



برآورد بیزی مدل رگرسیون بتا و مدل‌بندی سهم مهاجرت‌های استانی

زهرة فلاح محسن خانی^۱، محسن محمدزاده^۱

^۱دانشگاه تربیت مدرس، mohsen_m@modares.ac.ir zohrehf@yahoo.com

چکیده: مدل رگرسیون بتا که در سال‌های اخیر در منابع مورد استفاده قرار گرفته است یک انتخاب مناسب برای متغیرهای پیوسته روی بازه (۰ و ۱) است. در این مدل متغیر پاسخ دارای توزیع بتا در نظر گرفته می‌شود و پارامترهای میانگین و دقت متغیر پاسخ از طریق تابع‌های پیوند مناسب با الگویی خطی به متغیرهای تبیینی مرتبط می‌شوند. در این مقاله، ضمن معرفی مدل رگرسیون بتا نحوه برآورد بیزی پارامترهای مدل ارائه می‌شود. سپس نحوه کاربست این مدل در مدل‌بندی سهم مهاجرت‌پذیری استان‌های کشور طی فاصله دو سرشماری متوالی نشان داده می‌شود. در مدل‌بندی میانگین متغیر پاسخ، متغیرهای تبیینی سهم شاغلین استان و سرانه عملکرد اعتبارات هزینه‌ای دستگاه‌های اجرایی از محل درآمد عمومی استان به عنوان متغیرهای موثر بر سهم مهاجران وارد شده به استان و در مدل‌بندی دقت متغیر پاسخ، متغیر تبیینی سرانه ارزش افزوده استان در نظر گرفته شد. با توجه به مقادیر برآورد شده پارامترها، استان‌های مهاجرپذیر از سهم شاغلین بالاتری برخوردار هستند. بنابراین فراهم آوردن اشتغال پایدار و ایجاد فرصت‌های شغلی مناسب در استان‌های مهاجر فرست، می‌تواند مهمترین عامل بازدارنده مهاجرت بین استانی باشد.

واژه‌های کلیدی: مهاجرت، نسبت اشتغال، مدل رگرسیون بتا، توزیع بتا

کد موضوع بندی ریاضی (۲۰۱۰): 91D20. محور تخصصی: آمار رسمی (کد ۵۰۹)

۱ مقدمه

مهاجرت واژه‌ای است که علوم مختلف به تناسب نیاز خود به تعریف آن پرداخته‌اند و هر علمی با توجه به پیامدها و اثراتی که مهاجرت بر روی جوامع می‌گذارد در این باره بحث‌هایی ارائه نموده‌اند. واژه مهاجرت در فرهنگ فارسی معین، (معین، ۱۳۶۳) به معنی از موطن خود به جایی دیگر انتقال کردن، آمده است. همچنین در لغت‌نامه دهخدا، (دهخدا، ۱۳۵۲) به معنی ترک دیار گفتن و در مکان دیگر اقامت کردن می‌باشد. در لغت‌نامه جمعیت‌شناسی سازمان ملل، (زنجانی، ۱۳۷۶) واژه مهاجرت این گونه تعریف شده است: «شکلی از تحرک جغرافیایی یا ترک مکانی است که بین دو واحد جغرافیایی صورت می‌گیرد. این تحرک جغرافیایی تغییر اقامتگاه از مبدا یا

^۱ زهرة فلاح محسن خانی : zohrehf@yahoo.com

محل حرکت به مقصد یا محل ورود می‌باشد». مهاجرت از جنبه‌های بسیاری مطالعه شده است، میرزا مصطفی و همکاران (۱۳۹۲)، انگیزه‌های مهاجرت یکی از مواردی است که محققان بسیاری در این خصوص مطالعه نموده‌اند. به عنوان مثال، بوگ، به سه مورد انگیزه مهاجرت، عوامل موثر در انتخاب مقصد و شرایط اجتماعی، اقتصادی موثر در مهاجرت اشاره نموده است و تصمیم به مهاجرت را در چهار گروه، عوامل موثر مربوط به مبدا، عوامل موثر مربوط به مقصد، عوامل و ویژگی‌های فردی مهاجران و موانع عمده مهاجرت یا عوامل مداخله‌گر طبقه‌بندی کرده است.

در این مقاله عوامل موثر بر مقصد مهاجرت برای افرادی که از سرشماری نفوس و مسکن ۱۳۸۵ تا سرشماری ۱۳۹۰ محل سکونت خود را تغییر داده‌اند، با استفاده از مدل آماری، بررسی می‌شود. یکی از متغیرهای قابل بررسی تعداد مهاجران وارد شده به هر استان است. به منظور فراهم کردن امکان مقایسه میزان مهاجرت به استان‌ها با توجه به جمعیت آن‌ها، متغیر تعداد مهاجران وارد شده استان از سرشماری عمومی نفوس و مسکن ۱۳۸۵ تا سرشماری عمومی نفوس و مسکن ۱۳۹۰ تقسیم بر جمعیت استان در سرشماری عمومی نفوس و مسکن ۱۳۹۰ مورد بررسی قرار می‌گیرد. متغیر مذکور سهم مهاجران وارد شده هر استان نامیده می‌شود. از آن جا که دامنه این متغیر پذیرای اعداد بین صفر و یک است، لازم است مدل آماری مناسب برای تحلیل این نوع داده‌ها به کار رود. بنابراین در این مقاله ابتدا روش مناسب برای مدل‌بندی متغیر پاسخ سهم معرفی شده، سپس با استفاده از آن، عوامل موثر بر سهم مهاجران وارد شده هر استان مشخص می‌شوند.

۲ مدل آماری

از آنجا که سهم مهاجران وارد شده هر استان طی سرشماری‌های نفوس و مسکن ۱۳۸۵ و ۱۳۹۰ از تقسیم تعداد مهاجران بر جمعیت استان در سال ۱۳۹۰ به دست می‌آید، مقادیری بین صفر و یک را به خود اختصاص می‌دهد، بنابراین می‌بایست برای مدل‌بندی آن، روش‌های مدل‌بندی داده‌های نسبتی به کار رود. راه حل متداول، تبدیل متغیر مورد بررسی به گونه‌ای است که مقادیر آن در مجموعه اعداد حقیقی قرار گیرند تا با لحاظ فرض نرمال بودن آن متغیر، به‌کارگیری مدل رگرسیونی معمول امکان‌پذیر شود. این رویکرد دارای اشکالاتی است، از جمله این که برآورد پارامترهای مدل به راحتی قابل تفسیر نیستند. از طرفی چون توزیع اندازه‌های نسبت ممکن است متقارن نباشند استنباط بر اساس فرض نرمال بودن گمراه‌کننده است. پائولینو (۲۰۰۱)، مدل رگرسیون بتا را که در آن متغیر پاسخ دارای توزیع بتا است پیشنهاد کرد. برای مدل‌بندی متغیرهای پاسخ از جنس نسبت، فرض می‌شود که متغیر پاسخ از توزیع بتا با تابع چگالی

$$\pi(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1} \quad 0 < y < 1, a, b > 0$$

پیروی می‌کند، که میانگین و واریانس آن به ترتیب عبارتند از: $E(Y) = \frac{a}{a+b}$ و $Var(Y) = \frac{a(1-a)}{(a+b)^2 \times (b+a+1)}$ با توجه به این که در مدل‌های رگرسیونی، الگوی رفتار میانگین متغیر پاسخ مشروط بر متغیرهای تبیینی، مورد بررسی قرار می‌گیرد، فراری و کریباری (۲۰۰۴)، توزیع بتای باز پارامتریده را برای مطالعه نسبت‌ها پیشنهاد کردند. آن‌ها پارامترهای توزیع بتا را به‌گونه‌ای بازنویسی کردند که مدل رگرسیونی براساس میانگین متغیر پاسخ معین شود. برای این منظور با قرار دادن $\mu = \frac{a}{a+b}$ و $\phi = a+b$ توزیع بتا به صورت:

$$\pi(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1} \quad 0 < y < 1$$

بازپارامتریده می‌شود، که در آن μ میانگین و ϕ پارامتر دقت است. آن‌ها با در نظر گرفتن الگوی خطی برای متغیرهای تبیینی به برازش مدل پرداختند. در این مدل پارامترهای میانگین و دقت متغیر پاسخ، از طریق تابع پیوند مناسب به ترکیبی خطی از متغیرهای تبیینی به

صورت:

$$g(\mu_i) = \sum_{i=1}^k x_i \beta_i = \eta_i$$

$$h(\phi_i) = \sum_{i=1}^k x_i c_i = \tau_i$$

مربوط می‌شوند، که در آن β و c پارامترهای رگرسیونی هستند. در این مطالعات معمولاً برای میانگین، تابع پیوند لجوجیت و برای دقت، تابع پیوند لگاریتم با در نظر گرفتن الگوی خطی به متغیرهای تبیینی ربط داده شده‌اند. فیگورا و همکاران (۲۰۱۳)، پارامترها را به صورت $\mu_i = \frac{\exp(X_i \beta)}{1 + \exp(X_i \beta)}$ و $\phi_i = c_i \exp(X_i)$ مدل‌بندی کردند. فرض کنید $y = (y_1, \dots, y_n)$ و $\eta = (\eta_1, \dots, \eta_n)$ و $\tau = (\tau_1, \dots, \tau_n)$ بردار مشاهدات باشند، با فرض استقلال پارامترها، چگالی پسینی توام عبارت است از:

$$f(\beta, \eta, \tau, c; y) \propto \left[\prod_{i=1}^n f(y_i | \tau_i, \eta_i) \right] \left[\prod_{i=1}^n f(\eta_i | \beta) \right] \left[\prod_{i=1}^n f(\tau_i | c) \right] f(\beta) f(c)$$

از آن جا که به دست آوردن توزیع‌های پسینی حاشیه‌ای بسیار پیچیده است از الگوریتم زنجیر مارکوف مونت کارلو (MCMC) و نمونه گیر گیبز استفاده می‌شود. نمونه‌های پسینی از توزیع‌های شرطی کامل پارامترها به دست می‌آید. توزیع‌های شرطی کامل عبارتند از:

$$P(\beta | \eta, \tau, c; y) \propto \prod_{i=1}^n f(y_i | \beta) \pi_0(\beta)$$

$$P(c | \eta, \tau, \beta; y) \propto \prod_{i=1}^n f(y_i | c) \pi_0(c)$$

محاسبات الگوریتم (MCMC) شامل نمونه‌گیری گیبز و الگوریتم متروپولیس هستینگ با استفاده از بسته R2WINBUGS که نرم افزار WINBUGS را به نرم افزار R متصل می‌نماید انجام می‌شود، (اسپیگل هالتر، ۲۰۰۳).

۳ مدل‌بندی و تحلیل داده‌های مهاجرت

در این بخش برای بررسی عوامل موثر بر سهم مهاجران وارد شده هر استان طی سرشماری‌های نفوس و مسکن ۱۳۸۵ و ۱۳۹۰ مدل‌بندی انجام می‌شود. برای متغیر پاسخ سهم مهاجران وارد شده هر استان با توجه به آمارهای رسمی موجود، متغیرهای کمکی سهم افراد دارای تحصیلات عالی استان، سهم شاغلین هر استان، سرانه ارزش افزوده و سرانه عملکرد اعتبارات هزینه‌ای دستگاه‌های اجرایی از محل درآمد عمومی استان در نظر گرفته شده است که مقادیر آن در جدول ۲ ارائه شده است. برای پارامترها، توزیع پیشین نرمال با میانگین صفر و واریانس ۱۰۰ در نظر گرفته می‌شود. علت بزرگ فرض کردن واریانس، ناآگاهی بخش بودن این توزیع به علت عدم وجود اطلاعات مورد نیاز توزیع پیشین می‌باشد. در این مقاله لجوجیت میانگین مهاجران وارد شده هر استان طی سرشماری‌های نفوس و مسکن ۱۳۸۵ و ۱۳۹۰ براساس متغیرهای کمکی به دست می‌آید و لگاریتم دقت مهاجران وارد شده هر استان نیز بر اساس متغیرهای کمکی در نظر گرفته می‌شود. لجوجیت میانگین مهاجران وارد شده هر استان، بر اساس متغیرهای کمکی سهم شاغلین هر استان و سرانه عملکرد اعتبارات هزینه‌ای دستگاه‌های اجرایی از محل درآمد عمومی استان به صورت:

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

مدل‌بندی می‌شود، که در آن β_0 عرض از مبدا و β_1 و β_2 پارامترهای رگرسیونی مدل هستند و اندیس i معرف استان مورد نظر است. همچنین لگاریتم دقت مهاجران نیز بر اساس متغیر کمکی سرانه ارزش افزوده استان به صورت:

$$\ln(\phi_i) = c_0 + c_1 x_i$$

در نظر گرفته شده است. برآورد پارامترها با ۵۰۰۰۰۰ تکرار مونت کارلو و بعد از صرف نظر کردن از ۲۵۰۰۰۰ نمونه اول، به دست می‌آید. نمودارهای خودهمبستگی نشان می‌دهد که زنجیر هر پارامتر مانا و ناهمبسته است که این بررسی برای رسیدن به برآورد مناسب پارامترهای مدل ضروری است. خلاصه‌های پسین که بیانگر ضرایب مدل رگرسیون بتا است، در جدول ۱ ارائه شده است:

جدول ۱: برآورد ضرایب رگرسیونی پارامترهای مدل

پارامتر	میانگین	انحراف استاندارد	۲/۵%	۹۷/۵%
β_0	-۴/۲۶	۰/۴۶	-۵/۱۷۷	-۳/۴۰۹
β_1	۴/۵۶	۱/۶۱	۱/۶۰۶	۷/۹۲۴
β_2	۰/۳۹۹	۱/۱۷	۰/۰۶۱	۰/۷۴۳
c_0	۱/۵۲۸	۰/۰۲۶	۱/۴۷۷	۱/۵۷۸
c_1	۱/۱۲۸	۰/۰۰۵	۰/۱۱۸	۰/۱۳۸

پارامترهای β_0 ، β_1 و β_2 به ترتیب عرض از مبدا، ضریب متغیر تبیینی سهم شاغلین و ضریب متغیر تبیینی سرانه عملکرد اعتبارات هزینه‌ای دستگاه‌های اجرایی از محل درآمد عمومی در لوجیت میانگین و پارامترهای c_0 و c_1 به ترتیب عرض از مبدا و ضریب متغیر تبیینی سرانه ارزش افزوده در لگاریتم دقت است.

از میان متغیرهای تبیینی مورد بررسی در لوجیت میانگین، متغیرهای سهم شاغلین و سرانه عملکرد اعتبارات هزینه‌ای دستگاه‌های در مدل‌بندی پارامتر دقت از میان متغیرهای تبیینی مورد بررسی، متغیر سرانه ارزش افزوده با اثر مثبت وارد مدل می‌شود. بعبارتی با بالا اجرایی از محل درآمد عمومی استان با اثر مثبت وارد مدل رگرسیونی میانگین می‌شوند و عرض از مبدا با اثر منفی وارد مدل می‌شود. رفتن سرانه ارزش افزوده استان‌ها پارامتر دقت افزایش می‌یابد یعنی تغییرات سهم مهاجران وارد شده به استان‌ها کاهش می‌یابد و عرض از مبدا نیز با اثر مثبت در مدل رگرسیونی دقت تاثیرگذار است. از آن جا که متغیر تبیینی سهم شاغلین استان نسبت به دیگر متغیرهای تبیینی از اثرگذاری بیشتری برخوردار است می‌توان گفت مهمترین دلیل مهاجرت در افرادی که از سرشماری نفوس و مسکن ۱۳۸۵ تا سرشماری ۱۳۹۰ محل سکونت خود را تغییر داده‌اند امید به یافتن یا بهتر شدن اشتغال در استان مقصد بوده است. بنابراین برنامه‌ریزی برای ایجاد اشتغال و فرصت‌های شغلی مناسب و پایدار در استان‌های مهاجر فرست مهمترین عامل بازدارنده‌ی جریان‌های مهاجرت بین استانی است.

بحث و نتیجه‌گیری

پدیده مهاجرت به لحاظ تاثیری که بر جامعه مبدا و مقصد می‌گذارد و باعث تغییر ساختار جمعیتی و پارامترهای اجتماعی، اقتصادی و فرهنگی جامعه‌ها می‌شود موضوع مهمی است که در علوم مختلف به آن پرداخته شده است. بدین منظور در این مقاله عوامل موثر بر سهم مهاجران وارد شده هر استان طی سرشماری‌های نفوس و مسکن ۱۳۸۵ و ۱۳۹۰ با استفاده از مدل رگرسیون بتا بررسی شده است و دلیل اصلی مهاجرت پذیری استان‌ها مشخص شده است.

جدول ۲: سهم مهاجران وارد شده مقادیر متغیرهای تبیینی به تفکیک استان، ۱۳۹۰

استان	سهم مهاجران		سهم افراد دارای		سرانه ارزش	سرانه عملکرد
	وارد شده	تحصیلات عالی	سهم شاغلین	افزوده		
آذربایجان شرقی	۰/۰۵۶۲	۰/۱۲۷۰	۰/۳۰۲۷	۰/۰۵۶۰	۰/۴۸۳۴	اعتبارات هزینه ای
آذربایجان غربی	۰/۰۵۷۶	۰/۰۹۹۱	۰/۲۷۳۱	۰/۰۴۱۸	۰/۵۵۵۳	
اردبیل	۰/۰۵۶۸	۰/۱۲۳۸	۰/۲۸۴۱	۰/۰۴۷۱	۰/۶۹۲۱	
اصفهان	۰/۰۷۷۳	۰/۱۵۹۲	۰/۲۸۶۴	۰/۰۸۶۱	۰/۴۷۹۴	
البرز	۰/۱۲۵۹	۰/۱۶۲۸	۰/۲۷۵۹	۰/۰۶۷۴	۰/۳۲۶۰	
ایلام	۰/۰۷۶۷	۰/۱۷۶۴	۰/۲۶۲۲	۰/۱۱۷۶	۱/۳۹۱۹	
بوشهر	۰/۱۲۰۵	۰/۱۳۳۳	۰/۳۲۲۳	۰/۲۰۲۰	۱/۱۱۵۶	
تهران	۰/۰۸۰۳	۰/۲۰۰۱	۰/۲۸۸۲	۰/۱۱۹۲	۰/۳۰۱۰	
چهار محال بختیاری	۰/۰۶۰۵	۰/۱۲۵۳	۰/۲۶۹۰	۰/۰۴۷۸	۰/۸۷۸۰	
خراسان جنوبی	۰/۱۲۷۸	۰/۱۲۳۷	۰/۳۱۵۴	۰/۰۴۴۶	۱/۱۹۴۵	
خراسان رضوی	۰/۰۶۸۰	۰/۱۱۳۱	۰/۲۸۸۵	۰/۰۵۵۷	۰/۴۶۳۶	
خراسان شمالی	۰/۰۹۴۲	۰/۱۰۴۷	۰/۲۹۶۹	۰/۰۴۱۱	۰/۷۹۷۶	
خوزستان	۰/۰۵۳۴	۰/۱۱۷۱	۰/۲۱۰۶	۰/۱۸۳۵	۰/۵۴۳۷	
زنجان	۰/۰۷۴۴	۰/۱۲۷۹	۰/۳۰۲۱	۰/۰۵۲۵	۰/۸۴۹۸	
سمنان	۰/۱۱۴۲	۰/۱۸۷۸	۰/۲۸۷۳	۰/۰۸۹۷	۱/۳۴۹۶	
سیستان و بلوچستان	۰/۰۳۹۶	۰/۰۵۹۲	۰/۱۵۲۳	۰/۰۳۰۵	۰/۵۴۸۷	
فارس	۰/۰۷۹۳	۰/۱۴۳۱	۰/۲۶۷۲	۰/۰۵۹۰	۰/۵۴۴۶	
قزوین	۰/۰۷۴۲	۰/۱۲۴۶	۰/۲۸۲۷	۰/۰۷۰۷	۰/۶۳۴۶	
قم	۰/۰۵۹۲	۰/۱۵۱۷	۰/۲۴۶۹	۰/۰۵۱۴	۰/۵۱۰۲	
کردستان	۰/۰۸۸۱	۰/۰۹۳۴	۰/۲۸۱۱	۰/۰۴۲۴	۰/۷۰۵۴	
کرمان	۰/۰۵۷۴	۰/۱۲۹۵	۰/۲۵۶۹	۰/۰۵۶۲	۰/۵۸۹۶	
کرمانشاه	۰/۰۵۸۴	۰/۱۲۱۷	۰/۲۴۲۵	۰/۰۵۵۰	۰/۷۳۰۶	
کهگیلویه و بویر احمد	۰/۰۹۹۰	۰/۱۵۲۱	۰/۲۲۵۴	۰/۲۱۸۸	۱/۲۱۸۳	
گلستان	۰/۰۸۵۱	۰/۱۱۰۴	۰/۲۷۷۴	۰/۰۴۰۱	۰/۷۱۴۰	
گیلان	۰/۰۹۱۶	۰/۱۳۷۱	۰/۳۰۶۳	۰/۰۵۱۶	۰/۷۶۳۵	
لرستان	۰/۰۴۱۱	۰/۱۲۴۶	۰/۲۴۲۶	۰/۰۴۰۵	۰/۶۴۹۸	
مازندران	۰/۰۷۶۹	۰/۱۶۹۷	۰/۳۰۸۸	۰/۰۶۶۵	۰/۷۰۹۴	
مرکزی	۰/۰۷۸۱	۰/۱۳۱۷	۰/۲۸۹۵	۰/۰۸۸۱	۰/۶۷۷۵	
هرمزگان	۰/۰۷۶۷	۰/۰۹۵۶	۰/۲۵۴۶	۰/۰۸۴۱	۰/۷۷۷۲	
همدان	۰/۰۶۲۱	۰/۱۱۴۸	۰/۲۸۲۲	۰/۰۵۱۲	۰/۶۵۸۰	
یزد	۰/۰۹۶۲	۰/۱۶۹۱	۰/۳۰۰۱	۰/۱۰۱۲	۰/۹۲۸۷	

منبع: سالنامه‌ی آمار کشور (۱۳۹۱)، نتایج تفصیلی سرشماری عمومی نفوس و مسکن ۱۳۹۰، (۱۳۹۱)

مراجع

- سالنامه آماری کشور، ۱۳۹۰. (۱۳۹۱). مرکز آمار ایران، تهران.
- دهخدا، علی اکبر. (۱۳۵۲). لغت‌نامه دهخدا. موسسه انتشارات و چاپ دانشگاه تهران، تهران.
- زنجانی، حبیب‌اله (۱۳۷۶). تحلیل جمعیت‌شناختی. انتشارات سمت، تهران.
- میرزا مصطفی، سید مهدی، قاسمی، پروانه. (۱۳۹۲). عوامل موثر بر مهاجرت‌های استانی با استفاده از مدل جاذبه، فصلنامه سیاست‌های مالی و اقتصادی، سال اول، شماره ۳، پاییز ۱۳۹۲.
- معین، محمد. (۱۳۶۳). فرهنگ معین. انتشارات امیرکبیر، تهران.
- نتایج تفصیلی سرشماری عمومی نفوس و مسکن ۱۳۹۰، (۱۳۹۱). مرکز آمار ایران، تهران.
- Ferrari, S. and Cribari, F. (2004), Beta Regression for Modelling Rates and Proportions, *Journal of Applied Statistics*, **31**, 799-815.
- Figuroa-Zúñiga, J. I., Arellano-Valle, R. B. and Ferrari, S. L. (2013), Mixed Beta Regression: A Bayesian Perspective, *Computational Statistics & Data Analysis*, **61**, 137– 147.
- Paolino, P. (2001), Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables, *Political Analysis*, **9**, 325-346.
- Spiegelhalter, D., Thomas, Andrew., Best, Nicky., Lunn, Dave. (2003), *Maximum WINBUGS user manual*, <http://www.mrc-bsu.cam.ac.uk/bugs>.