توزیع سری لگاریتمی آمیخته و کاربرد آن در مدل بندی اثر تغذیه بر بروز عارضه در بیماران دیابتی

پرویز نصیری[1*]، محبوبه میقانی[1]، امیر انشکانی[1]

1) گروه آمار، دانشکده علوم، دانشگاه پیام نور تهران، تهران، ایران

تاریخ دریافت: ۹۳/٤/۷ تاریخ پذیرش: ۹۳/۸/۵

چکیده

مقدمه: مدل بندی یکی از اساسی ترین روش های تبیین متغیرهای آماری است که با استفاده از آن می توان به چگونگی توزیع متغیر پاسخ مورد نظر پی برد. توزیع آمیخته تا حدود زیادی می تواند نیکوئی برازش مدل را بهبود بخشد؛ لذا هدف از این مطالعه معرفی توزیع سری لگاریتمی آمیخته و مدل بندی اثر تغذیه بر بروز عارضه بیماران دیابتی توسط آن می باشد.

مواد و روش ها: در این مطالعه پس از معرفی توزیع آمیخته سری لگاریتمی، مدل مناسب برازش داده شده به داده های مربوط به تغذیه بیماران دیابتی بررسی می گردد. در این مطالعه ۳۲ متغیر مربوط به انواع و نحوه مصرف مواد غذایی به عنوان متغیرهای مستقل و عارضه به عنوان متغیر پاسخ برای مدل بندی آماری در نظر گرفته شدند.

یافته های پژوهش: نتایج مطالعه نشان داد با توجه به مقدار آماره کای دو حاصل از توزیع سری لگاریتمی آمیخته، مدل برازش داده شده با نسبت های آمیختگی $\frac{1}{32}$ به تمام متغیرهای مستقل، مدل مناسبی می باشد.

بحث و نتیجه گیری: با توجه به این که داده های مربوط به بروز عارضه بیماران دیابتی فاقد چگالی صفر می باشند و هم چنین مقایسه مقدار آماره χ^2 حاصل از توزیع سری لگاریتمی آمیخته و توزیع پواسون آمیخته، در می یابیم که توزیع سری لگاریتمی آمیخته مدل مناسب تری به داده ها برازش می دهد.

واژه های کلیدی: شناسایی پذیری، تابع درست نمایی، توزیع سری لگاریتمی آمیخته، دیابت، تغذیه، عارضه

* **نویسنده مسئول:** گروه آمار، دانشکده علوم، دانشگاه پیام نور تهران، تهران، ایران

Email: pnasiri45@yahoo.com

[Downloaded from sjimu.medilam.ac.ir on 2024-03-28]

مقدمه

در بسیاری از مسائل کاربردی تشخیص مدل مناسب برای توزیع صفت جامعه مورد بررسی از اهمیت ویژه‌ای برخوردار است. پدیده‌های طبیعی معمولاً پدیده‌هایی چند متغیره هستند که تحت تاثیر عوامل مختلف بوده و از نوعی ناهمگنی برخوردارند. بررسی این پدیده‌ها به صورت یک جامعه واحد و همگن خصوصاً در حالت چند متغیره می تواند به نتایج کاملاً گمراه کننده ای منجر شود. کاربرد و استفاده از توزیع های آمیخته به دلیل انعطاف پذیری بالای آن ها سابقه بسیار طولانی در اغلب زمینه های علمی از جمله پزشکی، کشاورزی، هواشناسی، بازاریابی، مدیریت و غیره دارد. مزیت استفاده از این گونه توزیع ها زمانی که داده ها دارای پراکندگی زیاد و مجموعه داده ها حاوی مشاهدات گمشده حتی دور افتاده باشند، به خوبی معلوم است. این نوع مسائل آماری اولین بار مورد توجه پیرسن قرار گرفت و پس از او مک لاکلان و باسفورد، مک لاکلان و پیل، مک لاکلان و کریشنان(۴-۲) و سایر محققان آماری در مقالات متعددی روی این موضوع بحث و بررسی کردند. توزیع های آمیخته را می توان بر پایه بسیاری از توزیع های شناخته شده مثل یکنواخت، نمایی، نرمال، پواسن، دو جمله ای و غیره مدل بندی کرد. هدف از هر مدل بندی آماری برآورد پارامترها است، برای برآورد پارامترها روش های گوناگونی از جمله گشتاورها، ماکسیمم درستنمایی، الگوریتم مونت کارلو، EM و MCMC وجود دارد(۸).

دیابت یا «قاتل خاموش» در دنیا به دلیل شیوه نادرست زندگی رو به افزایش است و به عنوان یکی از مهم ترین علت مرگ و میر انسان ها در آینده خواهد بود. افزایش فشارخون، دفع پروتئین در ادرار، احساس گز گز در دست و پا، زخم های طولانی، کم شدن قدرت بینایی، مشکلات کلیوی، عصبی و عفونت های مکرر از عوارض شایع این بیماری است. از آن جا که عوامل مساعدکننده بسیاری در ایجاد و تشدید این عوارض دخیل می باشند، بنا بر این جهت بالا بردن کیفیت زندگی بیماران و تامین رفاه این قشر از جامعه، محققان همواره به دنبال روش های جدید برای مدل بندی و برخورد با فاکتورهای تشدیدکننده بیماری دیابت بر اساس وضعیت تغذیه بیماران می باشند.

با توجه به این که مقادیر متغیر تصادفی توزیع سری لگاریتمی عدد صفر را در دامنه خود ندارد لذا هدف این مقاله، ارائه مدل آماری داده های مربوط به تغذیه اثرگذار بر بروز عارضه در بیماران دیابتی با استفاده از مدل سری لگاریتمی آمیخته می باشد.

توزیع آمیخته: از گذشته های بسیار دور محققان همواره به دنبال روش هایی برای دسته بندی و خلاصه نمودن داده ها بوده اند. این موضوع زمانی اهمیت پیدا نمود که آمار به کمک سایر رشته ها خصوصاً پزشکی، مدیریت، جامعه شناسی، روان پزشکی و مهندسی آمد. با ورود کامپیوتر به تحلیل های آماری و گذر از آمار بیز به آمار کلاسیک به مرور روش های بیشتری برای خلاصه بندی داده ها ابداع شد تا این که کارل پیرسن(۸) توزیع های آمیخته را معرفی نمود.

کاربرد مدل های آمیخته در مواردی است که جامعه آماری ترکیبی از چند زیر جامعه بوده، به گونه ای که وقتی نمونه ای از این جامعه گرفته می شود، به طور دقیق مشخص نیست که هر مشاهده به کدام زیر جامعه تعلق دارد.

تعریف: فرض کنید y_n و ... و y_2 و y_1 نشان دهنده یک نمونه تصادفی به حجم n باشد که y_j یک بردار تصادفی p بعدی با تابع چگالی احتمال $f(y_j;\Psi)$ در فضای R^{p} است. y_j شامل متغیرهای تصادفی مربوط به p مشخصه اندازه گیری شده بر روی j- امین عنصر نمونه است. آن گاه تابع چگالی y_j به صورت زیر نوشته می شود:

$$f(y_j;\Psi)=\sum_{i=1}^{g}\pi_i f_i(y_j;\theta_i) \qquad (۱)$$

که $f_i(y_j;\theta_i)$ تابع چگالی احتمال با بردار پارامترهای ناشناخته θ_i و π_i است، که:

$$\sum_{i=1}^{g}\pi_i=1 \quad ,0\le\pi_i\le 1$$
$$(i=1,2,\dots,g)$$

Ψ شامل تمام پارامترهای مجهول در مدل آمیخته است که به صورت

$$\Psi=(\pi_1,\pi_2,\dots,\pi_{g-1};\ \theta_1,\theta_2,\dots,\theta_g)^T$$

[Downloaded from sjimu.medilam.ac.ir on 2024-03-28]

نوشته می شود. از آن جایی که مجموع π_i ها یک است، یکی از آن ها از روی بقیه قابل محاسبه است، در نتیجه در تعریف Ψ، π_g کنار گذاشته شده است. در تعریف فوق $\pi_1, \pi_2, \dots, \pi_g$ نسبت های آمیخته $f(y_j;\theta_i)$ چگالی مولفه ها، g تعداد مولفه ها در مدل آمیخته و چگالی(۱) یک چگالی آمیخته g مولفه ای نامیده می شود.

یکی از اولین تحلیل های عمده که شامل استفاده از مدل های آمیخته بود، در بیش از صد سال قبل به وسیله زیست سنج مشهور کارل پیرسن(۱۸۹۴) انجام شد. او یک مدل آمیخته با دو مولفه چگالی نرمال را با میانگین های متفاوت μ_1, μ_2 و واریانس های σ_1^2 , σ_2^2 را با نسبت های π_1, π_2 به داده های ویلدن(۳-۱۸۹۲) برازش داد.

شناسایی پذیری: یکی از جنبه های مهم و گاه دشوار توزیع های آمیخته بررسی شناسایی پذیر بودن آن ها است. شناسایی پذیری یک توزیع آمیخته به معنی وجود مشخصه های یکتا برای آن توزیع است. برآورد پارامترها در شرایطی که توزیع آمیخته شناسایی پذیر نباشد، ممکن است منجر به نتایج کاملاً گمراه کننده ای شود. در ساده ترین حالت شناسایی پذیری توزیع آمیخته f با رابطه زیر بیان می شود.

$$f(y_j;\Psi) = f(y_j;\Psi^*) \Leftrightarrow \Psi = \Psi^* \quad (۲)$$

که در آن Ψو Ψ^* مقادیر فضای پارامتری توزیع آمیخته f هستند.

توزیع سری لگاریتمی آمیخته: توزیع سری لگاریتمی را فیشر، کوربت و ویلیامز(۱) برای بررسی توزیع پروانه ها در مالایان پنینسولا، ابداع کردند. این سری ها در نمونه گیری ربعی از گونه های گیاهی، توزیع گونه های جانوری، رشد جمعیت و کاربردهای اقتصادی مورد استفاده قرار گرفته است. تعداد دیگری از محققان از توزیع سری لگاریتمی برای نشان دادن توزیع تعداد آیتم های محصول خریداری شده توسط خریدار در یک دوره زمانی مشخص، استفاده کردند. آن ها به این نکته اشاره کردند که سری لگاریتمی این مزیت را دارند که تنها به یک پارامتر θ وابسته باشند. توزیع سری لگاریتمی از طریق بسط تابع لگاریتمی$-\ln(1-\theta)$ به عنوان سری توانی در θ به دست می آید. روش دیگر برای به دست آوردن آن، در نظر گرفتن آن به مورد محدودکننده توزیع دو جمله ای منفی بدون خطا(صفر بریده)، با میل کردن k به سمت صفر است. هر کدام را که انتخاب کنید، توزیع سری لگاریتمی توزیع بسیار مفیدی روی اعداد صحیح مثبت است. از آن جا که صفر برای متغیر تصادفی، مقدار ممکنی نیست که از توزیع سری لگاریتمی پیروی کند، این توزیع مدل جایگزینی برای آزمایشاتی است که در آن مقدار متغیر تصادفی نمی تواند صفر باشد.

با استفاده از معادله تابع مولد احتمال $P_0(t)$ را می توان به این صورت نوشت:

$$p_0(s) = E[t^x] = \frac{(1-(1-p)t)^{-k}-1}{1-p^{-k}}$$

وقتی $\theta = (1-\theta)$

$$p_0(s) = \frac{(1-(1-\theta t)^{-k}-1}{1-(1-\theta)^{-k}} \quad (۳)$$

با حد گرفتن از p_0 زمانی که k به سمت صفر میل می کند، فرمی نامشخص به دست می آوریم، بنا بر این قانون L' Hospital را به کار می بریم و رابطه (۴) به دست می آید.

$$\lim_{k\to 0} p_0(s) = \lim_{k\to 0} \frac{(1-\theta t)^{-k}-1}{1-(1-\theta)^{-k}} = \frac{-\ln(1-\theta t)}{-\ln(1-\theta)} = \frac{-1}{\ln(1-\theta)} \sum_{r=1}^{\infty} \frac{(\theta t)^k}{r}$$

اکنون، تابع مولد احتمال، توزیع احتمال متغیر تصادفی را مشخص می کند. یعنی اگر متغیر تصادفی x تابع مولد احتمالی داشته باشد که توسط $\Pi_0(t)$ به دست آمده باشد، آن گاه تابع تجمعی احتمال از این رابطه به دست می آید:

$$P[X=x] = \left[\frac{1}{x!}\frac{d^x}{d^y}\Pi_0(t)\right]_{t=0} = \frac{-1}{\ln(1-\theta)}\frac{\theta^x}{x} \quad x = 1,2,3,\dots$$

که به آن توزیع سری لگاریتمی گفته می شود. به راحتی می توان ثابت کرد که:

[Downloaded from sjimu.medilam.ac.ir on 2024-03-28]

$$E(X)=\frac{\alpha\theta}{1-\theta}\ ,\ V(X)=\frac{\alpha\theta}{(1-\theta)^2}\ ,\ \alpha=\frac{-1}{\ln(1-\theta)}$$

برای برآورد کردن θ فرض می‌کنیم $x_1,x_2,\cdots,x_n$ نمونه تصادفی از توزیع سری لگاریتمی با پارامتر θ است. میانگین نمونه به این شکل به دست می‌آید:

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

در بالا ثابت شد که در جامعه:

$$E(X)=\mu=\frac{\alpha\theta}{1-\theta}$$

با مساوی قرار دادن این دو، معادله زیر به دست می آید:

$$\overline{x}=\frac{\alpha\theta}{1-\theta}$$

که

$$\frac{\alpha}{\overline{x}}=\frac{1-\theta}{\theta}$$

یا

$$\hat{\theta}=\frac{\overline{x}}{\overline{x}+\alpha}$$

ضرورت دارد که مقدار $\hat{\theta}$ متعلق به بازه $(0,1)$ باشد.

در ضمن تابع چگالی توزیع پواسن به صورت زیر می باشد.

$$f(x)=\frac{e^{-\theta}\theta^{x}}{x!}\qquad x>\circ$$

تابع چگالی توزیع آمیخته برای مشاهدات $i=1,2,\ldots,n$ به صورت زیر است:

$$f(y_i|\theta)=\sum_{j=1}^{k}\pi_j f_j(y_i|\theta_j)$$

که در آن $\sum_{j=1}^{k}\pi_j=1$ و $\pi_j>0(j=1,2,\ldots,k)$ و بردار $\Psi=(\pi_1,\pi_2,\ldots.\pi_{k-1},\theta_1,\theta_2,\ldots.,\theta_k)$ تمام پارامترهای مدل می باشد. احتمالات آمیخته π_j می تواند به عنوان احتمالات غیر شرطی در نظر گرفته شود.

حال با جایگذاری توزیع سری لگاریتمی در معادله بالا، توزیع سری لگاریتمی آمیخته متناهی با احتمال غیرشرطی به دست می آید:

$$f(y_i|\theta)=\sum_{j=1}^{k}\pi_j\frac{-1}{In(1-\theta)}\frac{\theta^{y_i}}{y_i}$$

این مدل با استفاده از برنامه نوشته شده در نرم افزار *R* جهت تحلیل داده های مورد نظر به کار گرفته شده است. نیکوئی برازش مدل ها در توزیع با استفاده از آماره کای دو مورد بررسی قرار گرفته و بهبودی مدل در کاهش آماره کای دو بوده است.

داده های مورد بررسی در این مقاله، بیماران دیابتی تحت پوشش انجمن خیریه حمایت از بیماران دیابتی شهر اراک از مهرماه سال ۱۳۸۵ تا اردیبهشت ماه سال ۱۳۸۶ می باشند. حجم نمونه بر اساس یک مطالعه مقدماتی بر روی ۳۰ بیمار دیابتی که به طور تصادفی انتخاب شده بودند و با استفاده از فرمول

$$n=\frac{p(1-p)\mathrm{Z}^2(1-{}^{\alpha}/_{2})}{\mathrm{d}^2}$$
$$=\frac{0.9(1-0.9)(1.96)^2(1-\frac{0.05}{2})}{(0.03)^2}\geq 375$$

حدود ۴۰۰ نفر برآورد شد. نمونه ها به صورت تصادفی انتخاب شده اند؛ که برای آن ها عارضه به عنوان متغیر پاسخ و ۳۲ متغیر ذکر شده در ذیل به عنوان متغیرهای مستقل ذکر شدند. عارضه ناشی از بیماری دیابت که در این تحقیق مدنظر است، عبارت است از فشارخون بالا، دیده شدن پروتئین در ادرار، وجود زخم های دیر خوب شونده، گزگز شدن دست و پاها، کوری، عفونت های مکرر و بیماری قلبی و کلیوی. بیماری که یک یا چند تا از فاکتورهای فوق را دارا باشد، در این تحقیق عارضه دار محسوب می شود(جدول شماره ۱).

[Downloaded from sjimu.medilam.ac.ir on 2024-03-28]

جدول شماره ۱. متغیرهای مستقل مربوط به عارضه دیابت

K	۱	۲	۳	۴	۵	۶
نام متغیر	حبوبات	جنس	سن	نوع دیابت	وزن	مدت ابتلا
K	۷	۸	۹	۱۰	۱۱	۱۲
نام متغیر	مصرف صبحانه	سنگک	لواش	بربری	نان جو	چربی
K	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸
نام متغیر	تعداد وعده غذایی	آب پز	سرخ کردنی	کبابی	مغز	مصرف گوشت
K	۱۹	۲۰	۲۱	۲۲	۲۳	۲۴
نام متغیر	نوع روغن	نوع گوشت	دارچین	سبزی	شنبلیله	نخود فرنگی
K	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰
نام متغیر	سیب زمینی	اسفناج	هویج	لبنیات	میوه	تخم مرغ
K	۳۱	۳۲				
نام متغیر	میزان قند خون	میزان کلسترول خون				

یافته های پژوهش

کاربرد مدل بیان شده در این مقاله بر روی نمونه‌های انتخاب شده از بین بیماران دیابتی صورت گرفته است.در مطالعه انجام شده حاضر بر روی ۴۰۰ بیمار به این نتیجه رسیدیم که ۸۹ درصد بیماران دچار عارضه می باشند. با توجه به این که در مورد متغیر عارضه مقدار صفری دیده نمی شود، توزیع سری لگاریتمی که به مقدار صفر چگالی نسبت نمی دهد، توزیع مناسب تری جهت برازش داده ها خواهد بود.

چگالی مربوط به متغیر عارضه به صورت زیر است:

$$f(x) = \sum_{i=1}^{32} \pi_i f_i(x) \qquad , x = 1,2, \dots$$

که چگالی یک توزیع سری لگاریتمی آمیخته است. در این چگالی π_i (احتمالات آمیخته) وزنی است که به چگالی مربوط به متغیر i ام، $i = 1, \dots, 32$، نسبت داده شده است. لازم به ذکر است که $\sum_{i=1}^{32} \pi_i = 1$. در این مقاله با در نظر گرفتن مقادیر مختلفی برای π_i، $i = 1, \dots, 32$ به مقایسه بین چگالی های ایجاد شده برای متغیر عارضه، تحت توزیع های سری لگاریتمی و پواسون آمیخته، جهت تعیین بهترین مدل می پردازیم. از آن جا که بررسی برازش مدل برای داده‌های فوق مورد نظر می باشد، آماره آزمون برای نیکویی برازش تقریباً دارای توزیع کی دو است بدین منظور از آماره کای دو استفاده می کنیم و بین چگالی های پیشنهادی بر اساس مقادیر مختلف π_i $i = 1, \dots, 32$، چگالی برتری دارد که دارای مقدار آماره کای دو کمتری نسبت به بقیه باشد.

با توجه به نوع متغیرها ۷ مدل با احتمالات آمیخته(π_iها) مختلف بیان شده در ذیل در نظر می گیریم، سپس مقادیر آماره کای دو را برای هر ۶ مدل، تحت سری لگاریتمی آمیخته و پواسون آمیخته محاسبه و با هم مقایسه می کنیم.

مدل ۱: تخصیص وزن $\frac{1}{6}$ به متغیرهای مربوط به عادات غذایی(نحوه مصرف صبحانه، تعداد وعده غذایی، مصرف غذای سرخ کردنی، مصرف غذای کبابی، مصرف گوشت و نوع روغن مصرفی)، مدل ۲: تخصیص وزن ۱/۳ به متغیرهای مربوط به ویژگی های فردی(جنس، سن، وزن)، مدل ۳: تخصیص وزن ۱/۶ به متغیرهای مربوط به نان و غلات(حبوبات، سنگک، لواش، بربری، نان جو و مغز)، مدل ۴: تخصیص وزن ۱/۴ به متغیرهای مربوط به وضعیت سلامتی افراد(نوع دیابت، مدت ابتلاء، میزان قندخون و میزان کلسترول خون)، مدل ۵: تخصیص وزن ۱/۸ به متغیرهای مربوط به مصرف میوه و سبزیجات(دارچین، سبزی، شنبلیله، نخود فرنگی، سیب زمینی، اسفناج، هویج، میوه)، مدل ۶: تخصیص وزن ۱/۶ به متغیرهای مربوط به مواد پروتئینی(چربی، آب پز، نوع گوشت، میزان مصرف لبنیات و تعداد تخم مرغ مصرفی در هفته)، مدل۷: تخصیص وزن$\frac{1}{32}$ به ۳۲ متغیر مستقل مورد بررسی

[Downloaded from sjimu.medilam.ac.ir on 2024-03-28]

جدول شماره ۲ مقادیر آماره کای دو را برای هر کدام از مدل های فوق تحت سری لگاریتمی آمیخته و پواسون آمیخته نشان می دهد؛ که با استفاده از نرم افزار R محاسبه شده است.

جدول شماره ۲. آماره کای دو برآورد شده تحت سری لگاریتمی آمیخته و پواسون آمیخته

مدل های مورد بررسی	آماره کای دو تحت سری لگاریتمی	آماره کای دو تحت پواسون آمیخته
۱	۵/۱۵۵	۲۰۵/۶۹
۲	۷/۵۲۸	۲۲۸/۱۹
۳	۲/۰۷۲	۱۹۹/۸
۴	۳/۴۳۲	۲۰۰/۲۶
۵	۳/۲۲۲	۱۹۴/۴۶
۶	۱/۷۶۴	۱۸۱/۲۰
۷	۰/۱۳۱	۱۹۷/۸۶

بر اساس نتایج جدول شماره ۲ مقدار آماره کای دو توزیع سری لگاریتمی آمیخته برابر با ۰/۱۳۱ برای مدل هفتم با تفاوت زیادی کمتر از بقیه مدل ها است و این نشان می دهد که در نظر گرفتن چگالی آمیخته با وزن های برابر ۱/۳۲ برای همه متغیرهای مستقل مناسب ترین مدل برازش شده در بین مدل های مورد بررسی است. در مورد توزیع پواسون آمیخته نیز می توان گفت که مدل ششم با مقدار آماره کای دو برابر با ۱۹۷/۸۶ بهترین مدل ممکن می باشد، زیرا مقدار آماره کای دو مربوط به آن از بقیه کمتر است.

بحث و نتیجه گیری

با توجه به جدول شماره ۲ و مقادیر آماره کای دو تحت دو توزیع سری لگاریتمی و پواسون آمیخته می بینیم که توزیع پواسون آمیخته در تمامی موارد مقدار آماره کای دو بیشتری در مقایسه با توزیع سری لگاریتمی آمیخته دارد. پس نتیجه می گیریم توزیع سری لگاریتمی توزیع مناسب تری جهت برازش داده ها می باشد. علت این امر این است که توزیع پواسون به مقدار صفر چگالی مثبتی را نسبت می دهد، در حالی که در مورد متغیرهای مورد بررسی مقدار صفری دیده نمی شود. بنا بر این توزیع سری لگاریتمی که به مقدار صفر چگالی نسبت نمی دهد، توزیع مناسب تری جهت مدل بندی داده ها می باشد.

و بر این اساس مدل هفتم با احتمالات آمیخته برابر $\frac{1}{32}$ برای کلیه متغیرهای تاثیرگذار بر بروز عارضه بیماران دیابتی مورد بررسی، که مقدار آماره کای دو توزیع سری لگاریتمی آمیخته آن با تفاوت زیادی کمتر از بقیه موارد است، مناسب ترین مدل برازش شده می باشد.

References

1.Fisher RA, Cobert AS, Williams CB. The relationbetween the number of species and the number of individuals in a randomsample of an animal population. J Anim Ecol1943;12:42-57.

2.Jeffrey R, Wilson. Logarithmic Series Distribution and Its use in Analyzing Discrete Data. Arizona Sta Uni Tempe1990;22: 852-7.

3. Fong Y, Wakefield J, Rice K. Bayesian mixture modeling using a hybrid sampler with application to protein subfamily identification. Biostatistics2010;11:18-33.

4. Shaikh M, Mcnicholas PD, Desmond AF. A pseudo-EM algorithm for clustering incomplete longitudinal data. Int J Biostat2010;6:8.

5. Vanhavre Z, White N, Rousseau J, Mengersen K. Overfitting bayesian mixture models with an unknown number of components. PLoS One2015; 15;10:131-9..

6.Nasiri P, Aneshkani, A. Estimation Parameters of Multivariate Normal Pattern. Int J Math Stat2013;14: 9-21.

7.Nasiri P, Daraei MR. Finite mixture logarithmic series and its application to

[Downloaded from sjimu.medilam.ac.ir on 2024-03-28]

analysis of defaulters behavior. Int J Acad Res2011; 3:1175- 8.
8.Nasiri P. Estimation Parameter of Zero Truncated Mixed Poisson Models. International Journal of Mathematical Analysis2011;5:465-70.
9.Pearson K. Contribution to the theory of mathematical evolution. Royal Soc London1984;185, 71-110.
10. Beath KJ. A finite mixture method for outlier detection and robustness in meta-analysis. Res Synth Methods 2014;5:285-93.
11. Salter M, Murphy TB. Role Analysis in Networks using Mixtures of Exponential Random Graph Models. J Comput Graph Stat2015;24:520-38.

Mixture of Logarithmic Series Distribution and Its Application in Modeling of Effects of Nutrition on Complications in Diabetic Patients

Nasiri P[1], Mighani M[1], Aneshkani A[1]*

Abstract

Introduction: Modeling is one of the most basic methods to explain the statistical variables that by using it we can realize the response variable distribution. Mixture distribution can improve the goodness of fit largely, so, the aim of this study was to introduce the mixture of logarithmic series distribution and modeling of effects of nutrition on complications in diabetic patients by it.

Material & methods: This study examined appropriate model fitted related to diabetic patient's data after introducing the mixture of logarithmic series distribution. 32 variables have been considered which related to type and consumption instruction of food as independent variables and complication as dependent variable.

Findings: The results showed that according to chi-square value of Mixture logarithmic series distribution, fitted model with a mixing ratio of 1/32 to all independent variables are reasonable model.

Discussion & Conclusion: given that data of diabetic patients' complications have no zero density and also comparing χ^2 statistic from mixture of logarithmic series distribution combined and mixed Poisson distribution, we find that mixed logarithmic series distribution model gives a better fit to data.

Keywords: Ability to identify, Likelihood function, Mixture of logarithmic series distribution, Diabet, Nutrition, complication

1. Dept of Satatistics, faculty of Science, Tehran PayameNoor University, Tehran, Iran
**** Correspondin author Email: pnasiri45@yahoo.com***

[Downloaded from sjimu.medilam.ac.ir on 2024-03-28]