

# تحلیل آماری داده‌های طول مدت بیکاری رُندشده ایران با استفاده از مدل‌های مخاطره‌های رقابتی

مجری

مریم میرزایی

همکاران

تابان باغفلکی

اشکان شباک



ژئوشکده‌ی آمار

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

بهار ۱۴۰۲



# بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

کد شناسه: RP-0104

عنوان: تحلیل آماری داده‌های طول مدت بیکاری رُندشده ایران با استفاده از مدل‌های مخاطره‌های رقابتی

مجری: مریم میرزایی

همکاران: تابان باغفلکی، اشکان شباک

گروه پژوهشی: پردازش داده‌ها و اطلاع‌رسانی

تاریخ انتشار: بهار ۱۴۰۲

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- ۸۸۶۳۰۴۴۰ ۰۲۱

[www.srtc.ac.ir](http://www.srtc.ac.ir)



به نام خداوند جان و خرد

## پیش‌گفتار

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. 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. یکی از متغیرهای مورد علاقه در تحلیل داده‌های نیروی کار طول مدت بیکاری است، این متغیر 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، متغیرهای مختلفی از جمله جنسیت، سن، وضعیت تاهل، وضعیت تحصیلی، محل اقامت و غیره نیز برای هر فرد ثبت می‌شوند. دو مسئله‌ای که برای طول مدت بیکاری مطرح می‌شود انباشتگی و مخاطره‌های رقابتی است. در نظرسنجی نیروی کار افراد ممکن است اطلاعات خود را به صورت دقیق اظهار نکنند، یعنی ممکن است طول مدت بیکاری خود را به نزدیک‌ترین مقدار ممکن آن گرد یا رُند کنند. علاوه بر این دو علت اشتغال و غیر فعال شدن با یکدیگر به رقابت می‌پردازند تا فرد را از وضعیت بیکاری خارج کنند. به طور کلی می‌توان گفت داده‌های طول مدت بیکاری با دو مسئله انباشتگی و مخاطره‌های رقابتی مواجه هستند که برای مدل‌بندی آن‌ها روش‌های سنتی به دلیل نادیده گرفتن خطای حاصل از رُند شدن مناسب نیستند (در نظر نگرفتن خطای حاصل از رُند شدن منجر به نتایج غیردقیق می‌شود). هدف از طرح پژوهشی حاضر تعمیم مدل مخاطره‌های رقابتی برای تحلیل داده‌های طول مدت بیکاری رُند شده در ایران به منظور شناسایی عوامل تأثیرگذار بر طول مدت بیکاری است.

مطالعه حاضر به این صورت سازماندهی شده است؛ ابتدا مقدمات موضوع مورد بحث قرار می‌گیرد، سپس مفاهیم اصلی از جمله تابع بقا، نرخ خطر، توزیع‌های پیکاری در تحلیل داده‌های مدت زمان تا رخداد پیشامد و مدل‌های رگرسیونی برای تحلیل این داده‌ها مطرح می‌شوند. مفاهیم مخاطره‌های رقابتی و انباشتگی و همچنین مدل‌های مناسب در رابطه با مفاهیم مطرح شده، مورد بحث قرار می‌گیرند. پیشنهادی برای مدل‌بندی توأم مخاطره‌های رقابتی و انباشتگی ارائه می‌شود. به دنبال آن، مدل پیشنهادی با استفاده از مطالعه‌های شبیه‌سازی مورد بررسی قرار می‌گیرد و در نهایت برای تحلیل داده‌های نیروی کار مرکز آمار مورد استفاده می‌شود.

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی



# فهرست مطالب

۱	کلیات تحقیق
۱-۱-۱-۱	مقدمه
۳-۲-۱	تعاریف
۴-۳-۱	بیان مسئله و هدف تحقیق
۶-۴-۱	روش تحقیق
۸-۵-۱	توصیف داده‌ها

۱۷	مفاهیم اصلی
۱۷-۱-۲	مقدمه
۱۸-۲-۲	تحلیل داده‌های زمان تا رخداد پیشامد
۱۹-۲-۲-۱	توزیع‌های گسسته
۲۰-۲-۲-۲	توزیع‌های پیوسته
۲۲-۳-۲-۲	سانسور
۲۳-۴-۲-۲	برآوردگر کاپلان مهیر
۲۴-۱-۴-۲-۲	واریانس برآوردگر کاپلان مهیر
۲۵-۵-۲-۲	توزیع وایبل
۲۶-۶-۲-۲	توزیع لگ‌نرمال
۲۸-۷-۲-۲	توزیع لگ‌لوژستیک
۲۸-۸-۲-۲	مدل‌های رگرسیونی بقا برای طول مدت بیکاری
۲۹-۱-۸-۲-۲	مدل رگرسیونی بقا با پاسخ وایبل
۲۹-۲-۸-۲-۲	مدل رگرسیونی بقا با پاسخ لگ‌نرمال
۳۰-۳-۸-۲-۲	مدل رگرسیونی بقا با پاسخ لگ‌لوژستیک
۳۰-۹-۲-۲	مدل‌های رگرسیونی بقا برای طول مدت بیکاری با سانسور از راست و بازه‌ای
۳۱-۱۰-۲-۲	مدل خطرات متناسب کاکس
۳۱-۳-۲	مدل انباشتگی
۳۲-۱-۳-۲	مدل‌های رگرسیونی بقا برای مقادیر رُندشده بدون در نظر گرفتن مقادیر سانسور شده

۲-۳-۲- مدلهای رگرسیون بقا برای مقادیر رُندشده در حضور مقادیر سانسور شده: برای تحلیل داده‌های طول مدت بیکاری ایران	۳۳
۴-۲- معیارهای انتخاب مدل	۳۶
۱-۴-۲- معیار اطلاع آکائیکه	۳۶
۲-۴-۲- معیار آکائیکه تصحیح شده	۳۶
۳-۴-۲- معیار اطلاع بیزی	۳۶
۵-۲- معیارهای سنجش در شبیه‌سازی	۳۷
۱-۵-۲- معیار قدر مطلق ارزیابی	۳۷
۲-۵-۲- معیار قدر مطلق ارزیابی نسبی	۳۷
۳-۵-۲- معیار میانگین توان دوم خطا	۳۷
۴-۵-۲- احتمال پوشش	۳۷
۵-۵-۲- اعتبار سنجی متقابل	۳۸
۶-۲- آزمون نیکویی برازش	۳۸

## ۴۱ مدل‌های مخاطره‌های رقابتی

۱-۳- مقدمه	۴۱
۲-۳- تعاریف مختلف مخاطره‌های رقابتی	۴۳
۱-۲-۳- مخاطره‌های رقابتی به عنوان متغیر تصادفی دو متغیره	۴۳
۲-۲-۳- مخاطره‌های رقابتی به عنوان زمان‌های شکست پنهان	۴۴
۳-۳- مفاهیم اساسی	۴۴
۱-۳-۳- مخاطره‌های رقابتی به عنوان متغیر تصادفی دو متغیره	۴۴
۲-۳-۳- مخاطره‌های رقابتی به عنوان زمان‌های شکست پنهان	۴۶
۴-۳- روش‌های توصیفی برای داده‌های مخاطره‌های رقابتی	۴۶
۱-۴-۳- برآوردگر حد حاصلضربی و مخاطره‌های رقابتی	۴۶
۲-۴-۳- تابع وقوع تجمعی	۴۸
۱-۲-۴-۳- برآورد تجربی CIF	۴۸
۲-۲-۴-۳- برآورد ماکسیمم درست‌نمایی ناپارامتری CIF	۵۰
۳-۲-۴-۳- محاسبه برآوردگر CIF	۵۱
۴-۲-۴-۳- واریانس و بازه اطمینان برای برآوردگر CIF	۵۳
۵-۳- مدل خطر علت مشخص بر اساس مدل خطر متناسب کاکس	۵۴
۶-۳- مدل‌بندی داده‌های طول مدت بیکاری با استفاده از مخاطره‌های رقابتی	۵۵
۷-۳- مدل‌بندی داده‌های طول مدت بیکاری رُندشده با استفاده از مخاطره‌های رقابتی	۵۶

## ۵۹ مطالعه‌های شبیه‌سازی و تحلیل داده‌های نیروی کار

۱-۴- مقدمه	۵۹
۲-۴- مطالعه شبیه‌سازی مدل‌های مخاطره‌های رقابتی	۵۹
۳-۴- مطالعه شبیه‌سازی مدل‌های مخاطره‌های رقابتی برای داده‌های رُندشده	۶۷



۴-۴ - تحلیل داده‌های نیروی کار با استفاده از مدل پیشنهادی ..... ۷۷

۸۵ بحث و نتیجه‌گیری

۸۷ مراجع‌ها

۹۳ واژه‌نامه فارسی به انگلیسی

۹۷ واژه‌نامه انگلیسی به فارسی





# فهرست جدول‌ها

- جدول ۱-۱- تغییر وضعیت افراد بیکار بهار ۱۳۹۹ به اشتغال در تابستان ۱۳۹۹ ..... ۹
- جدول ۱-۲- فرمول‌های ریاضی پایه در تحلیل بقا ..... ۲۱
- جدول ۲-۲- توابع بقا متداول برای داده‌های طول مدت بیکاری ( $\Phi$  و  $\Phi$ ) به ترتیب بیانگر تابع چگالی و تابع توزیع  
تجمعی نرمال استاندارد هستند) ..... ۳۵
- جدول ۱-۳- داده‌های ترخیص یا مرگ ..... ۴۷
- جدول ۲-۳- داده‌های سقوط آسیب‌زا ..... ۵۱
- جدول ۳-۳- محاسبه *CIF* ..... ۵۲
- جدول ۱-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌لوژیستیک ..... ۶۱
- جدول ۲-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌لوژیستیک با سانسور ۳۰٪ ..... ۶۲
- جدول ۳-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌نرمال ..... ۶۳
- جدول ۴-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌نرمال با سانسور ۳۰٪ ..... ۶۴
- جدول ۵-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ وایبل ..... ۶۵
- جدول ۶-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ وایبل با سانسور ۳۰٪ ..... ۶۶
- جدول ۷-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ وایبل برای یک نقطه انباشتگی ..... ۶۸
- جدول ۸-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ وایبل با سه نقطه انباشتگی ..... ۶۹
- جدول ۹-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ وایبل برای پنج نقطه انباشتگی ..... ۷۰
- جدول ۱۰-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌لوژیستیک برای یک نقطه انباشتگی ..... ۷۱
- جدول ۱۱-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌لوژیستیک برای سه نقطه انباشتگی ..... ۷۲
- جدول ۱۲-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌لوژیستیک برای پنج نقطه انباشتگی ..... ۷۳
- جدول ۱۳-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌نرمال برای یک نقطه انباشتگی ..... ۷۴
- جدول ۱۴-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌نرمال برای سه نقطه انباشتگی ..... ۷۵
- جدول ۱۵-۴- نتایج مطالعه شبیه‌سازی مدل مخاطره رقابتی با پاسخ لگ‌نرمال برای پنج نقطه انباشتگی ..... ۷۶
- جدول ۱۶-۴- نتایج اعتبارسنجی متقابل مدل مخاطره رقابتی با پاسخ وایبل ..... ۷۷
- جدول ۱۷-۴- نتایج اعتبارسنجی متقابل مدل مخاطره رقابتی با پاسخ لگ‌لوژیستیک ..... ۷۷
- جدول ۱۸-۴- نتایج اعتبارسنجی متقابل مدل مخاطره رقابتی با پاسخ لگ‌نرمال ..... ۷۷
- جدول ۱۹-۴- معیارهای سنجش مدل مخاطره رقابتی ..... ۷۹

- جدول ۴-۲۰- نتایج اجرای مدل مخاطره رقابتی با پاسخ لگ‌لوژیستیک به علت شاغل شدن ..... ۷۹
- جدول ۴-۲۱- نتایج اجرای مدل مخاطره رقابتی با پاسخ لگ‌لوژیستیک به علت غیر فعال شدن ..... ۸۰
- جدول ۴-۲۲- نتایج احتمالات انباشتگی حاصل از اجرای مدل مخاطره رقابتی با پاسخ لگ‌لوژیستیک ..... ۸۰
- جدول ۴-۲۳- نتایج اجرای مدل مخاطره رقابتی با پاسخ لگ‌نرمال به علت شاغل شدن ..... ۸۱
- جدول ۴-۲۴- نتایج اجرای مدل مخاطره رقابتی با پاسخ لگ‌نرمال به علت غیر فعال شدن ..... ۸۱
- جدول ۴-۲۵- نتایج احتمالات انباشتگی حاصل از اجرای مدل مخاطره رقابتی با پاسخ لگ‌نرمال ..... ۸۲
- جدول ۴-۲۶- نتایج اجرای مدل مخاطره رقابتی با پاسخ وایبل به علت شاغل شدن ..... ۸۲
- جدول ۴-۲۷- نتایج اجرای مدل مخاطره رقابتی با پاسخ وایبل به علت غیر فعال شدن ..... ۸۳
- جدول ۴-۲۸- نتایج احتمالات انباشتگی حاصل از اجرای مدل مخاطره رقابتی با پاسخ وایبل ..... ۸۳

# فهرست شکل ها

- شکل ۱-۱- توزیع فراوانی خوداظهاری طول مدت بیکاری در بهار ۱۳۹۸ ..... ۷
- شکل ۱-۲- توزیع فراوانی خوداظهاری طول مدت بیکاری در بهار ۱۳۹۹ ..... ۸
- شکل ۱-۳- نمودار کاپلان مهیر طول مدت بیکاری به تفکیک جنسیت. .... ۱۰
- شکل ۱-۴- نمودار کاپلان مهیر طول مدت بیکاری بر اساس تحصیلات. .... ۱۰
- شکل ۱-۵- نمودار کاپلان مهیر طول مدت بیکاری بر اساس وضعیت تأهل. .... ۱۱
- شکل ۱-۶- نمودار کاپلان مهیر طول مدت بیکاری به تفکیک شهری یا روستایی بودن. .... ۱۱
- شکل ۱-۷- نمودار کاپلان مهیر طول مدت بیکاری بر اساس تعداد اعضای خانوار. .... ۱۲
- شکل ۱-۸-  $CIF$  برآورد شده برای دو نوع رویداد شاغل شدن و غیر فعال شدن. .... ۱۲
- شکل ۱-۹-  $CIF$  برآورد شده برای دو نوع رویداد شاغل شدن و غیر فعال شدن بر اساس جنسیت. .... ۱۳
- شکل ۱-۱۰-  $CIF$  برآورد شده برای دو نوع رویداد شاغل شدن و غیر فعال شدن بر اساس تحصیلات. .... ۱۳
- شکل ۱-۱۱-  $CIF$  برآورد شده برای دو نوع رویداد شاغل شدن و غیر فعال شدن بر اساس وضعیت تأهل. .... ۱۴
- شکل ۱-۱۲-  $CIF$  برآورد شده برای دو نوع رویداد شاغل شدن و غیر فعال شدن بر اساس شهری یا روستایی بودن. .... ۱۴
- شکل ۱-۱۳-  $CIF$  برآورد شده برای دو نوع رویداد شاغل شدن و غیر فعال شدن بر اساس تعداد اعضای خانوار. .... ۱۵
- شکل ۱-۲- نمودار تابع چگالی احتمال  $f(t)$ ، تابع توزیع  $F(t)$ ، تابع بقا  $S(t)$  و تابع خطر  $h(t)$  برای مدل  $W(\lambda, \alpha)$  برای  $\lambda = 1/5$  و مقادیر مختلف  $\alpha$ . .... ۲۶
- شکل ۲-۲- نمودار تابع چگالی احتمال  $f(t)$ ، تابع توزیع  $F(t)$ ، تابع بقا  $S(t)$  و تابع خطر  $h(t)$  برای مدل  $LN(\mu, \sigma^2)$  برای  $\mu = 1$  و مقادیر مختلف  $\sigma^2$ . .... ۲۷
- شکل ۳-۲- نمودار تابع چگالی احتمال  $f(t)$ ، تابع توزیع  $F(t)$ ، تابع بقا  $S(t)$  و تابع خطر  $h(t)$  برای مدل لگلوژیستیک برای  $\mu = 2$  و مقادیر مختلف  $\sigma$ . .... ۲۸
- شکل ۱-۳-  $CIF$  برآورد شده برای سه نوع رویداد. .... ۵۲
- شکل ۲-۳- تقسیم  $1 - S(t)$  به  $F_1(t)$ ،  $F_2(t)$  و  $F_3(t)$ . .... ۵۳

