پزشک اندازه‌گیری شده است.

۴ نتایج

در جدول ۴ میزان میانگین، میانه، انحراف معیار هر متغیر و همچنین میزان چولگی و کشیدگی متغیرهای موجود در مطالعه و همچنین میزان اختلاف این چولگی و کشیدگی از توزیع نرمال به ترتیب با استفاده از آزمون‌های دی آگوستینو (۱۹۷۰) و آزمون آنسکاپ-گلیان (۱۹۸۳) مورد بررسی قرار گرفته است. برای انجام دو آزمون گفته شده با استفاده از بسته "moments" و با استفاده از نرم افزار R به آزمون پرداختیم. همچنین از آنجایی که تعداد داده‌ها در این مطالعه تقریباً بزرگ بود

توزیع تشکیل یافته به شکل معادله (۱۸)، توزیع بتا-نرمال گفته می‌شود. از آنجایی که میزان چولگی توزیع بتا نرمال تحت دو پارامتر نامعلوم a و b یکنواخت نبوده و امکان دو مدی شدن و همچنین قابلیت پوشش توزیع‌های با چولگی زیاد را ندارد، برخی از محققین (فریا و استیل، ۲۰۰۴) ترجیح می‌دهند تا با قرار دادن شرط $a = b$ این توزیع را به توزیعی با یک پارامتر محدود کنند و به خصوصیات بهتری از این توزیع دست یابند. از دیگر راه‌کارهای تولید داده‌های چوله می‌توان به روش مقیاس عامل معکوس^{۱۸} (فرناندز و استیل، ۱۹۹۸) (معادله (۱۹)) و روش یکپارچه^{۱۹} (معادله (۲۰)) (فریا و استیل، ۲۰۰۶) اشاره کرد.

$$f_{a,b}(x) = [\Phi(x)]^{a-1} \times [1 - \Phi(x)]^{b-1} / 2B(a, b), \quad (18)$$

$$f_{\gamma}(x) = \frac{\gamma}{\gamma+1} \varphi(x\gamma^{-\text{sign}(x)}), \quad (19)$$

$$\gamma > 0,$$

$$f_p(x) = h(x)p[H(x)], \quad (20)$$

که در آن p یک تابع چگالی دلخواه است. شایان ذکر است که به غیر از روش‌های بیان شده، محققان مختلف بصورت مشخص به تعریف توزیع‌های چوله پرداخته‌اند که از آن جمله می‌توان به کارهای ناداریچ (۲۰۰۸) (معادله (۲۱))، کیم (۲۰۰۵) تحت عنوان توزیع نرمال چوله دو تکه‌ای^{۲۰} (معادله (۲۲)) و یا حالت تعمیم یافته آن (جمالزاده، علیپور و بالا کریشان، ۲۰۰۹) (معادله (۲۳)) و یا حالت ساده شده آن وقتی ($\rho = 0$) (معادله (۲۴)) اشاره کرد. در جداول ۱ تا ۳ توزیع‌های چوله متداول به همراه روش تعریف و خصوصیات آن ارائه شده است

$$f_{\alpha,a,b}(x) = Cx^{\alpha-1} \exp(-ax^2 - bx), \quad (21)$$

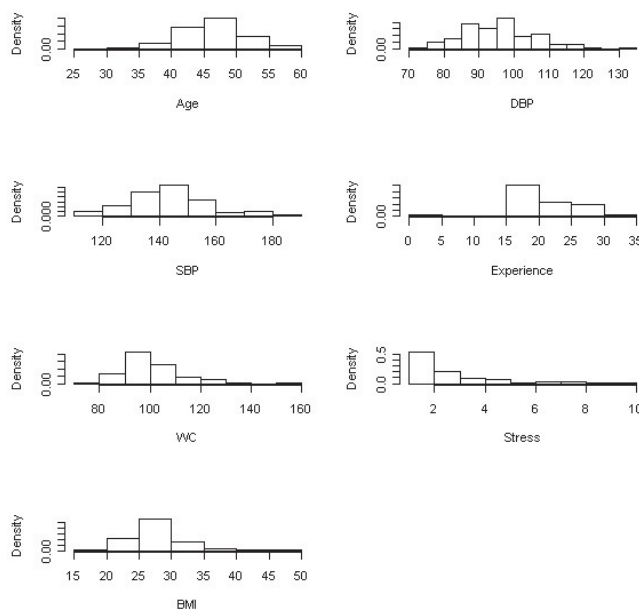
$$x > 0, a > 0, -\infty < b < +\infty.$$

که در آن

$$C = \int x^{\alpha-1} \exp(-ax^2 - bx),$$

¹⁸Inverse Scale Factors Method¹⁹Unified Method²⁰Two-Piece Skew-Normal²¹Body Mass Index (BMI)

نشود ولی میزان چولگی و کشیدگی داده مورد مطالعه در محدوده نرمال باشد باز توزیع نرمال چوله برازش بهتری را نسبت به توزیع نرمال نشان می‌دهد. این موضوع را از دیدگاه دیگر می‌توان اینگونه بیان نمود که از آنجایی که اثر عدم تقارن بر فرض نرمالیتی در کل جدی تر از دم سنگین بودن توزیع بواسطه وجود نقاط پرت است (مدهولکر و هیتسون، ۲۰۰۰). در این موارد توزیع نرمال چوله برازش بهتری را نسبت به نرمال نشان خواهد داد. همچنین در مورد متغیرهای سابقه کار، دور کمر، استرس و شاخص BMI از آنجایی که میزان چولگی و کشیدگی هر دو فاصله زیادی با توزیع نرمال داشتند از این رو استفاده از توزیعی که علاوه بر قابلیت مدل‌سازی چولگی، قابلیت بالایی در مدل‌سازی کشیدگی دارد مورد پیش‌بینی بود. از این رو توزیع تی چوله گزینه مناسبی برای این متغیرها به شمار می‌رفت که نتایج حاصل از برازش مدل (جدول ۵) نیز نشان دهنده این برازش بهتر در توزیع تی چوله نسبت به سه توزیع دیگر بود. در پایان و براساس نتایج این مطالعه می‌توان راهکار زیر را جهت مدل‌سازی به فرم جدول ۶ ارائه نمود. به این ترتیب که ابتدا براساس آزمون‌های شاپیرو، دی آگوستینو و آنسکاپ-گلیان به ترتیب به بررسی نرمال بود، چولگی و کشیدگی متغیر مورد مطالعه با توزیع نرمال پرداخته و سپس با استفاده از جدول ۶ به انتخاب توزیع مناسب از میان ۴ توزیع نرمال، تی، نرمال چوله و تی چوله دست می‌زنیم.



شکل ۱. نمودار هیستوگرام برای متغیرهای مورد مطالعه

($n = 166$) جهت بررسی هم‌توزیعی متغیرهای مورد مطالعه با توزیع نرمال از آزمون شاپیرو^{۲۲} (پاتریک، ۱۹۸۲) و با استفاده از نرم افزار R به این مهم پرداخته و نتایج را در جدول ۴ ارائه نمودیم. نمودار هیستوگرام متغیرهای مورد مطالعه نیز در شکل ۱ نمایش داده شده است. با توجه به نمودار ۱ و نتایج جدول ۴ می‌توان به خوبی به تخطی توزیع داده‌ها از فرضیه نرمال بودن مشاهدات پی برد. در جدول ۵ خلاصه نتایج برازش و همچنین AIC مدل که یک شاخص کفایت مدل است. (مقادیر کمتر آن نشان دهنده برازش بهتر است) حاصل از ۴ توزیع نرمال، تی، نرمال چوله و تی چوله ارائه شده است. برای برازش توزیع‌های نرمال و تی با استفاده از بسته "bbmle" و جهت برازش توزیع‌های نرمال چوله و تی چوله با استفاده از بسته "sn" و با استفاده از نرم افزار R به این منظور دست یافتیم.

۵ بحث و نتیجه‌گیری

یکی از مسائل فرا روی محققان در استفاده از توزیع‌های آماری، انتخاب درست این توزیع‌ها است. در این مقاله به منظور آشکار ساختن کاربرد توزیع‌های چوله به برازش ۴ توزیع نرمال، تی، نرمال چوله و بالاخره تی چوله در یکسری داده پزشکی با درجات مختلف چولگی و کشیدگی دست زدیم. همانگونه که در جدول ۴ مشاهده می‌نمائید از آنجایی که میزان چولگی متغیر سن در حد توزیع نرمال و میزان کشیدگی آن در حد این توزیع نیست می‌توان پیش از برازش مدل به مناسب بودن توزیع تی به عنوان بهترین توزیع جهت برازش مدل مشکوک بود. این فرضیه با استفاده از معیار شاخص AIC مورد بررسی قرار گرفت و براساس نتایج جدول ۵ این فرضیه مورد تأیید قرار گرفت. همچنین برای متغیرهای فشار خون سیستولیک و دیاستولیک با وجودی که میزان چولگی و کشیدگی در حد توزیع نرمال گزارش گردید اما آزمون نرمال بودن داده‌ها تأیید نشده و براساس مقادیر AIC توزیع نرمال چوله بهترین گزینه برای برازش برای این متغیرها در نظر گرفته شد. آزالینی نشان داده است که میزان پارامتر چولگی توزیع نرمال چوله در بازه زیر

$$(-0.995, 0.995)$$

تغییر می‌کند و این یعنی اینکه از توزیع نرمال چوله در مواقعی می‌توان استفاده کرد که چولگی داده‌ها زیاد نباشد (آزالینی، ۱۹۸۵؛ اسپینوسا، ۲۰۰۸). بنابراین در مواقعی که هم‌توزیعی داده‌ها با توزیع نرمال تأیید

²²Shapiro test

جدول ۱: برخی از توزیع‌های چوله معروف براساس نوع انتخاب و (معادله (۱))

نام توزیع	$w(\cdot)$	$g(\cdot)$	$H(\cdot)$
نرمال چوله یا نرمال-نرمال چوله	λx	نرمال	نرمال
کوشی چوله	λx	کوشی	کوشی
t -چوله	λx	تی	تی
چوله یکنواخت	λx	یکنواخت	یکنواخت
لاپلاس چوله	λx	لاپلاس	لاپلاس
لجستیک چوله	λx	لجستیک	لجستیک
کوشی-نرمال چوله	λx	نرمال	کوشی
نرمال t -چوله	λx	نرمال	تی
یکنواخت-نرمال چوله	λx	نرمال	یکنواخت
لاپلاس-نرمال چوله	λx	نرمال	لاپلاس
نرمال-لجستیک چوله	λx	نرمال	لجستیک
تی - نرمال چوله	λx	*تی	نرمال
یکنواخت - نرمال چوله	λx	یکنواخت	نرمال
لاپلاس - نرمال چوله	λx	لاپلاس	نرمال
لجستیک - نرمال چوله	λx	لجستیک	نرمال
بالاکریشنان چوله	λx	نرمال	$c_n(\lambda) \Phi^n(\cdot)$
توزیع چوله نرمال تعمیم یافته نوع ۱	$\frac{\lambda_1 x}{\sqrt{1+\lambda_2 x^2}}$	نرمال	نرمال
توزیع چوله نرمال تعمیم یافته نوع ۲	$\lambda x + \gamma x^3$	نرمال	نرمال
توزیع چوله نرمال تعمیم یافته نوع ۴	$\frac{\lambda_1 x + \lambda_2 x^3}{\sqrt{1+\lambda_2 x^2}}$	نرمال	نرمال
توزیع چوله تی-نرمال تعمیم یافته نوع ۱	$\frac{\lambda_1 x}{\sqrt{1+\lambda_2 x^2}}$	تی	نرمال
توزیع چوله تی-نرمال تعمیم یافته نوع ۲	$\lambda x + \gamma x^3$	تی	نرمال
توزیع چوله تی-نرمال تعمیم یافته نوع ۳	λx	تی	$c_{n,v}(\lambda) \Phi^n(\cdot)$
توزیع چوله تی-نرمال تعمیم یافته نوع ۴	$\frac{\lambda_1 x + \lambda_2 x^3}{\sqrt{1+\lambda_2 x^2}}$	تی	نرمال
توزیع چوله لاپلاس-نرمال تعمیم یافته	$\frac{\lambda_1 x}{\sqrt{1+\lambda_2 x^2}}$	لاپلاس	نرمال

جدول ۲: برخی از توزیع‌های چوله معروف براساس نوع انتخاب U (معادله (۱۱))

نام توزیع	U
تی چوله	$\chi_{(n)}^2, (n > 1)$
کوشی چوله	$\chi_{(1)}^2$
چوله بریده شده	$Beta(v, 1)$
چوله آلوده به نرمال	$vI(\gamma) + (1-v)I(1)$

جدول ۳: تابع برخی از توزیع‌های چوله معروف دیگر

نام توزیع	تابع چگالی
توزیع بتا نرمال دو پارامتری	$\frac{[\Phi(.)]^{a-1}[1-\Phi(.)]^{b-1}}{\sqrt{B(a,b)}}$
توزیع بتا نرمال تک پارامتری	$\frac{[\Phi(.)]^{a-1}[1-\Phi(.)]^{\frac{1}{a}-1}}{\sqrt{B(a,\frac{1}{a})}}$
توزیع عامل معکوس	$f(x) = \frac{\gamma}{\gamma+1} \varphi(x\gamma^{-sign(x)})$
توزیع چوله ناداریچ	$Cx^{\alpha-1} \exp(-ax^2 - bx)$
توزیع نرمال چوله دو تکه‌ای	$\frac{1}{\pi + \sqrt{2} \tan^{-1}(\lambda)} \varphi(x) \Phi(\lambda x)$
توزیع نرمال چله دو تکه‌ای تعمیم یافته	$\frac{1}{\pi + \sqrt{2} \tan^{-1}(\lambda_1)} \varphi(x) \Phi(\lambda_1 x) \Phi(\lambda_2 x)$

جدول ۴: میزان میانگین، میانه، انحراف معیار، چولگی و کشیدگی متغیرهای موجود در مطالعه و میزان اختلاف این چولگی و کشیدگی از توزیع نرمال

پارامتر	سن	DBP*	SBP**	سابقه کار	WC++	استرس	BMI++
میانگین	۴۶/۴۷	۹۶/۸۹	۱۴۴/۶۵	۲۱/۳۰	۱۰۰/۴۷	۲/۸۷	۲۷/۶۶
میانه	۴۷	۹۶	۱۴۴	۲۰	۹۹	۲	۲۷/۰۳
انحراف معیار	۵/۴۴	۱۰/۵۹	۱۴/۵	۴/۷۱	۱۱/۴۴	۲	۴/۰۷
چولگی	-۰/۲۵	۰/۴۰	۰/۵۳	-۰/۹۸	۱/۳۵	۱/۴۷	۱/۶۱
آماره آزمون †	۰/۸۹۹	۱/۴۰	۱/۸۱	-۳/۰۵	۳/۸۴	۴/۰۸	۴/۳۲
†Pvalue	۰/۳۶۸	۰/۱۶	۰/۰۷	۰/۰۰۲	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱
کشیدگی	۳/۹۸	۳/۲۹	۳/۲۷	۵/۹۸	۵/۹۲	۵/۰۲	۷/۷۳
آماره آزمون ‡	۲/۱۹۷	۰/۹۷	۰/۹۴	۴/۰۹	۴/۰۵	۳/۳۶	۵/۰۱
‡ Pvalue	۰/۰۲۸	۰/۳۳۰۹	۰/۳۴۸۸	۰/۰۰۰۴	۰/۰۰۰۱	۰/۰۰۰۸	۰/۰۰۰۱
آماره آزمون ○	۰/۹۸۱	۰/۹۸۳	۰/۹۷۰	۰/۸۷۳	۰/۹۰۹	۰/۸۹۲	۰/۸۹۳
○ Pvalue	۰/۰۱۸	۰/۰۴۱	۰/۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱

* : فشار خون دیاستولیک، ** : فشار خون سیستولیک، + : دور کمر، ++ : شاخص توده بدنی،

† : آماره آزمون و همچنین Pvalue برای آزمون دی آگوستینو جهت مقایسه چولگی متغیر با چولگی توزیع نرمال

‡ : آماره آزمون و همچنین Pvalue برای آزمون آنسکاپ- گلیان جهت مقایسه کشیدگی متغیر با کشیدگی توزیع نرمال

○ : آماره آزمون و همچنین Pvalue برای آزمون شاپیرو جهت مقایسه توزیع داده‌ها با توزیع نرمال

جدول ۵: پارامترهای مدل برای متغیرهای مورد مطالعه براساس ۴ مدل نرمال، تی، نرمال چوله و تی چوله

مدل	پارامتر	سن	DBP*	SBP**	سابقه کار	WC+	استرس	BMI++
نرمال	μ	۴۶٫۳۷ ($\pm ۰٫۱۵$)	۹۶٫۸۹ ($\pm ۰٫۴۲$)	۱۴۴٫۶۵ ($\pm ۰٫۸۲$)	۲۱٫۳۰ ($\pm ۰٫۱۲$)	۱۰۰٫۴۷ ($\pm ۰٫۸۸$)	۲٫۸۷ ($\pm ۰٫۱۵$)	۲۷٫۶۶ ($\pm ۰٫۳۱$)
	σ	۵٫۴۲ ($\pm ۰٫۳۰$)	۱۰۰٫۵۶ ($\pm ۰٫۵۸$)	۱۴٫۴۵ ($\pm ۰٫۷۹$)	۴٫۶۹ ($\pm ۰٫۲۶$)	۱۱٫۴۱ ($\pm ۰٫۶۳$)	۱٫۸۹ ($\pm ۰٫۱۰$)	۴٫۰۶ ($\pm ۰٫۲۲$)
	AIC	$\pm ۱۰۳۶٫۴۰$	$\pm ۱۲۵۷٫۷۷$	$\pm ۱۳۶۱٫۸۹$	$\pm ۹۸۸٫۲۷$	$\pm ۱۲۸۳٫۳۶$	$\pm ۶۸۷٫۰۲$	$\pm ۹۴۰٫۲۱$
	μ	۴۶٫۵۳ ($\pm ۰٫۴۱$)	۹۶٫۷۳ ($\pm ۰٫۸۴$)	۱۴۴٫۰۵ ($\pm ۰٫۱۷$)	۲۱٫۳۵ ($\pm ۰٫۳۳$)	۹۸٫۵۲ ($\pm ۰٫۷۳$)	۲٫۳۸ ($\pm ۰٫۱۳$)	۲۶٫۹۵ ($\pm ۰٫۲۶$)
تی	σ	۴٫۸۱ ($\pm ۰٫۴۲$)	۱۰۰٫۰۶ ($\pm ۰٫۸۳$)	۱۴٫۹۷ ($\pm ۰٫۱۵$)	۳٫۴۳ ($\pm ۰٫۳۱$)	۷٫۴۰ ($\pm ۰٫۸۰$)	۱٫۱۹ ($\pm ۰٫۱۳$)	۲٫۵۷ ($\pm ۰٫۲۷$)
	df	۹٫۲۲	۲۱٫۱۸	۹٫۷۴	۴٫۲۱	۲٫۹۵	۲٫۷۲	۲٫۹۶
	AIC	$\pm ۱۰۳۶٫۸۸$	$\pm ۱۲۵۹٫۳۲$	$\pm ۱۳۶۲٫۷۶$	$\pm ۹۸۸٫۷۹$	$\pm ۱۲۶۰٫۹۹$	$\pm ۶۸۸٫۲۶$	$\pm ۹۰۹٫۵۸$
	μ	۵۰٫۴۳ ($\pm ۰٫۲۲$)	۸۶٫۶۵ ($\pm ۰٫۷۹$)	۱۴۹٫۸۷ ($\pm ۰٫۰۶$)	۲۵٫۶۹ ($\pm ۰٫۳۵$)	۸۷٫۷۰ ($\pm ۰٫۷۲$)	۰٫۹۹ ($\pm ۰٫۰۰۱۴$)	۲۳٫۲۲ ($\pm ۰٫۲۵$)
نرمال چوله	σ	۶٫۷۲ ($\pm ۰٫۳۰$)	۱۴۰٫۷۱ ($\pm ۰٫۵۷$)	۲۰٫۶۷ ($\pm ۰٫۷۷$)	۶٫۴۲ ($\pm ۰٫۲۵$)	۱۷٫۱۲ ($\pm ۰٫۵۳$)	۳٫۰۲ ($\pm ۰٫۰۰۱۱$)	۶٫۰۱ ($\pm ۰٫۰۸$)
	λ	-۱٫۱۰ ($\pm ۰٫۱۵$)	۱٫۷۹ ($\pm ۰٫۱۷$)	۲٫۰۶ ($\pm ۰٫۱۴$)	-۱٫۷۶ ($\pm ۰٫۱۷$)	۳٫۵۵ ($\pm ۰٫۰۸$)	۱۵۳۳٫۷۷ ($\pm ۰٫۰۰۰۰۱$)	۳٫۴۹ ($\pm ۰٫۰۸$)
	AIC	$\pm ۱۰۳۷٫۰۵$	$\pm ۱۲۵۵٫۴۱$	$\pm ۱۳۵۶٫۰۸$	$\pm ۹۷۹٫۵۰$	$\pm ۱۲۵۳٫۲۱$	$\pm ۵۷۷٫۱۷$	$\pm ۹۰۶٫۸۱$
	μ	۴۸٫۲۶ ($\pm ۰٫۷۴$)	۸۶٫۶۶ ($\pm ۰٫۸۹$)	۱۴۶٫۵۷ ($\pm ۰٫۳۷$)	۱۸٫۸۶ ($\pm ۰٫۲۰$)	۹۰٫۲۶ ($\pm ۰٫۶۲$)	۰٫۹۹ ($\pm ۰٫۰۰۱$)	۲۴٫۱۷ ($\pm ۰٫۵۰$)
تی چوله	σ	۲۵٫۸۲ ($\pm ۰٫۲۰$)	۲۱۶٫۱۴ ($\pm ۰٫۱۱$)	۲۰۰٫۳۹ ($\pm ۰٫۴۶$)	۴٫۳۲ ($\pm ۰٫۲۲$)	۱۳۶٫۰۹ ($\pm ۰٫۲۰$)	۳٫۳۹ ($\pm ۰٫۱۱$)	۱۴٫۸۴ ($\pm ۰٫۱۶$)
	λ	-۰٫۴۴ ($\pm ۰٫۶۷$)	۱٫۷۸ ($\pm ۰٫۴۴$)	۰٫۷۰ ($\pm ۰٫۹۳$)	۱٫۵۳ ($\pm ۰٫۴۶$)	۲٫۰۹ ($\pm ۰٫۴۹$)	۳۱۸۹۳۲۲۸ (± ۱۸۵۸۸)	۲٫۰۹ ($\pm ۰٫۴۹$)
	df	۹٫۶۶	۲۵۱۹٫۲۱	۸٫۱۶	۱٫۱۵	۴٫۵۹	۳٫۳۰	۳٫۹۱
	AIC	$\pm ۱۰۳۶٫۶۰$	$\pm ۱۲۵۷٫۴۱$	$\pm ۱۳۶۱٫۴۸$	$\pm ۹۵۷٫۱۶$	$\pm ۱۲۴۷٫۵۷$	$\pm ۵۵۸٫۲۷$	$\pm ۸۹۴٫۴۵$

*: فشار خون دیاستولیک، **: فشار خون سیستولیک،

+: دور کمر، ++: شاخص توده بدنی،

جدول ۶: تابع برخی از توزیع‌های چوله معروف دیگر

نرمال	چولگی	کشیدگی	توزیع انتخابی
عدم اختلاف معنی‌دار از هم‌توزیعی با توزیع نرمال	همواره عدم اختلاف معنی‌دار از چولگی توزیع نرمال	همواره عدم اختلاف معنی‌دار از کشیدگی توزیع نرمال	نرمال
اختلاف معنی‌دار از هم‌توزیعی با توزیع نرمال	عدم اختلاف معنی‌دار از چولگی توزیع نرمال	اختلاف معنی‌دار از کشیدگی توزیع نرمال	تی
	عدم اختلاف معنی‌دار از چولگی توزیع نرمال	عدم اختلاف معنی‌دار از کشیدگی توزیع نرمال	نرمال چوله
	اختلاف معنی‌دار از چولگی توزیع نرمال	عدم اختلاف معنی‌دار از کشیدگی توزیع نرمال	نرمال چوله
	اختلاف معنی‌دار از چولگی توزیع نرمال	اختلاف معنی‌دار از کشیدگی توزیع نرمال	تی چوله

مراجع

- [1] Abtahi, A., M. Towhidi and J. Behboodian (2009). An appropriate empirical version of skew-normal density. *Stat Papers*.
- [2] Anscombe, F. J. and W. J. Glynn (1983). Distribution of kurtosis statistic for normal statistics. *Biometrika*, **70(1)**, 227-234.
- [3] Arellano-Valle, R., H. Go'mez and F. Quintana (2004). A new class of skew-normal distributions. *Communications in Statistics: Theory and Methods*, **33(7)**, 1465-1480.
- [4] Arellano-Valle, R. B., H. Bolfarine and V. H. Lachos (2005). Skew-normal linear mixed models. *Journal of Data Science*, **3**, 415-438.
- [5] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist. Theory and Applications*, **12(2)**, 171-178.
- [6] Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scand. J. Statist. Theory and Applications*, **32(2)**, 159-200.
- [7] Azzalini, A. and A. Capitanio (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, **61**(Series B), 579-602.
- [8] Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **65**, 367-389.
- [9] Azzalini, A. and A. DallaValle (1996). The multivariate skew-normal distribution. *Biometrika*, **83(4)**, 715-726.
- [10] Azzalini, A., Genton, M., (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, **76**, 106-129.
- [11] Balakrishnan, N. (2002). Discussion on Skew Multivariate Models Related to Hidden Truncation and/or Selective Reporting by B. C. Arnold and R. J. Beaver. *Test*, **11**, 37-39.
- [12] Bandyopadhyay, D., V. H. Lachos, C. A. Abanto-Valle and P. Ghosh (2010). Linear mixed models for skew-normal/independent bivariate responses with an application to periodontal disease. *Statist. Med*, **29**, 2643-2655.
- [13] D'Agostino, R. B. (1970). Transformation to Normality of the Null Distribution of G1. *Biometrika*, **57(3)**, 679-681.

- [14] Dey, D. (2010). *Estimation of the Parameters of Skew Normal Distribution by Approximating the Ratio of the Normal Density and Distribution Functions*. *Applied Statistics*, CALIFORNIA, UNIVERSITY OF CALIFORNIA RIVERSIDE. PHD.
- [15] Elal-Olivero, D., H. W. Gomez and F. A. Quintana (2009). Bayesian modeling using a class of bimodal skew-elliptical distributions. *Journal of Statistical Planning and Inference*, **139(4)**, 1484-1492.
- [16] Eugene, N., C. Lee and F. Famoye (2002). Beta-normal distribution and its applications. *Commun Stat Theory Methods*, **31**, 497-512.
- [17] FERNANDEZ, C. and J. STEEL MF (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359-371.
- [18] Ferreira, J. T. A. S. and J. STEEL MF (2004). Discussion of Mc Jones, Families of distributions arising from distributions of order statistics. *Test*, **13**, 30-33.
- [19] Ferreira, J. T. A. S. and J. STEEL MF (2006). A constructive representation of univariate skewed distributions. *J Am Stat Assoc*, **101**, 823-829.
- [20] Genton, M. G. (2004). *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality*, USA, Chapman & Hall/CRC.
- [21] Gómez, H. W., O. Venegas and H. Bolfarine (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, **18(4)**, 395-407.
- [22] Gupta, A. K. (2003). Multivariate skew t-distributions. *Statistics*, **37**, 359-363.
- [23] Gupta, A. K., F. C. Chang and W. J. Huang (2002). Some skew-symmetric models. *Random Operators and Stochastic Equations*, **10(2)**, 133-140.
- [24] Gupta, A. K. and S. Nadarajah (2004). On the moments of the beta-normal distribution. *Commun Stat Theory Methods*, **33(1)**, 1-13.
- [25] Hasanlipour, P. and M. Sharafi (2010). A new generalized Balakrishnan skew-normal distribution. *Stat Papers*.
- [26] Jamalizadeh, A., A. R. Arabpour and N. Balakrishnan (2009). A generalized skew two-piece skew-normal distribution. *Stat Papers*.
- [27] Jara, A., F. Quintana and E. San Martin (2008). Linear mixed models with skew-elliptical distributions: a Bayesian approach. *Computational Statistics and Data Analysis*, **52**, 5033-5045.

- [28] Kheradmandi, A., M. Mohammadzadeh and N. Sanjari Farsipour (2010). Generalizations of the Skew t-Normal Distribution and their Properties. *Journal of Sciences, Islamic Republic of Iran*, **21(4)**, 343-351.
- [29] KIM, H. J. (2005). On a class of two-piece skew-normal distributions. *Statistics*, **39**, 537–553.
- [30] Lachos, V. H., D. K. Dey and V. G. Canchoc (2009). Robust linear mixed models with skew-normal independent distributions from a Bayesian perspective. *Journal of Statistical Planning and Inference*, **139**, 4098 - 4110.
- [31] Lee, J. C., T. I. Lin, K. J. Lee and Y. L. Hsu (2005). Bayesian analysis of Box-Cox transformed linear mixed models with ARMA(p,q) dependence. *Journal of Statistical Planning and Inference*, **133(2)**, 435-451.
- [32] Lin, T. I. and J. C. Lee (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in Medicine*, **27**, 1490–1507.
- [33] Ma, Y., M. G. Genton and M. Davidian (2004). *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality. Linear mixed models with flexible generalized skew-elliptical random effect*, Chapman & Hall, CRC Press, Boca Raton, FL,.
- [34] Montazeri, A., A. Harirchi, M. Shariati, G. Garmaroudi, M. Ebadi and A. Fateh (2003). The 12 item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version. *Health and Quality of life Outcome*, **1**, 66-70.
- [35] Mudholkar, G. S. and A. D. Hutson (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, **83(2)**, 291-309
- [36] Nadarajah, S. (2008). A new model for symmetric and skewed data. *Probab Eng Inf Sci*, **22**, 261–271.
- [37] Nadarajah, S. and S. Kotz (2003). Skewed distributions generated by the normal kernel. *Statist. Probab. Lett.*, **65(3)**, 269-277.
- [38] Nekoukhrou, V. and M. H. Alamatsaz (2011). A family of skew-symmetric-Laplace distributions. *Stat Papers*.
- [39] Patrick, R. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics*, **31**, 115–124.
- [40] Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *J. Appl. Statist*, **27(7)**, 859-870.
- [41] Sahu, S. K., D. K. Dey and M. D. Branco (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, **31**, 129–150.

- [42] SHARAFI, M. and J. BEHBOUDIAN (2006). "New skew normal density." *Journal of Statistical Research Of Iran (JSRI)*, **3(1)**, 47-61.
- [43] Spinosa, J. S. (2008). *Inference for the Multiparameter Skew Normal Distribution*, Applied statistic. California, University of California Riverside. PHD.
- [44] Tsung, I. L. and J. C. Lee (2003). On modelling data from degradation sample paths over time. *Australian and New Zealand Journal of statistics*, **45(3)**, 257-270.
- [45] Wang, J. and M. Genton (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference*, **136**, 209–220.
- [46] Yadegari, I., A. GERAMI and M. J. Khaledi (2008). A generalization of the Balakrishnan skew-normal distribution. *Stat Probab Lett*, **78**, 1165–1167.

ضمیمه

فرمول‌های تکمیلی

$$f_{\lambda_1, \lambda_2, \rho}(x) = \frac{\pi}{\cos^{-1}\left(\frac{-(\rho + \lambda_1 \lambda_2)}{\sqrt{1 + \lambda_1^2} \sqrt{1 + \lambda_2^2}}\right) + \cos^{-1}\left(\frac{-(\rho - \lambda_1 \lambda_2)}{\sqrt{1 + \lambda_1^2} \sqrt{1 + \lambda_2^2}}\right) + \pi \tan^{-1}(\lambda_2)} \times \varphi(x) \Phi_{\pi}(\lambda_1 x, \lambda_1 |x|, \rho), \quad -\infty < x, \lambda_1, \lambda_2 < +\infty \quad (23)$$

$$f_{\lambda_1, \lambda_2}(x) = \frac{1}{\pi + \pi \tan^{-1}(\lambda_2)} \varphi(x) \Phi(\lambda_1 x) \Phi(\lambda_2 |x|), \quad -\infty < x, \lambda_1, \lambda_2 < +\infty. \quad (24)$$

فرم داده‌ها در نرم افزار R

```

Age<-c(48.5, 53.5, 41.5, 48, 48.5, 39, 38.5, 54.5, 42.5, 41.5, 45, 44, 54, 45.5,
50, 51, 48.5, 40.5, 49.5, 43.5, 47.5, 50, 48, 50.5, 49.5, 39.5, 58.5, 51.5, 47.5,
51.5, 45.5, 46.5, 50.5, 49.5, 50.81, 45.5, 48.5, 40.5, 43.5, 40.5, 52.5, 33.48,
52.5, 50, 50.5, 53, 48.5, 49, 51, 51, 43.5, 48.5, 44.5, 50.42, 40.33, 41.5, 42.5,
43, 48, 49.5, 44.5, 47.01, 46, 43.5, 38.5, 45.5, 44, 41, 46.5, 57.5, 47.5, 46.5, 42,
41.5, 25.5, 32.5, 49.5, 59, 54, 48, 49.5, 48.5, 47.5, 47.5, 58.5, 51, 55.5, 51.5, 47,
48.5, 45.5, 47, 40, 40.5, 48.5, 50.5, 39.5, 59.5, 46, 42.5, 46.5, 44, 46.5, 42, 49.5,
40.5, 41.5, 40.5, 48, 39.5, 41.5, 41, 41, 43.5, 43.5, 47.5, 57, 47.5, 49, 47, 41.94,
46.5, 48, 48, 40.41, 32.5, 36.5, 49.5, 54, 48.5, 50.5, 55, 54.5, 50, 49.5, 52.5, 53,
47.5, 48.5, 43.5, 53.5, 43, 41.5, 48, 41.5, 45.5, 46.04, 48.86, 43.5, 41.5, 43,
48.5, 43.5, 48.5, 46.5, 45.5, 43.39, 40, 44.5, 50.5, 47, 60, 39, 40, 47.5, 36.5)

bmi<-c(29.07, 31.22, 37.08, 29.22, 28.82, 28.44, 25.5, 21.15, 30.84, 27.43,
27.76, 28.04, 25.87, 28.34, 26.07, 24.15, 22.44, 25.87, 24.96, 26.01, 25.52,
27.37, 23.11, 34.1, 27.31, 25.66, 22.68, 30.24, 28.85, 27.86, 27.8, 25.15, 24.89,
32.19, 24.54, 28.02, 29.42, 33.97, 24.49, 27.13, 30.14, 27.13, 27.75, 33.58,
27.24, 22.78, 26.13, 26.15, 23.1, 25.61, 28.05, 31.55, 32.19, 25.64, 25.52, 27.99,
28.07, 23.45, 25.58, 35.03, 23.19, 27.99, 28.91, 25.95, 26.47, 25.93, 24.44,
26.43, 25.38, 26.02, 28.51, 25.41, 23.56, 18.8, 26.96, 28.73, 31.15, 26.52, 30.05,
24.91, 24.96, 25.64, 23.03, 49.02, 31.27, 24.24, 26.73, 28.3, 24.61, 27.93, 26.55,
26.82, 28.2, 27.08, 32.7, 23.14, 25.67, 28.52, 29.81, 39.99, 30.12, 27.28, 25.86,
23.78, 29.22, 23.65, 27.55, 25.94, 26.97, 33.89, 28.68, 25.36, 26.51, 25.94,
31.73, 27.57, 31.99, 23.17, 23.79, 22.3, 27.8, 25.96, 33.3, 29.37, 27.55, 39.12,
37.93, 30.46, 24.18, 25.88, 23.58, 25, 25.27, 24.36, 26.99, 40.42, 23.85, 25.61,
21.13, 29.04, 28.87, 25.49, 26.51, 25.96, 32.67, 25.72, 28.73, 27.53, 23.42,
36.97, 27.24, 37.52, 31.18, 24.54, 28.56, 28.06, 29.37, 24.47, 30.39, 24.56,
33.28, 27.26, 33.66, 25.85, 25.27, 20.98)

DBP<-c(72, 74, 76, 78, 78, 78, 80, 80, 80, 80, 83, 84, 84, 84, 84, 84, 84, 84,
84, 84, 84, 84, 86, 86, 86, 86, 88, 88, 88, 88, 88, 88, 90, 90, 90, 90, 90, 90,
90, 90, 90, 90, 90, 90, 90, 90, 90, 90, 90, 90, 92, 92, 92, 92, 92, 92, 92,
92, 94, 94, 94, 94, 94, 94, 94, 94, 94, 94, 94, 94, 94, 94, 94, 94, 94, 96, 96, 96,
96, 96, 96, 96, 96, 97, 98, 98, 98, 98, 98, 98, 98, 100, 100, 100, 100, 100, 100,
100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100,
100, 102, 102, 102, 104, 104, 104, 104, 104, 104, 104, 104, 104, 104, 104, 104,
105, 106, 106, 107, 108, 108, 108, 108, 108, 108, 110, 110, 110, 110, 110, 110,
110, 110, 110, 112, 114, 114, 114, 114, 114, 114, 114, 118, 118, 120, 120, 120, 120, 122,
134)

SBP<-c(122, 114, 120, 118, 128, 160, 120, 126, 120, 120, 143, 130, 152, 120,
142, 124, 136, 130, 140, 128, 136, 144, 134, 124, 146, 148, 156, 144, 142, 130,
144, 122, 132, 150, 142, 150, 124, 126, 134, 136, 138, 144, 140, 130, 136, 138,
140, 150, 128, 150, 148, 120, 138, 140, 142, 142, 138, 138, 142, 132, 142, 138,
134, 136, 152, 132, 140, 126, 144, 138, 136, 134, 134, 144, 144, 146, 134, 132,
124, 123, 142, 144, 148, 156, 136, 146, 131, 157, 186, 160, 144, 160, 166, 138,
132, 150, 154, 140, 144, 144, 174, 158, 154, 150, 154, 144, 160, 140, 150, 132,
148, 142, 152, 150, 138, 140, 150, 160, 152, 144, 152, 146, 144, 140, 152, 144,
144, 154, 150, 144, 152, 154, 160, 127, 146, 138, 151, 132, 146, 140, 150, 164,
146, 142, 150, 150, 176, 158, 180, 158, 150, 170, 160, 170, 182, 174, 172, 152,
178, 146, 160, 182, 170, 170, 172, 180)

Experience<-c(28, 20, 19, 21, 27, 19, 27, 17, 19, 21, 20, 20, 19, 26, 26, 26, 25,
20, 19, 20, 24, 20, 26, 30, 31, 26, 19, 27, 26, 22, 20, 26, 19, 26, 26, 20, 19,
18, 25, 28, 20, 19, 19, 19, 20, 20, 26, 25, 20, 5, 25, 23, 20, 20, 19, 24, 19, 20,
20, 31, 25, 25, 20, 20, 23, 15, 14, 19, 21, 19, 26, 22, 4, 26, 19, 26, 20, 9, 27, 26,
20, 19, 19, 19, 17, 20, 20, 19, 25, 25, 19, 19, 19, 4, 19, 29, 8, 24, 19, 20, 19,
19, 20, 20, 27, 27, 26, 20, 20, 28, 18, 25, 25, 22, 21, 23, 20, 20, 19, 20, 4, 19,
22, 20, 25, 26, 19, 20, 17, 22, 20, 17, 27, 31, 19, 25, 21, 20, 19, 20, 24, 11, 19,
20, 26, 30, 27, 23, 22, 20, 20, 20, 28, 22, 25, 20, 25, 20, 18, 19, 26, 16, 25, 19)

WC<-c(107, 93, 96, 105, 92.48, 86, 113.77, 87, 98, 86, 97, 104, 84, 94, 96, 90.5,
98, 91, 99, 109, 91, 102, 97, 112.47, 104, 114, 89, 98, 98, 102, 99, 127, 103,
99.47, 99, 102, 93, 92, 92, 99, 86.44, 94, 97, 103, 109, 113.5, 98.5, 97.39, 85,
105, 90, 105, 99, 90, 128.5, 98, 101.5, 93, 101, 84.5, 101.5, 93, 100, 105, 94,
90, 95, 97, 90, 95, 97, 94, 84, 86, 90, 100.11, 93, 120, 133.5, 110.5, 90, 99, 102,
111, 99.5, 118, 89, 105, 95, 101, 90, 116, 101, 100.5, 94, 127, 122.5, 105,
92.43, 93.5, 80, 94, 95, 113, 95, 107, 99.5, 97, 154, 102, 106, 130, 118, 94, 98,
117, 132, 106.54, 120, 102.6, 128.8, 101, 79, 100, 107, 91.5, 100, 100.5, 93,
103.5, 100.5, 92, 111, 105, 102, 104.5, 102, 101, 88, 96, 125, 92, 99.5, 91,
126.79, 98, 99.52, 108, 104, 92, 93.5, 90, 95, 99.5, 96.5, 82.5, 102, 90, 99, 92,
83, 110.32, 107.5, 104, 92, 99)

```