

کاربرد مدل رگرسیون پواسنی تعمیم یافته در تحلیل داده‌های باروری زنان روستایی استان فارس

نجف زارع^۱، مهرباب صیادی^۲، الهام رضائیان فرد^۳، هاله قائم^۴

^۱ دانشیار گروه آمار زیستی، دانشکده پزشکی، دانشگاه علوم پزشکی شیراز، ایران

^۲ مشاور و تحلیلگر آماری، کارشناس ارشد آمار زیستی، دانشگاه علوم پزشکی شیراز، ایران

^۳ کارشناس بهداشت خانواده، دانشگاه علوم پزشکی شیراز، ایران

^۴ مربی گروه اپیدمیولوژی، دانشکده بهداشت و تغذیه، دانشگاه علوم پزشکی شیراز، ایران

نویسنده رابط: مهرباب صیادی؛ دانشگاه علوم پزشکی شیراز، دانشکده پزشکی، گروه آمار زیستی. همراه: ۰۹۱۷۱۳۳۲۹۱۷، شماره: ۰۲۳۴۷۳۱۵-۰۷۱۱، پست الکترونیک:

sayadi_me@yahoo.com

تاریخ دریافت: ۱۳۸۸/۴/۴؛ پذیرش: ۱۳۸۸/۱۱/۳

مقدمه و اهداف: مدل‌بندی آماری تغییرات مشاهده شده در داده‌ها را از طریق معادلات ریاضی تبیین می‌نماید. در حالتی که متغیر پاسخ گسسته باشد مدل پواسن مورد استفاده قرار می‌گیرد و در صورتی که شرایط مدل پواسن برقرار نباشد، بهتر است از تعمیم یافته آن استفاده کرد. لذا هدف از این مطالعه، تاکید و توجه به ساختار داده‌ها، معرفی مدل رگرسیون پواسنی تعمیم یافته و بکار بردن این مدل جهت برآورد ضرایب عوامل مؤثر بر تعداد فرزندان و مقایسه آن با مدل رگرسیون پواسن معمولی است.

روش کار: در این مطالعه ضمن معرفی مدل پواسنی تعمیم یافته کاربرد آن در تحلیل داده‌های باروری بکار رفته است. این داده‌ها از یک نمونه ۱۰۱۹ نفری زنان روستایی استان فارس بصورت مقطعی و با استفاده از روش نمونه‌گیری طبقه‌بندی بدست آمد. متغیر تعداد فرزند از زن متولد شده یک زن به عنوان متغیر پاسخ شمارشی جهت کاربرد مدل در نظر گرفته شده است.

نتایج: میانگین فرزندان هر زن $2/88 \pm 4/3$ بود. مقدار آماره آزمون، $\log\text{-likelihood}$ برای مدل پواسنی معمولی $1950/93-$ و برای مدل پواسنی تعمیم یافته $1946/93-$ بود.

نتیجه‌گیری: نتایج نشان داد که داده‌ها Over Dispersion دارد. و بر اساس معیارهای انتخاب بهترین مدل، مدل پواسنی تعمیم یافته جهت تحلیل این داده‌ها مناسب است و می‌تواند ضرایب عوامل مؤثر بر تعداد فرزند را دقیق‌تر برآورد نماید.

واژگان کلیدی: داده‌های باروری، توزیع پواسن، توزیع پواسن تعمیم یافته، تعداد فرزند

مقدمه

پواسن، معادل بودن میانگین و واریانس متغیر پاسخ است اگر این شرط برقرار نباشد مدل توزیع پواسن تعمیم یافته مناسب خواهد بود (۹-۷).

داده‌های شمارشی معمولاً دارای توزیع نرمال نمی‌باشند و چوله به راست یا چپ هستند. بنابراین روش‌های آماری مبتنی بر توزیع نرمال برای تحلیل چنین داده‌هایی مناسب نیست، در این حالت استفاده از مدل‌های تعمیم یافته مناسب‌تر است (۱۰). یکی از فرضیات مدل رگرسیون پواسن معمولی این است که احتمال اتفاق هر پیشامد در هر دوره، مستقل از هم است اما در بعضی داده‌های شمارشی وقوع اولیه یک پیشامد ممکن است باعث افزایش یا کاهش احتمال پیشامد در آینده شود. در داده‌های باروری تعداد

مدل‌بندی آماری یکی از روش‌های تبیین تغییرات در داده‌های مشاهده شده از طریق معادلات ریاضی است که با استفاده از آن می‌توان به چگونگی تغییرات داده‌ها پی برد. در حالتی که متغیر پاسخ (وابسته) گسسته و نامنفی است مدل پواسنی مورد استفاده قرار می‌گیرد (۴-۱).

در سال‌های اخیر مدل پواسنی برای پاسخ‌های شمارشی بطور فراوان مورد استفاده قرار گرفته است. Winkelmann و King مدل‌های شمارشی تعمیم یافته را برای توزیع‌های پواسن، دو جمله‌ای و دو جمله‌ای منفی توسعه دادند و نشان دادند که مفروضات مدل پواسن معمولی برای بعضی از داده‌های شمارشی محدودیت ایجاد می‌کند (۵، ۶). شرط اصلی استفاده از مدل

مورد آن برقرار باشد.

الف: در هر لحظه زمانی فقط یک پیشامد داشته باشیم.

ب: احتمال اتفاق هر پیشامد در هر دوره مستقل از هم باشند.

ج: در لحظه شروع هر دوره پیشامدی نداشته باشیم آنگاه Y_i دارای توزیع پواسن است که تابع احتمال آن به شکل زیر است:

$$f(y_i, \lambda_i, t_i) = \frac{e^{-\lambda_i t_i} (\lambda_i t_i)^{y_i}}{y_i!}; \quad t_i > 0 \quad \lambda_i > 0$$

امید ریاضی و واریانس توزیع پواسن با هم برابرند و هر دوی آن‌ها λ_i است یعنی $E(Y_i) = Var(Y_i) = \lambda_i$ و مدل رگرسیون پواسن: $\lambda_i = \exp(x_i \beta)$ که بردار x_i بردار K بعدی از متغیرهای برون‌زا می‌باشد و β بردار $1 \times K$ بعدی پارامترها است.

توزیع پواسن تعمیم یافته:

ساختار توزیع پواسن تعمیم یافته به شرح زیر است:

اگر فرضیات توزیع پواسن برقرار نباشد واریانس این توزیع با میانگین آن برابر نمی‌شود. وقتی که واریانس بزرگتر از میانگین باشد **Over Dispersion** پدیدار می‌شود که منجر به کم برآوردی خطای معیار شده و مقدار آماره آزمون (ملاک آزمون) هر کدام از ضرایب را افزایش می‌دهد. وقتی واریانس کوچک‌تر از میانگین باشد **Under Dispersion** ظاهر شده که باعث بیش برآوردی خطای معیار و در نتیجه کاهش مقدار آماره آزمون (ملاک آزمون) هر کدام از ضرایب می‌شود.

فرض کنید که Y_i متغیرهای تصادفی باشند که دارای توزیع پواسن تعمیم یافته با **Over Dispersion** باشد. در این صورت تابع احتمال آن‌ها بصورت زیر است:

$$f(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1 + \alpha \mu_i} \right)^{y_i} \frac{(1 + \alpha \mu_i)^{y_i - 1}}{y_i!} \exp \left(- \frac{\mu_i (1 + \alpha \mu_i)}{1 + \alpha \mu_i} \right)$$

که $y_i = 0, 1, 2, \dots$ تعداد پیشامدها (در داده‌های باروری تعداد فرزندان هر زن) را نشان می‌دهد. α پارامتر **Dispersion** می‌باشد

$$E(y_i) = \mu_i$$

$$Var(y_i) = \mu_i (1 + \alpha \mu_i)^2$$

$$\mu_i = \exp(x_i, \beta)$$

که x_i بردار K بعدی از متغیرهای تعیین کننده است و β بردار $1 \times K$ بعدی پارامترهای رگرسیونی هستند و برای برآورد آن‌ها شبیه رگرسیون لجستیک از روش درست نمایی ماکزیمم استفاده شده و برای خطی سازی مدل از تابع اتصال لگاریتمی بهره

فرزندها از یک زن به زن دیگر مستقل است. ولی ممکن است تعداد فرزندان یک زن مستقل از هم نباشد. یعنی زوجینی که در یک زمان معین این تعداد فرزند را کافی نمی‌دانند تصمیم به داشتن فرزند بعدی می‌گیرند. اگر فرض بالا برقرار نباشد استفاده از مدل پواسن منجر به برآورد نا صحیحی از ضرایب رگرسیونی می‌گردد. توزیع‌های پواسن تعمیم یافته در حالت‌های عملی و کاربردی بسیار مطرح است و بیشتر در خصوص داده‌های که پراکندگی دارند مورد استفاده قرار می‌گیرد (۱۱، ۱۲، ۱۳).

در داده‌های باروری تعداد فرزندها به عنوان متغیر پاسخ، پراکندگی زیادی دارد. یک راهبرد عملی جهت تجزیه و تحلیل چنین داده‌هایی حذف مقادیر خیلی پرت با استفاده از تصحیح کننده‌های آماری است که این راهبرد جهت تبیین متغیر پاسخ با استفاده از متغیرهای مستقل، نتایج گمراه کننده‌ای را ارائه می‌دهد. ولی استفاده از مدل‌های پواسن تعمیم یافته در این خصوص مناسب‌تر است (۱۴).

نتیجه اینکه توجه به ساختار داده‌ها و یافتن توزیع مناسب برای متغیر پاسخ از اصول مهم مدل‌بندی آماری است. بعد از در نظر گرفتن توزیع مناسب متغیر پاسخ، برای تبیین این متغیر بر اساس متغیرهای مستقل از رگرسیون آن استفاده خواهد شد.

از آنجا که بعضی از داده‌های سیستم بهداشتی و درمانی ماهیت شمارشی دارند مدل پواسن تعمیم یافته جهت تجزیه و تحلیل چنین داده‌هایی مفید می‌باشد. تا کنون در ایران تجزیه و تحلیل چنین داده‌هایی بدون توجه به ساختار اصلی داده‌ها بوده است، بخصوص داده‌های باروری، که به لحاظ مشکلات افزایش جمعیت و اثرات نامطلوبی که این افزایش بر جامعه می‌گذارد از اهمیت بیشتری برخوردار است. در بعد تحلیل آماری این کار مستلزم این است که متغیرهای مؤثر بر باروری با دقت بیشتری تبیین و تفسیر شود. لذا هدف اصلی این مطالعه نشان دادن پتانسیل مدل‌های پواسن تعمیم یافته در تحلیل دقیق داده‌های باروری و مشابه آن است. در داده‌های حاضر تعداد فرزند به عنوان متغیر پاسخ و برخی از عوامل دموگرافیک به عنوان متغیرهای مستقل در نظر گرفته شده است.

روش کار

توزیع پواسن

ساختار توزیع پواسن معمولی به شکل زیر است:

اگر Y_i متغیرهای تصادفی مستقل باشند که سه فرض زیر در

جدول شماره ۱- توزیع فراوانی تعداد فرزندان، میانگین و انحراف معیار متغیرهای دموگرافیک

تعداد فرزندان	توزیع فراوانی		سن زن		سن ازدواج زن		طول دوره زناشویی	
	تعداد	%	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار
۰	۱۳	۱/۳	۳۲/۲	۱۰/۸	۲۰/۱	۶/۸	۱۲/۱	۸/۸
۱	۱۹۳	۱۸/۹	۲۳	۵/۱	۱۸/۱	۳/۴	۴/۹	۴/۱
۲	۱۴۸	۱۴/۵	۲۶/۸	۵/۳	۱۸/۲	۳/۷	۸/۶	۴/۲
۳	۱۲۴	۱۲/۲	۳۰/۳	۵/۲	۱۷/۸	۲/۷	۱۲/۵	۴/۷
۴	۱۲۲	۱۲	۳۳/۲	۵/۲	۱۷/۶	۳/۶	۱۵/۵	۴/۵
۵	۸۱	۷/۹	۳۶/۴	۵/۵	۱۷/۶	۳/۵	۱۸/۷	۴/۹
۶	۸۷	۸/۵	۳۸/۹	۴/۴	۱۶/۸	۲/۸	۲۲	۴/۳
۷	۸۹	۸/۷	۴۱/۷	۴/۵	۱۷/۳	۳	۲۴/۴	۴
۸	۶۷	۶/۶	۴۳/۶	۴/۷	۱۷/۲	۳/۳	۲۶/۶	۴/۷
۹	۴۴	۴/۳	۴۵/۲	۴	۱۶/۸	۲/۶	۲۸/۴	۳/۷
۱۰ و +	۵۱	۵	۴۴/۲	۳/۹	۱۵/۸	۲/۲	۲۸/۳	۴/۱
جمع	۱۰۱۹	۱۰۰	۳۳/۲	۹	۱۷/۶	۳/۳	۱۵/۶	۹/۲

توجهی از زنان دارای ۱۰ فرزند یا بیشتر هستند (۵/۱ درصد). همچنین میانگین و انحراف معیار متغیرهای سن زن، سن ازدواج زن و طول دوره زناشویی آورده شده است. چنانچه در جدول نیز مشاهده می شود با افزایش سن زن، تعداد فرزندان افزایش می یابد که این امری بدیهی است چرا که زنان مسن تر فرصت کافی برای داشتن فرزندان بیشتر داشته اند و لذا از وارد کردن سن زن و متغیر دیگر وابسته به زمان از جمله طول دوره زناشویی در این مدل ها صرف نظر گردید.

جدول شماره ۲ بیانگر تحلیل مدل پواسن معمولی و پواسنی تعمیم یافته و مقایسه آن با مدل دو جمله ای منفی است. در این جدول برآورد ضرایب هر کدام از متغیرهای مستقل همراه با خطای استاندارد آن و مقدار احتمال (pvalue) آورده شده است. در توزیع پواسنی تعمیم یافته و دو جمله ای منفی مقدار α نشان دهنده میزان پراکندگی (Dispersion) نیز گزارش شده است.

نتایج نشان داد که مقدار α مثبت است و اختلاف معنی داری با صفر دارد ($P < 0/001$) و نشان می دهد که داده ها دارای ساختار Over Dispersion هستند. مقدار log-likelihood نیز برای هر سه مدل بیان شده است. مقدار کمتر این آماره از نظر قدر مطلق دلالت بر بهتر بودن مدل برازشی دارد. بر اساس جدول شماره ۲ در هر سه مدل متغیرهای تحصیلات زن و همسر، شغل زن، سن ازدواج زن، وضعیت اقتصادی و متوسط فاصله گذاری بین فرزندان معنی دار شدند. اضافه بر این متغیر متوسط شیردهی فرزند فقط در مدل پواسنی معنی دار گردید و شاخص نسبت فرزند پسر به کل فرزندان در هیچ کدام از مدل ها معنی دار نگردید.

می جوید (۶،۱۳،۱۵). نیکوئی برآزش مدل ها در توزیع با استفاده از آماره log-likelihood آزمون می گردد.

کاربرد: داده های مورد استفاده، مطالعه ای بود که در سال ۸۵ انجام گرفت و جمعیت تحت مطالعه تمام زنان شوهر نموده ۴۹-۱۵ ساله مناطق روستایی استان فارس بود که در سن باروری قرار داشتند. یک نمونه ۱۰۱۹ نفری بصورت مقطعی و با استفاده از نمونه گیری طبقه بندی از جمعیت مذکور انتخاب گردید. که در اینجا گروه های سنی ۵ ساله به عنوان طبقه در نظر گرفته شد و زنان ۱۵ تا ۴۹ سال در ۷ طبقه قرار گرفتند. سپس سعی شد در هر طبقه بطور مساوی، با استفاده از روش نمونه گیری سیستماتیک و از روی پرونده های موجود در خانه های بهداشت تعداد نمونه های مورد نظر انتخاب گردند. روش جمع آوری داده ها پرسشنامه خود ساخته ای بود که اطلاعات بصورت مصاحبه ای توسط کارشناسان آموزش دیده بهداشت خانواده در آن ثبت می گردید. این پرسشنامه شامل اطلاعات دموگرافیک (سن زن، سن شوهر، قد زن، تحصیلات زن، شغل زن، تحصیلات شوهر، سن ازدواج زن، وضعیت اجتماعی و اقتصادی) و تعداد فرزندان بود.

یافته ها

در این مطالعه داده های ۱۰۱۹ زن روستایی مورد بررسی قرار گرفت. میانگین فرزند ان $4/3 \pm 2/88$ بود. همچنین میانگین تعداد فرزندان ۲/۱۸ بدست آمد. جدول شماره ۱ توزیع تعداد فرزندان را نشان می دهد. بر اساس این جدول مد توزیع فراوانی تعداد فرزندان ۲ است. حدود ۴۱ درصد زنان بیش از ۴ فرزند و درصد قابل

جدول شماره ۲- تعیین عوامل مؤثر بر تعداد فرزندان با استفاده از روش‌های رگرسیون پواسنی و پواسنی تعمیم یافته و دوجمله‌ای منفی

متغیرها	پواسن			پواسن تعمیم یافته			دوجمله‌ای منفی		
	برآورد	خطای استاندارد	احتمال	برآورد	خطای استاندارد	احتمال	برآورد	خطای استاندارد	احتمال
تحصیلات زن	-۰/۰۱۰۵	۰/۰۰۵۰	۰/۰۳۵	-۰/۰۱۰۶	۰/۰۰۵۳	۰/۰۴۷	-۰/۰۱۰۶	۰/۰۰۵۳	۰/۰۴۷
شغل زن	۰/۱۸۸۲	۰/۰۴۹۴	≤ ۰/۰۰۱	۰/۱۹۱۵	۰/۰۵۳۶	≤ ۰/۰۰۱	۰/۱۹۱۵	۰/۰۵۳۶	≤ ۰/۰۰۱
تحصیلات همسر	-۰/۰۷۱۸	۰/۰۰۴۱	≤ ۰/۰۰۱	-۰/۰۷۲۰	۰/۰۰۴۴	≤ ۰/۰۰۱	-۰/۰۷۲۰	۰/۰۰۴۴	≤ ۰/۰۰۱
سن ازدواج زن	-۰/۰۳۸۶	۰/۰۰۵۲	≤ ۰/۰۰۱	-۰/۰۳۸۶	۰/۰۰۵۵	≤ ۰/۰۰۱	-۰/۰۳۸۶	۰/۰۰۵۵	≤ ۰/۰۰۱
وضعیت اقتصادی	۰/۳۲۹۷	۰/۰۲۴۹	≤ ۰/۰۰۱	۰/۳۳۶۱	۰/۰۲۶۹	≤ ۰/۰۰۱	۰/۳۳۶۱	۰/۰۲۶۹	≤ ۰/۰۰۱
نسبت پسر به کل	-۰/۰۵۵۶	۰/۰۵۵۵	۰/۳۱۶	-۰/۰۵۵۸	۰/۰۵۹۹	۰/۳۵۲	-۰/۰۵۵۸	۰/۰۵۹۹	۰/۳۵۲
متوسط زمان شیردهی	-۰/۰۰۵۴	۰/۰۰۲۷	۰/۰۴۷	-۰/۰۰۵۲	۰/۰۰۲۹	۰/۰۷۶	-۰/۰۰۵۲	۰/۰۰۲۹	۰/۰۷۶
متوسط فاصله گذاری	-۰/۰۱۲۰	۰/۰۰۱۴	≤ ۰/۰۰۱	-۰/۰۱۲۱	۰/۰۰۱۵	≤ ۰/۰۰۱	-۰/۰۱۲۱	۰/۰۰۱۵	≤ ۰/۰۰۱
مقدار ثابت	۲/۲۱	۰/۱۳۷۸	≤ ۰/۰۰۱	۲/۲	۰/۱۴۷۸	≤ ۰/۰۰۱	۲/۲	۰/۱۴۷۸	≤ ۰/۰۰۱
$\ln(\alpha)$				-۳/۴۷	۰/۳۹۱	≤ ۰/۰۰۱	-۳/۴۷	۰/۳۹۱	≤ ۰/۰۰۲
شاخص نیکویی برازش	Log-likelihood = -۱۹۵۰/۹۳			Log-likelihood = -۱۹۴۶/۹۳			Log-likelihood = -۱۹۴۶/۹۳		
شاخص نیکویی برازش	CAIC=۳۹۳۷/۵			CAIC=۳۹۳۳/۴۶			CAIC=۳۹۳۳/۴۶		

بحث

هدف این مطالعه نشان دادن اهمیت توجه به ساختار داده‌ها (داده‌های باروری و مشابه) و برازش مدل مناسب جهت چنین داده‌هایی بود. نتایج مطالعه نشان داد که میانگین و واریانس متغیر پاسخ (تعداد فرزندان) معادل نیست و واریانس بیشتر از میانگین است و میانگین بیشتر از نما است بنابراین داده‌ها چوله به راست هستند. همچنین معنی‌دار بودن مقدار α نشان‌دهنده این است که داده‌ها ساختار Over Dispersion دارد.

آماره آزمون log-likelihood محاسبه گردید. مقدار این آماره از نظر قدر مطلق برای مدل پواسنی تعمیم یافته و دوجمله‌ای منفی کمتر از مدل پواسنی بود (جدول شماره ۲) و نشان می‌دهد که بکار بردن توزیع پواسنی تعمیم یافته و همچنین دو جمله‌ای منفی جهت این داده‌ها مناسب‌تر است.

مطالعات دیگر نیز در چنین مواقعی رگرسیون پواسنی تعمیم یافته را پیشنهاد کرده‌اند و نشان داده‌اند که مدل پواسنی مناسب چنین داده‌هایی نیست (۱،۲،۴). در ایران، در تحلیل مشابه بروی عوامل مؤثر بر حاملگی توسط کریمی و همکار نیز مدل‌های پواسنی معمولی را مناسب داده‌های با چنین ساختاری نیافتند و مدل دوجمله‌ای منفی پیشنهاد نمودند (۱۷).

در برازش سه مدل بر داده‌های متغیرهای تحصیلات زن و همسر، شغل زن، سن ازدواج زن، وضعیت اقتصادی و متوسط فاصله‌گذاری

بین فرزندان معنی‌دار شد. در نتیجه این عوامل بر تعداد فرزندان مؤثر هستند. تفسیر اینکه بطور مثال به ازای یک سال افزایش تحصیل زن، تعداد فرزندان حدود ۱ درصد کاهش می‌یابد و یا به ازای یک سال افزایش تحصیل همسر، تعداد فرزندان حدود ۷ درصد کاهش می‌یابد. اضافه بر این متغیر متوسط شیردهی فرزند فقط در مدل پواسنی معنی‌دار گردید. شاخص نسبت فرزند پسر به کل فرزندان در هیچ کدام از مدل‌ها معنی‌دار نشد. ضرایب و خطای استاندارد ضرایب در دو مدل پواسن و پواسن تعمیم یافته اختلافاتی با هم داشتند. در صورتیکه در دو مدل پواسن تعمیم یافته و دو جمله‌ای منفی این اختلاف‌ها وجود نداشت.

در مطالعه‌های مشابه بر داده‌های باروری (wang and famoye, 1997) مدل‌های پواسنی با پواسنی تعمیم یافته مقایسه شد که داده‌ها ساختار under Dispersion داشتند و در هر دو مدل تحصیلات زوجین، شاغل بودن زن، وضعیت اقتصادی، و محل زندگی بر تعداد فرزندان مؤثر بودند ولی ضرایب دو مدل با هم متفاوت بود (۲). در مطالعات انجام شده مدل‌های دیگری بروی داده‌های مشابه برازش شده است ولی پیشنهاد شده در صورتیکه داده‌ها Dispersion دارند از مدل‌های تعمیم یافته استفاده گردد (۱،۴،۶،۱۶).

نتیجه‌گیری

در این مطالعه مشخص شد که مدل رگرسیون پواسنی تعمیم

تشکر و قدردانی

از معاونت محترم پژوهشی دانشگاه علوم پزشکی شیراز به خاطر حمایت مالی این طرح متشکریم. درضمن این مطالعه قسمتی از پایان نامه کارشناسی ارشد آقای مهربان صیادی است.

یافته بر داده‌های باروری بهتر از مدل رگرسیون پواسنی معمولی پرازش می‌گردد لذا توجه به ساختار متغیر پاسخ در اولویت مدل‌بندی آماری قرار دارد. پیشنهاد می‌گردد در تجزیه و تحلیل و تفسیر داده‌های مشابه چنانچه داده‌ها شمارشی باشند و شرایط مدل پواسنی برقرار نباشد از تعمیم یافته آن و یا دو جمله‌ای منفی استفاده گردد.

منابع

- 1- Famoye F, Wulu JT, Singh KP. On the Generalized Poisson Regression Model with an Application to Accident Data; *Journal of Data Science* 2, 287-95.
- 2- Wang W, Famoye F. Modeling household fertility decisions with generalized Poisson regression; *J Popul Econ* 1997; 10: 273-83.
- 3- Skrondal A, Rabe S. Some applications of generalized linear latent and mixed models in epidemiology: repeated measures, measurement error and multilevel modeling. *Norse Epidemiology* 2003; 13: 265-78.
- 4- Wang K, Kelvin KWY, Lee AH. A zero-inflated Poisson mixed model to analyze diagnosis related groups with majority of same-day hospital stays. *Compute Methods Programs Biomed* 2002; 68: 195-203.
- 5- King G. Variance specification in event count models: From restrictive assumptions to generalized estimator. *American journal of political science* 1989; 33: 762-94.
- 6- Winkelmann R. and Zimmerman F. Count data models for demographic data. *Mathematical population studies* 1994; 4: 205-21.
- 7- Ng SK, Yau KKW, Lee AH. Modeling inpatient length of stay by hierarchical mixture regression via the EM algorithm. *Math Compute Model* 2003; 37: 365-75.
- 8- Karlis D, Xekalaki E. Mixed Poisson distributions. *Int Stat Re* 2005; 73: 35-58.
- 9- Wulu JT, Singh KP, Famoye F, McGwin G. Regression analysis of count data. *Journal of the Indian Society of Agricultural Statistics* 2002; 55: 220-31.
- 10- Marazzi A, Paccaud F, Rueux C, et al. Fitting the distributions of length of stay by parametric models. *Med Care* 1998; 36: 915-27.
- 11- Marshall A. Length of stay-based patient flow models: Recent developments and future directions. *Health Care Manage Sci* 2005; 8: 213-20.
- 12- Consul PC. *Generalized Poisson Distributions: Properties and Applications*. Marcel Dekker 1989.
- 13- Consul PC, Famoye F. Generalized Poisson regression model. *Communications in Statistics, Theory and Methods* 1992; 21: 89-109.
- 14- Famoye F. Restricted generalized Poisson regression model. *Communications in Statistics, Theory and Methods* 1993; 22, 1335-54.
- 15- Xio J, Lee A, Vemuri S. Mixture distribution analysis of length of hospital stay for efficient funding. *J Socio- Econ Plan Sci* 1999; 33: 39-59.
- 16- Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; 34: 1-14.
- 17- Wulu JT, Singh KP, Famoye F, McGwin G. Regression analysis of count data. *Journal of the Indian Society of Agricultural Statistics* 2002; 55: 220-31.
- 18- Karimi Sh, Kazemnejad A. Application of negative binomial regression model in determining the effective factors of unwanted pregnancy. MA thesis, Tarbiat Modares university 2002.
- 19- Shojaei Tehrani H, Ebadie Fard Azar F. *Population, Family planning and fertility health*, 1st edition, Tehran, Majed: 1999.
- 20- Kahn JR. Anderson KE. "Intergenerational Pattern of Teenage Fertility" *Demography* 1992; 29: 39-57.
- 21- World Health Organization and UNICEF, Revised 1990. *Estimates of Maternal Mortality Geneva: WHO and UNICEF* 1996.