

## مقایسه کارایی مدل رگرسیون پواسن تعمیم یافته و چند سطحی با پواسن معمولی در تحلیل رفتار باروری زنان روستایی استان فارس

نجف زارع (PhD)<sup>۱</sup>، مهرباب صیادی (MS)<sup>۲\*</sup>، الهام رضائیان فرد (MS)<sup>۳</sup>، هاله قائم (MS)<sup>۴</sup>، مهرداد وثوقی (MS)<sup>۴</sup>

۱- گروه آمار زیستی دانشکده پزشکی دانشگاه علوم پزشکی شیراز

۲- دانشگاه علوم پزشکی شیراز

۳- دانشکده بهداشت و تغذیه دانشگاه علوم پزشکی شیراز

۴- دانشجو دکتری آمار زیستی دانشگاه علوم پزشکی شیراز

دریافت: ۸۸/۵/۱۰، اصلاح: ۸۸/۷/۸، پذیرش: ۸۸/۱۲/۱۹

### خلاصه

**سابقه و هدف:** در تحلیل داده های باروری از روشهای مختلف آماری می توان استفاده نمود. در حالی که متغیر پاسخ (وابسته)، گسسته باشد مدل پواسن مورد استفاده قرار می گیرد. در صورتی که شرایط مدل پواسن برقرار نباشد، از تعمیم یافته آن و یا در صورتی که متغیری به عنوان متغیر لایه (سطح) وجود داشته باشد بهتر است از پواسن چند سطحی استفاده گردد. هدف از این مطالعه، مقایسه برازندگی دو مدل تعمیم یافته و چند سطحی، نسبت به مدل پواسن معمولی در برآورد ضرایب عوامل موثر بر تعداد فرزندان است.

**مواد و روشها:** این مطالعه مقطعی بر روی یک نمونه ۱۰۱۹ نفری از زنان روستایی ۱۵-۴۹ سال که با استفاده از روش نمونه گیری طبقه بندی انتخاب شدند، انجام گردید. زنان در ۷ طبقه قرار گرفتند و در هر طبقه بطور مساوی با استفاده از روش نمونه گیری سیستماتیک تعداد نمونه های مورد نظر انتخاب شده سپس تحلیل داده ها با استفاده از رگرسیون پواسن معمولی، تعمیم یافته و چند سطحی انجام گرفت.

**یافته ها:** میانگین فرزندان هر زن ۴/۳ و واریانس آن ۸/۳ بود. تحصیلات زن و همسر، سن ازدواج زن، وضعیت اقتصادی و متوسط شیردهی و فاصله گذاری بین فرزندان در دو مدل تعمیم یافته و چند سطحی معنی دار شدند ( $P < 0.05$ ).

**نتیجه گیری:** نتایج مطالعه نشان داد که دو مدل تعمیم یافته و چند سطحی نسبت به معمولی جهت تحلیل این داده ها مناسب تر بوده و می توانند ضرایب عوامل موثر بر تعداد فرزند را دقیقتر برآورد نمایند.

**واژه های کلیدی:** رفتار باروری، مدل پواسن معمولی، مدل پواسن تعمیم یافته، مدل پواسن چندسطحی، تعداد فرزندان.

### مقدمه

مدلهای شمارشی تعمیم یافته را برای توزیع های پواسن، دو جمله ای و دو جمله ای منفی توسعه دادند و نشان دادند که مفروضات مدل پواسن معمولی برای بعضی از داده های شمارشی محدودیت ایجاد می کند (عوضی، ۵). شرط اصلی استفاده از مدل پواسن، معادل بودن میانگین و واریانس متغیر پاسخ می باشد اگر این شرط برقرار نباشد مدل پواسن تعمیم یافته مناسب خواهد بود و اگر تعداد پاسخ صفر

مدل بندی آماری یکی از روشهای تبیین تغییرات در داده های مشاهده شده از طریق معادلات ریاضی می باشد که با استفاده از آن می توان به چگونگی تغییرات داده ها پی برد. معمولی ترین مدل برای تحلیل داده های شمارشی توزیع پواسن است، در سالهای اخیر انواع مدل پواسن برای پاسخهای شمارشی بطور فراوان مورد استفاده قرار گرفته است (۱-۴). Winkelmann و King

این مقاله حاصل پایان نامه مهرباب صیادی دانشجوی رشته کارشناسی ارشد آمار زیستی دانشگاه علوم پزشکی شیراز می باشد.  
\* مسئول مقاله:

آدرس: شیراز، خیابان زند، ساختمان مرکزی دانشگاه علوم پزشکی شیراز، تلفن: ۰۷۱۱-۲۱۲۳۴۱۷

e-mail:sayadi\_me@yahoo.com

## مواد و روشها

این مطالعه مقطعی بر روی ۱۰۱۹ نفر از زنان متاهل ۴۹-۱۵ ساله مناطق روستایی استان فارس که در سن باروری قرار داشتند و با استفاده از نمونه گیری طبقه بندی انتخاب شدند، انجام گردید. 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 گردد. داده ها با استفاده از پرسشنامه خود ساخته ای که اطلاعات بصورت مصاحبه توسط کارشناسان آموزش دیده بهداشت خانواده در آن ثبت می گردید، جمع آوری شد.

پرسشنامه شامل اطلاعات دمو گرافیک (سن زن، سن شوهر، قد زن، تحصیلات زن، شغل زن، تحصیلات شوهر، سن ازدواج زن، وضعیت اجتماعی و اقتصادی) و تعداد فرزند ها بود. جهت تجزیه و تحلیل این داده ها از روشهای آماری رگرسیون پواسن استفاده گردید و  $p < 0.05$  معنی داری در نظر گرفته شد. ساختار توزیع پواسن معمولی به شکل زیر است:

اگر  $Y_i$  متغیرهای تصادفی مستقل باشند که سه فرض زیر در مورد آن برقرار باشد. در هر لحظه زمانی فقط یک پیشامد داشته باشیم، احتمال اتفاق هر پیشامد در هر دوره مستقل از هم باشند و در لحظه شروع هر دوره پیشامدی نداشته باشیم آنگاه  $Y_i$  دارای توزیع پواسن است.

میانگین و واریانس توزیع پواسن با هم برابرند و هر دوی آنها  $\mu_i$  است یعنی  $E(Y_i) = \mu_i$  و  $Var(Y_i) = \mu_i$  که  $\mu_i = \exp(x_i/\beta)$  که  $\beta$  بردار  $K \times 1$  بعدی پارامترها است. در صورتیکه تعداد پاسخ ها تحت تاثیر زمان باشند آنگاه  $\mu_i = t_i \exp(x_i/\beta)$  که  $t_i$  متغیر زمان و مدت زمان مواجهه فرد است و به عنوان شخص سال در تحلیل پواسنی ظاهر می شود. در تحلیل داده های باروری تعداد فرزندان به شدت تحت تاثیر طول دوره زناشویی فرد می باشد.

طبیعی است هر چه طول دوره زناشویی (طول دوره ازدواج) بیشتر باشد تعداد فرزندان بیشتر است بنابراین این متغیر باید به عنوان متغیر زمان (offset) در مدلها وارد می شود. اگر فرضیات توزیع پواسن برقرار نباشد واریانس این توزیع با میانگین آن برابر نیست. وقتی که واریانس بزرگتر از میانگین باشد، بیش پراکندگی پدیدار می شود که منجر به کم برآوردی خطای معیار شده و مقدار آماره آزمون (ملاک آزمون) هر کدام از ضرایب متغیر ها را افزایش می دهد و باعث خطای پذیرش آن متغیر در مدل می گردد. وقتی واریانس کوچکتر از میانگین باشد پراکندگی زیر میانگین ظاهر شده که باعث بیش برآوردی خطای معیار و در نتیجه کاهش مقدار آماره آزمون (ملاک آزمون) هر کدام از ضرایب متغیر ها شده و باعث خطای نپذیرفتن آن متغیر می شود. اگر  $Y_i$  (تعداد فرزندان) دارای توزیع پواسن تعمیم یافته با پراکندگی باشد. در این صورت:  $E(Y_i) = \mu_i$ ،  $Var(Y_i) = \mu_i(1 + \alpha\mu_i)^2$  که  $\mu_i = t_i \exp(x_i/\beta)$  و  $\alpha$  پارامتر پراکندگی می باشد. برای برآورد ضرایب پارامترهای مدلهای پواسنی شبیه رگرسیون لجستیک از روش درست نمایی ماکزیم استفاده شده و برای خطی سازی مدل از تابع اتصال لگاریتمی بهره می جویند (۱۷ و ۱۲ و ۱۶). نیکوئی برازش مدلها در توزیع با استفاده از آماره لگاریتم درستنمایی آزمون می گردد.

(تورم صفر) زیاد باشد از مدلهای پواسنی با تورم صفر استفاده می شود (۱۰-۷). در صورتیکه یک متغیر در داده ها به عنوان متغیر لایه (سطح) وجود داشته باشد باید از مدل پواسنی چند سطحی استفاده کرد (۳ و ۸). داده های شمارشی معمولاً دارای توزیع نرمال نمی باشند 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 مناسب تر است (۱۱). یکی از فرضیات مدل پواسن معمولی این است که احتمال اتفاق هر پیشامد در هر دوره مستقل از هم باشد اما در بعضی دادههای شمارشی وقوع اولیه یک پیشامد ممکن است باعث افزایش یا کاهش احتمال پیشامد در آینده شود. در دادههای باروری تعداد فرزنداها 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 نمی دانند و تصمیم به داشتن فرزند بعدی می گیرند. اگر فرض بالا برقرار نباشد استفاده از مدل پواسنی منجر به برآورد نا صحیحی از ضرایب رگرسیونی می گردد. توزیع های پواسنی تعمیم یافته در حالت های عملی و کاربردی بسیار مطرح می باشد و بیشتر در خصوص داده هایی که پراکندگی دارند، مورد استفاده قرار می گیرد (۱۳ و ۱۲). در بعضی از داده ها ممکن است یک متغیر ایجاد لایه نماید یعنی تعداد افرادی که در هر کدام از سطوح این متغیر قرار می گیرند رفتاری مشابه ولی متغیر با دیگر سطوح داشته باشند در رفتار باروری ممکن است این رفتار از روستایی به روستایی دیگر با توجه به قومیت، ملیت، مذهب، فرهنگ و جمعیت روستا متفاوت باشد بنابراین متغیر روستا ممکن است در تحلیل این داده ها ایجاد لایه نماید که باید آن را در نظر گرفت و از مدلهای چند سطحی استفاده کرد. در این مدل برای هر روستا یک مدل نوشته می شود و سپس یک مدل رگرسیون ترکیبی حاصل می شود (۱۵ و ۱۴).

در داده های باروری تعداد فرزند ها به عنوان متغیر پاسخ، پراکندگی زیادی دارد. یک راهبرد عملی جهت تجزیه و تحلیل چنین داده هایی حذف مقادیر خیلی پرت با استفاده از تصحیح کننده های آماری است که این راهبرد جهت تبیین متغیر پاسخ با استفاده از متغیر های مستقل، نتایج گمراه کننده ای را ارائه می دهد. ولی استفاده از مدلهای پواسن تعمیم یافته و چند سطحی در این خصوص مناسبتر است (۱۶ و ۱۳). توجه به ساختار داده ها و یافتن توزیع مناسب برای متغیر پاسخ از اصول مهم مدل بندی آماری است. بعد از در نظر گرفتن توزیع مناسب متغیر پاسخ، برای تبیین این متغیر بر اساس متغیر های مستقل از رگرسیون آن استفاده خواهد شد.

از آنجا که بعضی از داده های سیستم بهداشتی و درمانی ماهیت شمارشی دارند مدل های پواسنی جهت تجزیه و تحلیل چنین داده هایی مفید می باشد. تاکنون در ایران تجزیه و تحلیل چنین داده هایی بدون توجه به ساختار اصلی داده ها بوده است، بخصوص داده های باروری، که به لحاظ مشکلات افزایش جمعیت و اثرات نامطلوبی که این افزایش بر جامعه می گذارد از اهمیت بیشتری برخوردار است. در تحلیل آماری این کار لازم است که متغیر های موثر بر باروری با دقت بیشتری تبیین و تفسیر شود. لذا هدف اصلی این مطالعه نشان دادن پتانسیل مدلهای پواسنی تعمیم یافته و چند سطحی در تحلیل دقیق داده های باروری و مشابه آن است. در داده های حاضر تعداد فرزند به عنوان متغیر پاسخ و برخی از عوامل دموگرافیک به عنوان متغیر های مستقل در نظر گرفته شده است.

**یافته ها**

در این مطالعه داده های ۱۰۱۹ زن روستایی مورد بررسی قرار گرفت. میانگین فرزندان ۴/۳ و واریانس آن ۸/۳ بود. همچنین میانگین تعداد فرزندان ۲/۱۸ مد توزیع فراوانی تعداد فرزند ۲ بدست آمد (جدول ۱). حدود ۴۱٪ زنان بیش از ۴ فرزند و درصد قابل توجهی از زنان دارای ۱۰ فرزند یا بیشتر هستند (۵۱٪). همچنین میانگین و انحراف معیار متغیر طول دوره زناشویی نیز آورده شده است. میانگین تعداد فرزندان هر زن در بعضی از روستاها متفاوت است (جدول ۲). عوامل موثر بر تعداد فرزندان در سه مدل پراسن مورد مقایسه قرار گرفت (جدول ۳). در این جدول برآورد ضرایب هر کدام از متغیرهای مستقل همراه با خطای استاندارد آن و مقدار احتمال (p-value) آورده شده است. در توزیع پواسنی تعمیم یافته مقدار  $\alpha$  نشاندهنده میزان پراکندگی و در تحلیل چند سطحی پارامتر اثر تصادفی نیز گزارش گردید. برای ارزیابی مناسب بودن مدل، مقدار لگاریتم درستنمایی برای هر سه مدل بیان شده است. مقدار کمتر این آماره از نظر قدر مطلق دلالت بر بهتر بودن مدل برازشی می کند. بر اساس جدول ۳ هر دو مدل پواسنی تعمیم یافته و پواسنی چند سطحی مناسبتر از مدل پواسنی معمولی است که بر اساس نتایج دو مدل متغیرهای تحصیلات زن و همسر، سن ازدواج زن، وضعیت اقتصادی و متوسط شیر دهی و فاصله گذاری بین فرزندانها معنی دار شدند ( $p < 0.05$ ). شاخص نسبت فرزند پسر به کل فرزندان در هیچ کدام از مدلها معنی دار نگردید.

**جدول ۱. توزیع فراوانی تعداد فرزندان، میانگین و انحراف معیار طول دوره زناشویی**

تعداد فرزند	تعداد (%)	تجمعی (%)	طول دوره زناشویی
			(سال)
۰	۱۳ (۱/۳)	۱/۳	Mean±SD ۱۲/۱±۸/۸
۱	۱۹۳ (۱۸/۹)	۲۰/۲	۴/۹±۴/۱
۲	۱۴۸ (۱۴/۵)	۳۴/۷	۸/۶±۴/۲
۳	۱۲۴ (۱۲/۲)	۴۶/۹	۱۲/۵±۴/۷
۴	۱۲۲ (۱۲)	۵۸/۹	۱۵/۵±۴/۵
۵	۸۱ (۷/۹)	۶۶/۸	۱۸/۷±۴/۹
۶	۸۷ (۸/۵)	۷۵/۴	۲۲±۴/۳
۷	۸۹ (۸/۷)	۸۴/۱	۲۴/۴±۴
۸	۶۷ (۶/۶)	۹۰/۷	۲۶/۶±۴/۷
۹	۴۴ (۴/۳)	۹۵	۲۸/۴±۳/۷
۱۰ و +	۵۱ (۵)	۱۰۰	۲۸/۳±۴/۱

**جدول ۲. توزیع فراوانی افراد مورد بررسی در هر روستا، میانگین و انحراف معیار تعداد فرزندان**

شماره روستا	تعداد افراد	تعداد (%)	شماره روستا	تعداد فرزندان	تعداد (%)
۱	۸۷ (۸/۵)	۲±۱/۷	۹	۱۱۰ (۱۰/۸)	۴/۵±۱/۹
۲	۳۸ (۳/۷)	۱/۷±۰/۸	۱۰	۳۶ (۳/۵)	۴/۴±۱/۶
۳	۶۲ (۶/۱)	۱/۷±۰/۹	۱۱	۶۸ (۶/۷)	۳/۸±۱/۵
۴	۳۳ (۳/۲)	۲/۲±۱	۱۲	۵۷ (۵/۶)	۶±۲/۷
۵	۸۱ (۷/۹)	۱/۶±۰/۹	۱۳	۸۲ (۸)	۷/۳±۲/۵
۶	۳۴ (۳/۳)	۲/۳±۱/۲	۱۴	۱۰۸ (۱۰/۶)	۷/۲±۲/۵
۷	۳۹ (۳/۸)	۲/۴±۱/۲	۱۵	۹۸ (۹/۶)	۶/۹±۲/۴
۸	۸۶ (۸/۴)	۴±۲/۳			

**جدول ۳. تعیین عوامل موثر بر تعداد فرزندان با استفاده از روشهای رگرسیون پواسنی معمولی، چند سطحی و تعمیم یافته**

متغیرها	پواسن معمولی			پواسن چند سطحی			پواسن تعمیم یافته		
	برآورد	خطای استاندارد	مقدار احتمال	برآورد	خطای استاندارد	مقدار احتمال	برآورد	خطای استاندارد	مقدار احتمال
تحصیلات زن	-۰/۰۳۴	۰/۰۰۵	≤ ۰/۰۰۱	-۰/۰۲۴	۰/۰۰۵	≤ ۰/۰۰۱	-۰/۰۳۳	۰/۰۰۹	≤ ۰/۰۰۱
شغل زن	-۰/۱۹۷	۰/۰۴۹	≤ ۰/۰۰۱	۰/۰۷۳	۰/۰۵۰	≤ ۰/۰۰۱	۰/۳۰۹	۰/۰۹۷	≤ ۰/۰۰۱
تحصیلات همسر	-۰/۰۸۷	۰/۰۰۴	≤ ۰/۰۰۱	-۰/۰۲۵	۰/۰۰۴	≤ ۰/۰۰۱	-۰/۱۱۵	۰/۰۰۷	≤ ۰/۰۰۱
سن ازدواج زن	-۰/۰۶۴	۰/۰۰۵	≤ ۰/۰۰۱	-۰/۰۶۴	۰/۰۰۵	≤ ۰/۰۰۱	-۰/۰۰۷	۰/۰۰۹	≤ ۰/۰۰۱
وضعیت اقتصادی	۰/۳۷۰	۰/۰۲۵	≤ ۰/۰۰۱	۰/۱۶۰	۰/۰۲۶	≤ ۰/۰۰۱	۰/۵۱۷	۰/۰۴۸	≤ ۰/۰۰۱
نسبت پسر به کل	-۰/۱۰۸	۰/۰۷۳	۰/۱۴	-۰/۰۰۵	۰/۰۷۰	۰/۳۵۲	-۰/۰۹۲	۰/۱۱۰	۰/۴۰
متوسط زمان شیردهی	-۰/۰۲۰	۰/۰۰۳	≤ ۰/۰۰۱	-۰/۰۱۶	۰/۰۰۳	≤ ۰/۰۰۱	-۰/۰۲۶	۰/۰۰۵	≤ ۰/۰۰۱
متوسط فاصله گذاری	-۰/۰۰۹	۰/۰۰۱	۰/۵۷	-۰/۰۱۰	۰/۰۰۱	≤ ۰/۰۰۱	-۰/۰۰۸	۰/۰۰۲	≤ ۰/۰۰۱
مقدار ثابت	۳/۳۴	۰/۱۳۷	≤ ۰/۰۰۱	۲/۷۰	۰/۲۵	≤ ۰/۰۰۱	۲/۴۱	۰/۲۶۷	≤ ۰/۰۰۱
طول دوره زناشویی	offset	offset	offset	offset	offset	offset	offset	offset	offset
$\text{Ln}(\alpha)$									
پارامترهای اثر تصادفی									
شاخص نیکوتی برازش									

Log-likelihood = -۱۹۴۶/۹۳

Log-likelihood = -۲۱۲۵/۴

Log-likelihood = ۲۸۰۰/۶

## بحث و نتیجه گیری

نتایج مطالعه نشان داد که میانگین و واریانس متغیر پاسخ (تعداد فرزند ها) معادل نمی باشد و واریانس بیشتر از میانگین است و میانگین بیشتر از نما است بنابراین داده ها چوله به راست هستند. همچنین معنی دار بودن مقدار  $\alpha$  در مدل پواسنی تعمیم یافته و اثرهای تصادفی در مدل چند سطحی نشاندهنده این است که مدل پواسنی معمولی بر داده ها به درستی برازش نمی گردد و مقدار کمتر آماره آزمون لگاریتم درستنمایی از نظر قدر مطلق در هر دو مدل دلالت بر این امر می کند. بنابراین بکار بردن توزیع پواسنی تعمیم یافته و چند سطحی جهت این داده ها مناسبتر است. به جزء زنانی که به دلیل نازائی فرزند نداشته اند بقیه با افزایش طول دوره زناشویی بر تعداد فرزندانشان افزوده شده است. و این طبیعی میباشد که مدت زمان در معرض باروری بر تعدد فرزندان موثر است و زنان مسن تر فرصت کافی برای داشتن فرزندان بیشتر دارند بنا براین در تحلیل های هر سه مدل این متغیر لحاظ شده است. ولی از وارد کردن سن زن به دلیل همبستگی زیاد آن با طول دوره زناشویی در این مدلها صرف نظر گردید.

در این مطالعه مقدار  $\alpha$  مثبت بود که نشانگر بیش پراکندگی در داده هاست. مطالعات دیگر در شرایطی مشابه، رگرسیون پواسنی معمولی را ناکارآمد و مدل های دیگر از جمله پواسنی تعمیم یافته را پیشنهاد کرده اند (۱۹ و ۲۰). در ایران، در تحلیل مشابه بر روی عوامل موثر بر حاملگی توسط Karimi و همکاران نیز مدل های پواسنی معمولی را مناسب داده های با چنین ساختاری نیافتند و مدل دو جمله ای منفی را پیشنهاد نمودند (۱۸). در برآزش دو مدل تعمیم یافته و چند سطحی بر داده ها متغیرهای تحصیلات زن و همسر، سن ازدواج زن، وضعیت اقتصادی و متوسط شیر دهی و فاصله گذاری بین فرزند ها معنی دار شد. در نتیجه این عوامل بر تعداد فرزندان موثر هستند. یعنی به ازای

یک سال افزایش تحصیل زن، تعداد فرزندان حدود ۲/۴٪ و بر اساس مدل تعمیم یافته ۲/۳٪ کاهش می یابد و یا بر اساس مدل تعمیم یافته به ازای یک سال افزایش تحصیل همسر، تعداد فرزندان حدود ۰/۷ درصد کاهش می یابد. شغل زن فقط در مدل پواسنی تعمیم یافته معنی دار گردید. خطای استاندارد ضرایب در سه مدل پواسن و پواسن تعمیم یافته و چند سطحی اختلافاتی با هم دارند.

در مطالعه ای مشابه بر داده های باروری توسط Wang و همکاران مدل های پواسنی با پواسنی تعمیم یافته مقایسه شد که داده ها ساختار پراکندگی زیر میانگین داشتند و در هر دو مدل تحصیلات زوجین، شاغل بودن زن، وضعیت اقتصادی و محل زندگی بر تعداد فرزند ها موثر بودند ولی ضرایب دو مدل با هم متفاوت بود (۲). در مطالعات انجام شده مدل های متفاوتی بر داده های مشابه برآزش شده است ولی پیشنهاد شد در صورتیکه داده ها پراکندگی دارند از مدل های تعمیم یافته استفاده گردد (۱۹ و ۲۰). در این مطالعه مشخص شد که مدل رگرسیون پواسنی معمولی بر داده های باروری بطور مناسبی برازش نمی گردد و باید از مدل های مناسب تری استفاده شود. لذا توجه به ساختار متغیر پاسخ در اولویت مدل بندی آماری قرار دارد. پیشنهاد می گردد در تجزیه و تحلیل و تفسیر داده های مشابه چنانچه داده ها شمارشی باشند و شرایط مدل پواسنی برقرار نباشد از تعمیم یافته آن و یا دو جمله ای منفی استفاده گردد و در صورتیکه متغیری ایجاد لایه می نماید از مدل چند سطحی نیز استفاده گردد.

## تقدیر و تشکر

بدینوسیله از معاونت محترم تحقیقات و فناوری دانشگاه علوم پزشکی شیراز به دلیل حمایت مالی از این تحقیق قدردانی می گردد.

## Comparison of Generalized and Multilevel Poisson Regression Model with Poisson Model in Fertility Data in Rural of Fars Province (Iran)

N. Zare (PhD)<sup>1</sup>, M. Sayadi (MS)<sup>2\*</sup>, E. Rezaeyan Fard (MS)<sup>2</sup>, H. Ghaem (MS)<sup>3</sup>, M. Vosugh (MS)<sup>4</sup>

1. Department of Biostatistics, Medical school, Shiraz University of Medical Sciences, Shiraz, Iran

2. Shiraz University of Medical Sciences, Shiraz, Iran

3. Health and Nutrition School, Shiraz University of Medical Sciences, Shiraz, Iran

4. PhD Student in Biostatistics, Medical school, Shiraz University of Medical Sciences, Shiraz, Iran

---

Received: Aug 1<sup>st</sup> 2009, Revised: Sep 30<sup>th</sup> 2009, Accepted: Mar 10<sup>th</sup> 2010.

### ABSTRACT

**BACKGROUND AND OBJECTIVE:** Different statistical methods can be used to analyze fertility data. In cases that dependent variable is count, Poisson model is applied. If Poisson model is not applicable in a specific situation, it is better to apply generalized Poisson model and in cases that multilevel variable exists, it is better to use multilevel Poisson model. Our goal in this study is to compare both generalized and multilevel Poisson regression model with Poisson regression model in estimating of coefficient of the effective factors on the number of children.

**METHODS:** This is a cross-sectional study. A sample of 1019 women (15-49 years old) from rural area was selected by stratified sampling. The women were categorized into seven groups and in each group the intended samples were selected equally by systematic sampling. Data were analyzed by Poisson regression model, generalized and multilevel Poisson regression model.

**FINDINGS:** The sample mean and sample variance of the number of children were 4.3 and 8.3, respectively. There was a significant relationship between educational status of the spouses, age of marriage, feeding period, economical status and the interval between the children in generalized and multilevel Poisson regression model.

**CONCLUSION:** According to the results of this study, generalized and multilevel Poisson regression models were more suitable for data analysis and it can estimate coefficient effective of factors on the number of children exactly.

**KEY WORDS:** *Fertility behavior, Poisson model, Generalized poisson model, Multilevel poisson model, Number of children.*

---

\*Corresponding Author;

Address: Central Building, Shiraz University of Medical Sciences, Zand St., Shiraz, Iran

Tel: +98 711 2122417

E-mail: sayadi\_me@yahoo.com

## References

1. Famoye F, Wulu JT, Singh KP. On the generalized poisson regression model with an application to accident data. *J Data Sci* 2004; 2:287-95.
2. Wang W, Famoye F. Modeling household fertility decisions with generalized poisson regression. *J Popul Econ* 1997;10(3):273-83.
3. Skrondal A, Rabe Hesketh S. Some applications of generalized linear latent and mixed models in epidemiology: repeated measures, measurement error and multilevel modeling. *Nor Epidemiol* 2003;13(2):265-78.
4. Wang K, Yau KK, Lee AH. A zero-inflated poisson mixed model to analyze diagnosis related groups with majority of same-day hospital stays. *Comput Methods Programs Biomed* 2002;68(3):195-203.
5. King G. Variance specification in event count models: From restrictive assumptions to generalized estimator. *Am J Pol Sci* 1989;33(3):762-94.
6. Winkelmann R, Zimmerman F. Count data models for demographic data. *Math Popul Stud* 1994;4(3):205-21.
7. Ng SK, Yau KKW, Lee AH. Modeling inpatient length of stay by hierarchical mixture regression via the EM algorithm. *Math Comput Model* 2003;37(3-4):365-75.
8. Karlis D, Xekalaki E. Mixed Poisson distributions. *Int Stat Re* 2005;73(1):35-58.
9. Marazzi A, Paccaud F, Rueffux C, Beguin C. Fitting the distributions of length of stay by parametric models. *Med Care* 1998;36(6):915-27.
10. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; 34: 1-14.
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. New York, Marcel Dekker 1989; pp: 161-8.
13. Consul PC, Famoye F. Generalized Poisson regression model. *Commun Stat Theory Methods* 1992;2(1):89-109.
14. Sayadi M, Zare N. Application of multilevel model in fertility behavioral analysis. Shiraz University of Medical Sciences 2007. MA thesis [in Persian]
15. Kahn JR, Anderson KE. Intergenerational pattern of teenage fertility. *Demography* 1992;29(1):39-57.
16. Famoye F. Restricted generalized Poisson regression model. *Commun Stat Theory Methods* 1993;22:1335-54.
17. Xiao J, Lee AH, Vemurri SR. Mixture distribution analysis of length of hospital stay for efficient funding. *J Socio Econ Plann Sci* 1999;33(1):39-59.
18. Karimi Sh, Kazemnejad A. Application of negative binomial regression model in determining the effective factors of unwanted pregnancy. Tarbiat Modares University 2002. MA thesis. [in Persian]

This document was created with Win2PDF available at <http://www.daneprairie.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.