بسم الله الرحمن الرحیم

دانشگاه یزد

دانشکده ریاضی

گروه آمار

پایان‌نامه

برای دریافت درجه‌ی کارشناسی ارشد
آمار

مدل نرخ شکست متناسب برای زمان‌های متوالی بر پایه‌ی سانسور آگاهی‌بخش

استاد راهنما: دکتر حمزه ترابی

استاد مشاور: دکتر عیسی محمودی

پژوهش و نگارش: نعیمه دهقانی

مهر ماه ۱۳۹۲

تقدیم به

پدر و مادر مهربانم

تحفه ی ناچیز ذهن کوچکم را

به بلندای بی کران وجودتان پیش کش می کنم

به آرزوی آن که قطره گون وجودم را

به بی انتهایی دریای سخاوتتان پذیرا باشید.

سپاس

حمد و سپاس ایزد منان را که با الطاف بی کران خود، طلیعه ی علم آموزی را در وجودم نهاد و اشتیاق پوییدن این مسیر نه چندان سهل را به من عنایت فرمود. پس از او، سپاس گزارم از پدر و مادر عزیزم که فروغ نگاهشان امید بخش زندگی ام است. بوسه بر دستان پر مهرشان می زنم و از خداوند می خواهم وجودشان را همیشه سرسبز و استوار گرداند.

از استاد مشاور گران قدرم، جناب آقای دکتر عیسی محمودی که راهنمایی های ارزنده ی خود را در راستای اعتلای این پایان نامه از من دریغ ننمودند، کمال تشکر و امتنان دارم. هم چنین از اساتید گرامی، دکتر دولتی و دکتر فلاح زاده که زحمت مطالعه و داوری این پایان نامه را بر عهده داشته اند، تشکر می نمایم.

با نهایت احترام، از اساتید بزرگوار گروه آمار دانشگاه یزد، چه آنان که افتخار شاگردی ایشان را داشتم و چه آنان که از محضر پر فیضشان محروم ماندم، سپاس گزاری می کنم.

از دوست خوبم خانم مشکوتی و خواهر و برادرم که مایه ی امید و دلگرمی من بوده اند، سپاس گزارم و از پروردگار، حسن عاقبت، سلامت و سعادت را برا ایشان خواستارم.

و اما...

زانوی ادب در برابر استاد ارجمندی می زنم که لحظه ای مرا در مسیر پر فراز و نشیب این پایان نامه به حال خود وانگذاشتند. استاد فرهیخته ای که همیشه و همه وقت راهنمایی هایشان، روشنی دیده ی تاریک ذهنم و راهکارهایشان، کلیدی بر درهای بسته ی مقابلم بوده است.

زحمات استاد راهنمای بزرگوارم جناب آقای دکتر حمزه ترابی را ارج می نهم و این پژوهش را اگر قدری است، با احترام و ارادت، پیش کش ایشان نیز می نمایم.

چکیده

امروزه تحلیل داده‌های بقا، کاربرد وسیعی در رشته‌های مختلف علوم از جمله پزشکی و مهندسی دارد. یکی از روش‌های متداول برای پیش‌بینی احتمال بقا، استفاده از مدل نرخ شکست متناسب کاکس است. در این مدل، برقراری فرض استقلال میان زمان سانسور و زمان پیشامد، یکی از فرض‌های اساسی به شمار می‌رود. ولی گاهی برقراری این فرض امکان‌پذیر نیست؛ در این موارد از سانسور وابسته استفاده می‌شود.

در این پایان‌نامه مدل نرخ شکست متناسب برای داده‌های زمان شکست یک متغیره و دو متغیره، ارائه می‌گردد. سپس شیوه‌ی استنباط نیم‌پارامتری برای این داده‌ها استفاده می‌شود و این شیوه تحت طرح سانسور آگاهی‌بخش گسترش می‌یابد. در ادامه، ویژگی برآوردگرهای پارامترها و توابع نرخ شکست پایه در حالت دو متغیره با طرح سانسور آگاهی‌بخش مورد بررسی قرار می‌گیرد. سرانجام این برآوردگرها با استفاده از شبیه‌سازی ارزیابی می‌گردد و هم‌چنین روش استفاده از داده‌های واقعی، نشان داده می‌شود.

فهرست مطالب

فهرست شکل‌ها

فهرست جدول‌ها

نمادها

پیش‌گفتار

در بیشتر پژوهش‌های مربوط به داده‌های طول عمر، عوامل بازدارنده‌ای هم‌چون زمان و هزینه باعث می‌شوند که آزمایش‌ها قبل از شکست همه‌ی واحدها، به پایان برسند؛ به داده‌های برآمده از این آزمایش‌ها، داده‌های سانسور شده گفته می‌شود.

در مطالعات بقا، بررسی داده‌های زمان شکست سانسور شده اهمیت بسزایی دارد. در تحلیل این نوع داده‌ها معمولاً برای اطمینان از تشخیص تابع بقا، فرض می‌شود که متغیرهای زمان شکست مستقل از متغیرهای سانسورکننده باشند؛ این فرض به عنوان سانسور ناآگاهی‌بخش در نظر گرفته می‌شود. از این سانسور در ساخت بیشتر برآوردگرهای تابع بقا مانند برآوردگر کاپلان و مایر[۱] [۲۰] استفاده شده است. هم‌چنین کمبل و فولدز[۲] [۱۱] در سال ۱۹۸۰، تسای و همکاران[۳] [۳۲] در سال ۱۹۸۶، دبروسکا[۴] [۱۴] در سال ۱۹۸۸، تسای و کرولی[۵] [۳۱] در سال ۱۹۸۸، پرنتایس و کای[۶] [۲۶] در سال ۱۹۹۲ و لین و یینگ[۷] [۲۳] در سال ۱۹۹۳ برآوردگرهای تابع بقای دو متغیره بر اساس این نوع سانسور ارائه نموده‌اند.

برای بهبود برآوردگرهای تابع بقا می‌توان از اطلاعات کمکی استفاده نمود؛ از این‌رو لین[۸] [۲۲] در سال ۱۹۹۴، کای و پرنتایس [۱۰] در سال ۱۹۹۵، اسپیکرمن و لین[۹] [۳۰] در سال ۱۹۹۸ و سنکاران و سریجا[۱۰] [۲۸] در سال ۲۰۰۷، نتایج مربوط به برآورد بهبودیافته‌ی تابع بقا را با استفاده از اطلاعات کمکی بر پایه‌ی سانسور ناآگاهی‌بخش بررسی نموده‌اند.

---

[۱] Kaplan and Meier

[۲] Campbel and Földes

[۳] Tsai et al.

[۴] Dabrowska

[۵] Tsai and Crowley

[۶] Prentice and Cai

[۷] Lin and Ying

[۸] Lin

[۹] Spiekerman and Lin

[۱۰] Sankaran and Sreeja

کلبفلیش و پرنتایس[۱۱] [۱۹] در سال ۲۰۰۲ و لاولس[۱۲] [۲۱] در سال ۲۰۰۳، مرور جامعی بر برآوردهای تابع بقا انجام داده‌اند.

در اکثر مسائل کاربردیِ تحلیل بقا، سانسور ناآگاهی‌بخش واقع‌بینانه نیست؛ بنابراین بررسی مسائلی که فرض استقلال متغیر سانسور، در آن ممکن نباشد، اهمیت ویژه‌ای می‌یابد. مدت زمان دو پیشامد متوالی را در نظر بگیرید. چنین نمونه‌هایی برای بحث در مورد پیشرفت بیماری‌های دومرحله‌ای مفید است؛ مانند بیماری AIDS که از دوره‌ی نهفتگی تا دوره‌ی بالینی پیشرفت می‌کند. (فاصله‌ی بین انتقال ویروس HIV تا تشخیص AIDS، دوره‌ی نهفتگی نامیده می‌شود). در تحلیل زمان‌های دو پیشامد متوالی، طول اولین مدت زمان روی دومین مدت زمان تأثیر می‌گذارد. در این موارد به دلیل وابستگی این دو زمان، دومین مدت زمان به‌وسیله‌ی یک متغیر وابسته‌ی مرتبط با مدت زمان اول، سانسور می‌شود که بدان سانسور آگاهی‌بخش گفته می‌شود.

وانگ و ولز[۱۳] [۳۵] در سال ۱۹۹۸، برآورد ناپارامتری زمان‌های متوالی بر اساس سانسور آگاهی‌بخش ارائه داده‌اند. براکرز و وراوربک[۱۴] [۹] در سال ۲۰۰۵، مسائل رگرسیونی برای تحلیل داده‌های زمان شکست متوالی بر پایه‌ی سانسور آگاهی‌بخش جزئی در حالت یک متغیره را به کار برده‌اند.

استنباط پیرامون زمان‌های متوالی بر پایه‌ی سانسور آگاهی‌بخش در حضور متغیرهای کمکی نیز، درخور توجه است. بدین منظور، اخیراً، سنکاران و سریجا [۲۹] در سال ۲۰۱۲، مدل نرخ شکست متناسب را برای چنین داده‌هایی ارائه نموده‌اند.

این پایان‌نامه شامل چهار فصل است. در فصل اول، مفاهیم و تعاریف اولیه‌ی مورد نیاز این پایان‌نامه بیان شده است. در فصل دوم، استنباط نیم‌پارامتری پیرامون مدل نرخ شکست متناسب برای داده‌های زمان شکست، در چهار حالت یک متغیره و دو متغیره، بدون در نظر گرفتن سانسور و با طرح سانسور انجام می‌شود. در فصل سوم، ویژگی برآوردگرها در حالت دو متغیره با طرح

[۱۱] Kalbfleisch and Prentice

[۱۲] Lawless

[۱۳] Wang and Wells

[۱۴] Braekers and Veraverbeke

سانسور آگاهی‌بخش بیان خواهد شد. در فصل چهارم، روش استنباطی معرفی شده برای داده‌های طول عمر دو متغیره با طرح سانسور آگاهی‌بخش، به کمک شبیه‌سازی و استفاده از داده‌های واقعی، ارزیابی شده است.

فصل ۱

مفاهیم و تعاریف اولیه

مقدمه

در این فصل برخی از مفاهیمی که در فصل‌های آتی مورد استفاده قرار می‌گیرند، به‌طور خلاصه ارائه شده است. در نخستین بخش، تحلیل بقا و مفاهیم مرتبط با آن، بیان شده و سپس در بخش ۲.۱ برخی از طرح‌های سانسور معرفی گردیده است. مفاهیم احتمالی، استنباطی و آنالیزی مورد نیاز به‌ترتیب در بخش‌های ۴.۱، ۵.۱ و ۶.۱ ارائه گردیده‌اند.

۱.۱ تحلیل بقا

به‌طور کلی، تحلیل بقا، مجموعه‌ای از تکنیک‌های آماری متنوع، جهت تحلیل متغیرهای تصادفی است که دارای مقادیر نامنفی هستند. مقدار این متغیر تصادفی، زمان شکست یک مؤلفه‌ی فیزیکی و یا زمان مرگ یک واحد بیولوژیک است. در تحلیل بقا با داده‌های سانسور شده مواجه هستیم که در بخش ۲.۱ معرفی می‌گردند. تابع بقا و تابع نرخ شکست از مفاهیم پایه‌ای در مطالعات بقا هستند که در ادامه به معرفی آن‌ها می‌پردازیم. مفاهیم بیان شده در این بخش برگرفته از مراجع [۱۵] و [۲۱] است.

۱.۱.۱ تابع بقا

احتمال این که واحدی بیشتر از زمان مشخص t عمر کند با تابع بقا محاسبه می‌شود و آن را با $\bar{F}(t)$ نمایش می‌دهیم؛ به عبارت دیگر،

$$\bar{F}(t) = P(T > t) = 1 - F(t).$$

زمانی که $f(.)$، تابع چگالی T، وجود داشته باشد، $\bar{F}(t)$ را می‌توان به‌صورت زیر نوشت:

$$\bar{F}(t) = \int_t^\infty f(u)du,$$

و تابع چگالی T از روی تابع بقا به‌صورت

$$f(t) = -\frac{d\bar{F}(t)}{dt},$$

است. تابع بقا، تابعی غیرنزولی و از چپ پیوسته است که

$$\bar{F}(0) = 1, \qquad \bar{F}(\infty) = \lim_{t\to\infty} \bar{F}(t) = 0.$$

همان‌طور که گفته شد تابع بقا نقش اساسی در مطالعات بقا ایفا می‌کند؛ زیرا به‌دست آوردن احتمالات بقا برای مقادیر مختلف t، اطلاعات مختصر و مفیدی از داده‌های بقا را فراهم می‌کند. تابع بقا، در مطالعات مربوط به طول عمر واحدهای صنعتی، **قابلیت اعتماد** نامیده می‌شود.

۲.۱.۱ تابع بقای دو متغیره

هرگاه واحدها شامل دو پیشامد باشند، تابع بقای توأم دو متغیر T_1 و T_2، به‌صورت زیر تعریف می‌شود:

$$\bar{F}(t_1, t_2) = P(T_1 > t_1, T_2 > t_2) = 1 - F(t_1) - F(t_2) + F(t_1, t_2).$$

قضیه ۱.۱.۱ فرض کنید (T_1, T_2) داده‌های دو متغیره با تابع بقای $\bar{F}_1(.)$ برای T_1 و $\bar{F}_2(.)$ برای T_2 باشند؛ در این صورت تابع بقای توأم به‌صورت زیر به‌دست می‌آید:

$$\bar{F}(t_1, t_2) = -\int_{u>t_1} \bar{F}_2(t_2|u) d\bar{F}_1(u),$$

که در آن $\bar{F}_2(t_2|u)$ تابع بقای شرطی T_2 به شرط T_1 است.

۳.۱.۱ تابع زیربقا

فرض کنید متغیرهای تصادفی $X_1, \cdots, X_n$ طول عمر n واحد تحت آزمایش در یک آزمون طول عمر باشند که هر واحد i با C_i از راست سانسور می‌شود و فرض کنید X_i و C_i از یکدیگر مستقل باشند. (T_i, δ_i) را بردار تصادفی از مشاهدات در نظر می‌گیریم که در آن $\delta_i = \mathrm{I}(X_i < C_i)$ و $T_i = \min(X_i, C_i)$، زمان پایان آزمایش هر کدام از واحدها است. فرض کنید F_i تابع توزیع T_i باشد؛ در این صورت توابع زیرتوزیع F_i به‌صورت

$$F^*_{1i}(t) = P(T_i < t, \delta_i = 1),$$

و

$$F^*_{2i}(t) = P(T_i < t, \delta_i = 0),$$

به‌دست می‌آید.

تعریف ۲.۱.۱ فرض کنید δ یک متغیر تصادفی دلخواه باشد که مقادیر ۰ و ۱ را اختیار می‌کند؛ در این صورت توابع زیربقا به‌صورت زیر تعریف می‌شوند:

$$\bar{F}^*_1(t) = P(T > t, \delta = 1),$$

$$\bar{F}^*_2(t) = P(T > t, \delta = 0).$$

قضیه ۳.۱.۱ فرض کنید توابع حقیقی $\bar{F}^*_1(.)$ و $\bar{F}^*_2(.)$ روی $[0, \infty)$، دو تابع زیربقا باشند؛ در این صورت

الف) $\bar{F}^*_1(.)$ و $\bar{F}^*_2(.)$ نامنفی و غیرصعودی هستند.

ب) $\bar{F}^*_1(.)$ و $\bar{F}^*_2(.)$ در نقطه‌ی صفر از راست پیوسته هستند و هم‌چنین $\bar{F}^*_1(0) < 1$ و $\bar{F}^*_2(0) < 1$.

ج) $\lim_{t\to\infty} \bar{F}^*_1(t) = 0$ و $\lim_{t\to\infty} \bar{F}^*_2(t) = 0$.

د) $\bar{F}^*_1(0) = 1 - \bar{F}^*_2(0)$.

۴.۱.۱ تابع نرخ شکست

یکی دیگر از مفاهیم اساسی در تحلیل بقا، تابع نرخ شکست است که نرخ لحظه‌ای مرگ یا شکست واحدی را در زمان t محاسبه می‌کند به شرط این که واحد مورد آزمایش بیشتر از زمان t عمر کند. تابع نرخ شکست به‌صورت زیر تعریف می‌شود:

$$\lambda(t) = \lim_{\Delta t\to\infty} \frac{P(t \leqslant T < t + \Delta t | T \geqslant t)}{\Delta t}. \qquad (1.1)$$

زمانی که $f(.)$، تابع چگالی T، موجود باشد، تابع نرخ شکست را می‌توان به‌صورت زیر نوشت:

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)} = -\frac{d\ln \bar{F}(t)}{dt}. \qquad (۲.۱)$$

لازم به یادآوری است که $\lambda(t)$ نامنفی است و $\int_{\circ}^{\infty} \lambda(u)du = \infty$. در ضمن تابع نرخ شکست انباشته در زمان t به‌صورت زیر محاسبه می‌گردد:

$$\Lambda(t) = \int_{\circ}^{t} \lambda(u)du.$$

روشن است که

$$\bar{F}(t) = \exp\big(-\Lambda(t)\big), \qquad (۳.۱)$$

و

$$f(t) = \lambda(t)\exp\big(-\Lambda(t)\big). \qquad (۴.۱)$$

۵.۱.۱ تابع نرخ شکست تکه‌ای-ثابت

در این زیربخش، تابع نرخ شکست تکه‌ای-ثابت تعریف می‌شود.

تعریف ۴.۱.۱ فرض کنید $\circ = a_{\circ} < a_{۱} < \cdots < a_m = \infty$ مقادیر مشخص و متمایز باشند. اگر تابع نرخ شکست متغیر T به‌صورت زیر باشد:

$$\lambda(t) = \lambda_j, \quad a_{j-۱} \leqslant t < a_j,$$

که λ_jها مقادیر مثبت برای $j = ۱, \cdots, m$ هستند، آنگاه T دارای تابع نرخ شکست تکه‌ای-ثابت است.

۶.۱.۱ مدل نرخ شکست متناسب

در بسیاری از شرایط کاربردی تحلیل بقا، طول عمر واحدها تحت تأثیر متغیرهای کمکی مانند سطح کلسترول یا فشار خون هستند و بررسی متغیر طول عمر واحدها بدون در نظر گرفتن متغیرهای کمکی و یافتن تابع نرخ شکست و تابع بقای آن کاری گمراه کننده است. در نتیجه به جای بررسی داده‌های طول عمر به یافتن عوامل مؤثر بر طول عمر می‌پردازیم. یکی از روش‌های بررسی رابطه‌ی