

تحليل آماری داده‌های طول مدت بیکاری رند شده ایران

مجری:

مریم میرزایی

همکاران:

اشکان شباک

مجتبی گنجعلی

تابان باغفلکی



پژوهشکده‌ی آمار

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

زمستان ۱۳۹۹

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

کد شناسه: RP-0004

عنوان: تحلیل آماری داده‌های طول مدت بیکاری رندشده ایران

مجری: مریم میرزایی

همکاران: اشکان شباک، مجتبی گنجعلی، تابان باغفلکی

گروه پژوهشی: پردازش داده‌ها و اطلاع‌رسانی

تاریخ انتشار: زمستان ۱۳۹۹

نوبت انتشار: اول

طراح جلد: ساناز مهندسی

صفحه‌آرا: نجمه ناظریان

❖ حق مالکیت معنوی این طرح پژوهشی متعلق به پژوهشکده‌ی آمار است و نقل مطالب فقط با ذکر مأخذ مجاز است.



پژوهشکده‌ی آمار

تهران، خیابان دکتر فاطمی، خیابان باباطاهر، خیابان سرتیپ فکوری،

شماره‌ی ۱۴۵

۳- ۰۲۱ ۸۸۶۳۰۴۴۰

www.srtc.ac.ir

پیش‌گفتار

بیکاری به عنوان یکی از عوامل عقب‌ماندگی کشور و جامعه، پیامدهای منفی بسیاری هم از نظر فردی و هم از نظر اجتماعی در پی دارد که کمترین آن‌ها عبارتند از: احساس بی‌ارزشی، فقر و تنگدستی، بزه‌کاری، و بیماری‌های جسمی و روانی.

از آنجایی که بازار کار هر کشوری نشان‌دهنده رشد و توسعه آن کشور در زمینه اقتصاد محسوب می‌شود، مطالعه و بررسی آن باید به صورت گسترده‌تر انجام شود و تنها به آمارهای توصیفی آن نباید بسنده کرد. طول مدت بیکاری که یکی از متغیرهای مهم است و در آمارگیری نیروی کار مرکز آمار ایران نیز ثبت می‌شود، بیانگر مدت زمانی است که یک فرد جویای کار بیکار است. برای پیدا کردن طول مدت بیکاری در دو فصل پیاپی از افراد طول مدت بیکاری را پرسیده و ثبت می‌کنند. همچنین، عامل‌های مختلفی نیز مانند جنسیت، سن، وضعیت تاهل، وضعیت تحصیلی، محل اقامت و غیره نیز ثبت می‌شوند. یکی از سؤال‌های مورد علاقه یافتن عوامل مؤثر بر طول مدت بیکاری است.

از طرفی، گرد کردن اطلاعات عددی پدیده‌ای است که در صورت دقیق اظهار نکردن اطلاعات رخ می‌دهد. یعنی فرد ممکن است طول مدت بیکاری خود را به نزدیک‌ترین مقدار ممکن آن گرد کند. با رخ دادن این پدیده روش‌های سنتی، به این دلیل که خطای حاصل از گرد کردن را در نظر نمی‌گیرد، ممکن است منجر به نتایج غیردقیق شود و روش‌های تعمیم‌یافته تحلیل داده‌های طول مدت بیکاری مورد استفاده قرار می‌گیرد تا نتایج دقیق‌تری حاصل شود. بنابراین هدف از این طرح پژوهشی تعمیم مدل‌های طول مدت برای تحلیل داده‌های طول مدت بیکاری در ایران با مقادیر گرد شده است.

در این طرح پژوهشی ابتدا مقدمه‌ای بر موضوع ارائه می‌شود. سپس مفاهیم اصلی مورد استفاده از جمله تابع بقا، نرخ خطر و توزیع‌های مهم مورد استفاده ذکر می‌شوند. مدل‌های معمول و مدل‌های با در نظر گرفتن انباشتگی (که به علت گرد کردن طول مدت بیکاری ایجاد می‌شود) مورد تحلیل قرار گرفته و در انتها این مدل‌ها برای تحلیل داده‌های شبیه‌سازی شده و داده‌های مرکز آمار استفاده می‌شوند.

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

ژوژوشکده‌ی آمار

فهرست مطالب

۱	کلیات تحقیق
۱-۱-۱	مقدمه
۳-۱-۲	بیان مسئله و هدف تحقیق
۴-۱-۳	روش تحقیق
۷	مفاهیم اصلی
۷-۱-۲	مقدمه
۸-۲-۲	طول مدت بیکاری
۹-۳-۲	تحلیل داده‌های زمان تا رخداد پیشامد
۹-۱-۳-۲	تابع بقا
۱۰-۲-۳-۲	تابع نرخ خطر
۱۱-۳-۳-۲	داده‌های سانسور شده
۱۲-۴-۳-۲	توزیع‌های طول مدت بیکاری
۱۹-۴-۲	معیارهای انتخاب مدل
۱۹-۱-۴-۲	معیار اطلاع آکائیکه
۲۰-۲-۴-۲	معیار آکائیکه تصحیح شده
۲۰-۳-۴-۲	معیار اطلاع بیزی
۲۰-۵-۲	بهینه‌سازی
۲۱-۶-۲	معیارهای سنجش در شبیه‌سازی
۲۱-۱-۶-۲	معیار اریبی انباشتگی القاشده
۲۲-۲-۶-۲	معیار قدر مطلق اریبی
۲۲-۳-۶-۲	معیار قدر مطلق اریبی نسبی
۲۲-۴-۶-۲	معیار میانگین توان دوم خطا
۲۳	مدل‌سازی عوامل تأثیرگذار بر طول مدت بیکاری

۲۳	۱-۳-۱- مقدمه
۲۴	۲-۳-۱- مدل خطی تعمیم‌یافته
۲۵	۳-۳-۱- مدل‌های رگرسیونی بقا برای طول مدت بیکاری
۲۵	۳-۳-۱-۱- مدل رگرسیونی بقا با پاسخ نمایی
۲۶	۳-۳-۲- مدل رگرسیونی بقا با پاسخ گاما
۲۷	۳-۳-۳- مدل رگرسیونی بقا با پاسخ لگ‌نرمال
۲۷	۳-۳-۴- مدل رگرسیونی بقا با پاسخ وایبول
۲۸	۳-۳-۵- مدل رگرسیونی بقا با پاسخ لگ‌لوژیستیک
۲۹	۴-۳-۱- مدل‌های مورد بررسی برای طول مدت بیکاری با مقادیر انباشتگی و سانسور
۲۹	۳-۴-۱- انباشتگی
۳۰	۳-۴-۲- تعریف مدل

۳۵ مطالعه‌های شبیه‌سازی و تحلیل داده‌های نیروی کار

۳۵	۱-۴-۱- مقدمه
۳۵	۲-۴-۱- مطالعه شبیه‌سازی
۳۵	۴-۲-۱- مطالعه شبیه‌سازی ۱: داده‌های کامل
۴۲	۴-۲-۲- مطالعه شبیه‌سازی ۲: با در نظر گرفتن سانسور در داده‌ها
۵۳	۴-۳-۱- تحلیل داده‌های نیروی کار با استفاده از مدل پیشنهادی
۵۳	۴-۳-۱-۱- توصیف داده‌ها
۵۵	۴-۳-۲- تحلیل داده‌ها

۶۳ بحث و نتیجه‌گیری

۶۵ مرجع‌ها

۶۹ واژه‌نامه

فهرست جدول‌ها

- جدول ۴-۱- نتایج مطالعه شبیه‌سازی مدل رگرسیون نمایی با انباشتگی با اندازه نمونه ۱۰۰۰ ۳۸
- جدول ۴-۲- نتایج مطالعه شبیه‌سازی مدل رگرسیون گاما با انباشتگی با اندازه نمونه ۱۰۰۰ ۳۹
- جدول ۴-۳- نتایج مطالعه شبیه‌سازی مدل رگرسیون وایبول با انباشتگی با اندازه نمونه ۱۰۰۰ ۴۰
- جدول ۴-۴- نتایج مطالعه شبیه‌سازی مدل رگرسیون لگ نرمال با انباشتگی با اندازه نمونه ۱۰۰۰ ۴۱
- جدول ۴-۵- نتایج مطالعه شبیه‌سازی مدل رگرسیون نمایی با سه نقطه انباشتگی و نرخ سانسور ۱۰ درصد با اندازه نمونه مختلف ۴۳
- جدول ۴-۶- نتایج مطالعه شبیه‌سازی مدل رگرسیون لگ نرمال با سه نقطه انباشتگی و نرخ سانسور ۱۰ درصد با اندازه نمونه مختلف ۴۴
- جدول ۴-۷- نتایج مطالعه شبیه‌سازی مدل رگرسیون وایبول با سه نقطه انباشتگی و نرخ سانسور ۱۰ درصد با اندازه نمونه مختلف ۴۴
- جدول ۴-۸- نتایج مطالعه شبیه‌سازی مدل رگرسیون لگ لوژستیک با سه نقطه انباشتگی و نرخ سانسور ۱۰ درصد با اندازه نمونه مختلف ۴۵
- جدول ۴-۹- نتایج مطالعه شبیه‌سازی مدل رگرسیون نمایی با سه نقطه انباشتگی و نرخ سانسور ۱۵ درصد با اندازه نمونه مختلف ۴۵
- جدول ۴-۱۰- نتایج مطالعه شبیه‌سازی مدل رگرسیون نمایی با سه نقطه انباشتگی و نرخ سانسور ۲۰ درصد با اندازه نمونه مختلف ۴۶
- جدول ۴-۱۱- نتایج مطالعه شبیه‌سازی مدل رگرسیون لگ نرمال با سه نقطه انباشتگی و نرخ سانسور ۱۵ درصد با اندازه نمونه مختلف ۴۷
- جدول ۴-۱۲- نتایج مطالعه شبیه‌سازی مدل رگرسیون لگ نرمال با سه نقطه انباشتگی و نرخ سانسور ۲۰ درصد با اندازه نمونه مختلف ۴۸
- جدول ۴-۱۳- نتایج مطالعه شبیه‌سازی مدل رگرسیون وایبول با سه نقطه انباشتگی و نرخ سانسور ۱۵ درصد با اندازه نمونه مختلف ۴۹
- جدول ۴-۱۴- نتایج مطالعه شبیه‌سازی مدل رگرسیون وایبول با سه نقطه انباشتگی و نرخ سانسور ۲۰ درصد با اندازه نمونه مختلف ۵۰
- جدول ۴-۱۵- نتایج مطالعه شبیه‌سازی مدل رگرسیون لگ لوژستیک با سه نقطه انباشتگی و نرخ سانسور ۱۵ درصد با اندازه نمونه مختلف ۵۱

- جدول ۴-۱۶- نتایج مطالعه شبیه‌سازی مدل رگرسیون لگ‌لوژیستیک با سه نقطه انباشتگی و نرخ سانسور ۲۰ درصد با اندازه نمونه مختلف..... ۵۲
- جدول ۴-۱۷- تغییر وضعیت افراد بیکار بهار ۱۳۹۸ به اشتغال در تابستان ۱۳۹۸..... ۵۴
- جدول ۴-۱۸- آمار توصیفی داده‌های کامل طول مدت بیکاری سال ۱۳۹۸..... ۵۵
- جدول ۴-۱۹- درصد انباشتگی طول مدت بیکاری خوداظهاری شده در بهار ۱۳۹۸..... ۵۵
- جدول ۴-۲۰- برآورد پارامترها و خطای استاندارد مدل با پاسخ چندجمله‌ای برای سن و تحصیلات..... ۵۸
- جدول ۴-۲۱- معیارهای سنجش مدل‌های رگرسیونی بقا با انباشتگی..... ۵۸
- جدول ۴-۲۲- نتایج اجرای مدل رگرسیون نمایی با انباشتگی روی داده‌های طول مدت بیکاری..... ۵۹
- جدول ۴-۲۳- نتایج اجرای مدل رگرسیون وایبول با انباشتگی روی داده‌های طول مدت بیکاری..... ۶۰
- جدول ۴-۲۴- نتایج اجرای مدل رگرسیون لگ‌نرمال با انباشتگی روی داده‌های طول مدت بیکاری..... ۶۱
- جدول ۴-۲۵- نتایج اجرای مدل رگرسیون لگ‌لوژیستیک با انباشتگی روی داده‌های طول مدت بیکاری..... ۶۲

فهرست شکل‌ها

- شکل ۱-۱- توزیع فراوانی خوداظهاری طول مدت بیکاری در بهار ۹۵..... ۵
- شکل ۲-۱- توزیع فراوانی خوداظهاری طول مدت بیکاری در بهار ۹۶..... ۵
- شکل ۳-۱- توزیع فراوانی خوداظهاری طول مدت بیکاری در بهار ۹۷..... ۶
- شکل ۴-۱- توزیع فراوانی خوداظهاری طول مدت بیکاری در بهار ۹۸..... ۶
- شکل ۱-۴- توزیع فراوانی خوداظهاری طول مدت بیکاری در بهار ۱۳۹۸..... ۵۶

کلیات تحقیق

۱-۱- مقدمه

بیکاری یکی از معضلات جامعه به شمار می‌رود که می‌تواند نتایج جبران ناپذیری را هم بر خود فرد و هم بازار کار و اقتصاد جامعه بگذارد. به صورتی که هزینه‌های اقتصادی بیکاری و طول مدت آن در سطح کلان و جامعه نیز قابل تأمل و تأثیرگذار است. برای نمونه به درآمد فرد شاغل، مالیات تعلق می‌گیرد و بر این اساس دولت می‌تواند سهمی از تولید نهایی وی را اخذ کند، اما وقتی فرد بیکار می‌شود، نمی‌توان مالیاتی از وی اخذ کرد. در واقع تجربه بیکاری برای کسانی که به داشتن شغل ثابت و مطمئن عادت کرده‌اند، ناراحت کننده است. نتیجه مستقیم بیکاری، از دست دادن درآمد است و فشار اقتصادی زیادی به افراد وارد می‌کند. افرادی که برای یک دوره زمانی نسبتاً کوتاه بیکار می‌شوند، استانداردهای زندگی آنها کاهش می‌یابد و علاوه بر عواقب اقتصادی، پیامدهای اجتماعی نیز با طول مدت بیکاری افزایش می‌یابد.

بیکاری از لحاظ طول مدت آن، معمولاً به دو نوع کوتاه مدت و بلندمدت تقسیم می‌شود و در شاخص‌های بین‌المللی بیکاری با طول مدت حداکثر ۶ ماه را کوتاه مدت و با مدت حداقل ۱۲ ماه را بلندمدت تعریف کرده‌اند. در مجموع بیکاری کوتاه‌مدت نگرانی خاصی ندارد به ویژه وقتی که فرد، تحت پوشش بیمه بیکاری و حمایت‌هایی از این قبیل است. اما بیکاری بلندمدت تأثیرات نامطلوبی دارد برای نمونه موجب از دست دادن عایدی و کاهش اشتغال‌پذیری کارجو می‌شود. اگر بیکاری کوتاه مدت باعث شود کارجو شغل مورد علاقه و دلخواه خود را بیابد، این مدت زمان، حتی مطلوب هم است؛ در حالیکه در بیکاری بلندمدت این چنین نیست. مطالعه بازار کار تنها با تجزیه و تحلیل متغیرهای ثابت^۱ مانند نرخ‌های اشتغال و بیکاری و رسم شکل‌های توصیفی آن کافی نیست و این موضوع باید عمیق‌تر مورد تحلیل و بررسی قرار گیرد.

¹ Static

یکی از متغیرهای مهم در این زمینه طول مدت بیکاری است. طول مدت بیکاری متغیر مهمی است که می‌تواند تغییرات در بازار را توضیح دهد (بلانچارد و دیاموند، ۱۹۹۲).

طول مدت بیکاری مدت زمانی است که یک فرد بیکار، و جویای کار است. بنابراین باید به بررسی طول مدت بیکاری و عوامل مؤثر بر آن بپردازیم تا بتوانیم مشکلات ناشی از این پدیده را با چاره‌اندیشی‌های مناسب رفع کنیم.

در سال‌های اخیر مدل‌های تحلیل زمان تا رخداد پیشامد^۲ مختلف و تکنیک‌های زمان رخداد پیشامد متفاوتی برای مدل‌بندی طول مدت بیکاری در جامعه‌های مختلف استفاده شده است. توسعه‌های جدید در روش‌های اقتصادی برای تحلیل داده‌های نیروی کار توسط مفیت (۱۹۹۹) انجام شده است. بسیاری از محققان به مطالعه اثر تحصیلات بر طول مدت بیکاری پرداختند.

در یک مطالعه توسط الیکانن (۲۰۰۳) بر طول مدت بیکاری در فنلاند سطح تحصیلات به عنوان یک عامل به شدت مثبت به ویژه برای زنان گزارش شده است. طبق کیتونن (۱۹۹۷) سطح تحصیلات پایین احتمال دوباره استخدام شدن را افزایش می‌دهد، اما در سطح تحصیلات بالاتر این ارتباط برعکس می‌شود. این در حالی است که، سطح تحصیلات بالا طول مدت بیکاری را در بریتانیا، بلژیک و ایرلند کوتاه می‌کند، در حالیکه در یونان و اسپانیا به نظر می‌رسد که سطح تحصیلات تأثیر زیادی بر طول مدت بیکاری ندارد (داگستینو و میلی، ۲۰۰۰). تعدادی دیگر از محققین به تفسیر بیکاری در بعد اجتماعی پرداختند و یافتند که بیکاری بر رضایتمندی از زندگی (وینکلمن، ۱۹۹۸، کلارک و اوزوالد، ۱۹۹۴)، دستمزد (روهام، ۱۹۹۱ و آرولاپالم ۲۰۰۱) و احتمال بیکاری در آینده (آرولاپالم و همکاران، ۲۰۰۰) تأثیر می‌گذارد.

کارول (۲۰۰۶) نیز متغیرهای مؤثر بر اشتغال مجدد فرد بیکار را با استفاده از چارچوب مدل‌بندی طول مدت بررسی کرده است. او دریافت که متغیرهای مرتبط با افزایش دستمزد، با نرخ‌های بالای خروج از بیکاری به اشتغال مرتبط هستند. این متغیرهای دستمزد، شامل تجربه‌ی شغلی و مدرک تحصیلی است. موکویاما و شاهین (۲۰۰۹) علل افزایش طول مدت بیکاری در ایالات متحده در سال‌های اخیر را بررسی کرده‌اند. آن‌ها نشان دادند که هم کاهش بیکاری و هم افزایش در پراکندگی دستمزد می‌توانند به طور معنی‌داری باعث طولانی شدن متوسط طول مدت بیکاری شوند. همچنین، کاوکلیور و همکاران (۲۰۰۹) به مطالعه اثر سن، جنس، سطح تحصیلات و محل اقامت به عنوان عوامل مؤثر در طول مدت بیکاری در رومانی، اتریش، اسلوانی، کرواسی و مقدونیه پرداختند و در این راستا از یک مدل متناسب کاکس استفاده کردند و به این نتیجه رسیدند که افراد بیکار با سطح تحصیلات بالاتر، موقعیت بهتری در بازار کار دارند.

در هنگام ثبت و گردآوری داده‌ها ممکن است افراد مصاحبه‌شونده از اظهار اطلاعات طول مدت بیکاری دقیق خود خودداری کنند و در واقع اطلاعات خود را گرد کنند، که در این صورت استفاده از روش‌های معمول و سنتی برای تحلیل اینگونه داده‌ها به دلیل اینکه خطای حاصل از گردشدن را در نظر نمی‌گیرد، صحیح نیست و باید مدل‌های مناسب برای تحلیل چنین داده‌هایی مورد استفاده قرار گیرد. در همین جهت به معرفی و تعمیم مدل‌های تحلیل زمان تا رخداد پیشامد برای داده‌های گردشده پرداخته می‌شود.

² Time to Event Analysis

۱-۲- بیان مسئله و هدف تحقیق

بیکاری موضوعی است که همواره از آن به عنوان معضل اقتصادی و اجتماعی یاد شده است و سعی می‌شود دلایل بروز آن با استفاده از تئوری‌های متفاوتی بیان گردد اما آنچه بیش از این موضوع اهمیت دارد طول مدتی است که طی آن فرد بیکار به دنبال شغل است. بنابر نظر کارشناسان یکی از شاخص‌های مهم در شناخت معضل بیکاری، «طول مدت بیکاری» افراد بیکار است. منظور از طول مدت بیکاری، مدت زمانی است که یک فرد بیکار، جویای کار است.

برای تحلیل طول مدت بیکاری از مدل‌های تحلیل زمان تا رخداد پیشامد استفاده می‌شود. نکته قابل توجه این است که مدل‌های زمان تا رخداد پیشامد در تحلیل داده‌های سایر علوم هم به وفور استفاده می‌شود. این پیشامد در پزشکی، طول زمان تا پیشرفت یک بیماری، زمان تا پاسخ به اثر یک درمان خاص، زمان تا عود یک بیماری، مرگ و یا هر پیشامد مورد علاقه دیگر در هر شاخه‌ای است. روش‌های تحلیل داده‌های زمان تا رخداد پیشامد علاوه بر اینکه در علوم پزشکی کاربرد دارد، در قابلیت اعتماد صنعتی، علوم اجتماعی و تجارت نیز کاربرد دارند. مثال‌هایی از داده‌های زمان تا رخداد پیشامد در سایر علوم عبارتند از طول عمر وسایل الکترونیک در مؤلفه‌ها و یا سیستم‌ها (قابلیت اعتماد صنعتی)، طول مدت بازگشت به جرم برای یک زندانی با عفو مشروط آزاد شده (جرم‌شناسی)، طول مدت زمان اولین ازدواج (اجتماعی)، طول مدت اشتراک روزنامه یا مجله (بازاریابی)، طول مدت ادعای گرامت کارگران (بیمه) و طول مدت بیکاری (اقتصاد).

یکی از ویژگی‌های مهم و متمایز این داده‌ها آن است که معمولاً شامل مشاهدات ناکامل هستند. این حالت زمانی رخ می‌دهد که مقدار دقیق طول مدت بیکاری فرد مصاحبه‌شده ثبت نشده و تنها اطلاع موجود از آن این است که از یک مقداری بزرگتر است. به این مشاهدات سانسور شده می‌گویند. زمانی که مشاهدات سانسور شده وجود ندارد، زمان‌های تا رخداد پیشامد کامل هستند.

در بحث گردآوری و ثبت اطلاعات مربوط به طول مدت بیکاری افراد بسیاری از آن‌ها از ثبت اطلاعات صحیح خود اجتناب می‌کنند و باعث نابهنجاری در اطلاعات صحیح یا به عبارتی باعث گردش داده‌ها یا اطلاعات خود می‌شوند که ممکن است این پدیده به صورت تصادفی رخ دهد اما در اکثر موارد به صورت عمدی رخ می‌دهد که در این صورت اطلاعات به دست آمده غیر دقیق هستند و به این پدیده‌ای که رخ می‌دهد انباشتگی یا گردش می‌گویند.

اگر داده‌ها به اشتباه تصور شوند و بدون خطای حاصل از گردش در نظر گرفته شوند تحلیل بدون در نظر گرفتن این نکته ممکن است منجر به نتایج اریب شود. برای تحلیل آماری چنین داده‌هایی استفاده از روش‌های سنتی مناسب به نظر نمی‌رسد به دلیل اینکه این مقادیر را مانند بقیه مقادیر در نظر گرفته و اریبی ناشی از گردش را در نظر نمی‌گیرند.

جهت مدل‌بندی داده‌های طول مدت بیکاری چه در حضور داده‌های سانسور شده و چه بدون داده‌های سانسور شده پیشنهادها بسیاری صورت گرفته که این مدل‌ها را برای داده‌های گردشده نمی‌توان مورد استفاده قرار داد زیرا خطای حاصل از گردش داده‌ها را در نظر نمی‌گیرد.