بسم الله الرحمن الرحيم

دانشگاه اصفهان

دانشکده علوم

گروه آمار

پایان نامه‌ی دکتری رشته‌ی آمار

مطالعه‌ای در مقایسه‌ی توزیع های طول‌عمر باقیمانده

استاد راهنما:

دکتر مجید اسدی

پژوهشگر:

ولی زاردشت

شهریور ماه ۱۳۸۹

دانشگاه اصفهان

دانشکده علوم

گروه آمار

پایان‌نامه‌ی دکتری رشته‌ی آمار

آقای ولی زاردشت

تحت عنوان

مطالعه ای در مقایسه توزیع های طول عمر باقیمانده

در تاریخ **۸۹/۶/۱۵** توسط هیأت داوران زیر بررسی با درجه **بسیار خوب** به تصویب نهایی رسید.

۱- استاد راهنمای پایان‌نامه دکتر مجید اسدی با مرتبه‌ی علمی استاد امضاء

۲ – استاد داور داخل گروه دکتر محمد حسین علامت ساز با مرتبه‌ی علمی استاد امضاء

۳- استاد داور داخل گروه دکتر نصراله ایران پناه با مرتبه‌ی علمی استادیار امضاء

۴- استاد داور خارج از گروه دکتر احمد پارسیان با مرتبه ی علمی استاد امضاء

۵- استاد داور خارج از گروه دکتر محمد خنجری صادق با مرتبه ی علمی استادیار امضاء

امضای مدیر گروه

تقديم بە

خانوادەی عزیزم

فهرست مطالب

فهرست شکل‌ها

فهرست جدول‌ها

مخفف‌ها

نرخ شکست نزولی	DFR
نرخ شکست صعودی	IFR
متوسط نرخ شکست نزولی	DFRA
متوسط نرخ شکست صعودی	IFRA
مستقل و هم‌توزیع	i.i.d.
میانگین طول عمرباقیمانده	MRL
نو بهتر از کهنه	NBU
نو بدتر از کهنه	NWU
میانگین مربعات خطا	MSE
میانگین انتگرال مربعات خطا	MISE
اعتبارسنجی متقابل حداقل مربعات	LSCV

علامت‌ها و نمادها

تابع نرخ شکست متغیر X	$h_X(t)$
تابع نرخ شکست تجمعی توزیع F	$\Lambda_F(t)$
تابع نرخ شکست معکوس متغیر X	$r_X(t)$
نسبت نرخ‌های شکست X و Y	$\rho(t) = \frac{h_Y(t)}{h_X(t)}$
نسبت نرخ‌های شکست معکوس X و Y	$\varrho(t) = \frac{r_X(t)}{r_Y(t)}$
متغیر طول عمر باقیمانده	X_t
میانگین طول عمر باقیمانده متغیر X	$m_X(t)$
متغیر مدت از کار افتادگی	X^t
میانگین مدت از کار افتادگی متغیر X	$m^*_X(t)$
آماره‌ی ترتیبی kام از نمونه‌ی تصادفی n تایی	$X_{k:n}$
همگرائی قریب به یقین	$\xrightarrow{a.s}$
همگرائی در احتمال	$\xrightarrow{p}$
همگرائی در توزیع	$\xrightarrow{d}$
تابع هسته	$k(x)$
برآوردگر تجربی $R(t)$	$\hat{R}(t)$
برآوردگر هموار $R(t)$	$\tilde{R}(t)$
$P(X > Y)$	R
$P(X_t > Y_t)$	$R_{XY}(t) = R(t)$
$P(X^t < Y^t)$	$L(t)$
$P(\min\{X,Y\} = X \mid \min\{X,Y\} \leq t)$	$M(t)$
$P(X_{k:n} = X_j \mid X_{n:n} \leq t)$	$L^j_{k,n}(t)$

پیشگفتار

در مباحث علم آمار، مقایسه‌ی بین کمیت‌ها و جامعه‌های مختلف همواره از جایگاه مهمی برخوردار بوده است. در حقیقت می‌توان گفت معرفی روش‌های مقایسه، بخش قابل توجهی از مباحث آماری را به خود اختصاص داده است. این موضوع، در تحلیل بقا و به ویژه در نظریه‌ی قابلیت اعتماد از اهمیت بیشتری برخوردار بوده و روش‌های مقایسه‌ی متنوعی، مورد مطالعه پژوهشگران قرار گرفته است.

ترتیب‌بندی تصادفی نوع خاصی از مقایسه است که در این میان نقش گسترده‌تر و پرکاربردتری ایفا می‌کند. ترتیب تصادفی معمولی، ترتیب در نرخ شکست و ترتیب در میانگین طول عمر باقیمانده از جمله‌ی این نوع مقایسه‌ها هستند. برای مثال، فرض کنید X و Y به ترتیب دارای توابع بقای $\bar{F}(x)$ و $\bar{G}(y)$ باشند. گوییم X در ترتیب تصادفی معمولی از Y کوچک‌تر است اگر برای هر x، $\bar{F}(x) \leq \bar{G}(x)$. در علوم مختلف، مصادیق و کاربردهای عملی بسیاری برای ترتیب‌های تصادفی وجود دارد و در این زمینه پژوهش‌های متعدد و متنوعی صورت گرفته است. برای نمونه، در فرایند تولید یک محصول، می‌توان گفت بین طول عمر محصول وقتی فرایند تحت کنترل و وقتی خارج از کنترل است؛ یک ترتیب تصادفی معمولی برقرار است.

شکل دیگر مقایسه‌ی بین دو متغیر که در دهه‌های اخیر بسیار مورد توجه پژوهشگران قرار گرفته و به عنوان یک مدل احتمالی در مسائل مختلف استفاده شده است؛ در نظر گرفتن مقدار $R = P(X > Y)$ است. این احتمال بیشتر با نام مدل تنش ـ مقاومت مشهور است. اگر متغیر X را مقاومت (مدت زمان دوام و طول عمر) دستگاهی در نظر بگیریم که در محیط مورد استفاده، تحت تنش تصادفی Y قرار دارد؛ در این صورت، مقدار R در حقیقت میزان قابلیت اعتماد دستگاه مورد نظر را نشان خواهد داد. استفاده از احتمال R در مقایسه‌ی طول عمر دو دستگاهی که از دو روش مختلف تولید شده‌اند؛ و نیز آماره آزمون هم‌توزیعی در آزمون ویلکاکسون ـ من ـ ویتنی (با توجه به این که مقدار این احتمال وقتی X و Y مستقل و هم‌توزیعند؛ برابر ۰/۵ است) از کاربردهای مهم این معیار مقایسه است.

اساس معیارها و روش‌های قابلیت اعتماد برای مقایسه‌ی بین متغیرهای طول عمر، در نظر گرفتن خواص احتمالی متغیرها در طول زمان و سنجش تاثیر گذر زمان بر عمر واحد مورد مطالعه است. بر مبنای این اصل و با توجه به اهمیت کاربردی مدل احتمالی فوق، به کارگیری این مدل به شکلی که گذر زمان را

نیز در مقایسه‌ی بین دو متغیر لحاظ کند؛ و در واقع ارائه‌ی مدلی شبیه مدل تنش ـ مقاومت پویا و وابسته به زمان، هدف این پایان‌نامه است. متغیرهای تصادفی مستقل X و Y را طول عمر دو دستگاهی در نظر بگیرید که در لحظه‌ی t در حال کار هستند. در این صورت، طول عمر باقیمانده‌ی دو دستگاه به ترتیب برابر $X_t = X - t|X > t$ و $Y_t = Y - t|Y > t$ خواهد بود. در این پایان‌نامه، تابع $R(t) = P(X_t > Y_t)$ را به عنوان معیار مقایسه‌ی وابسته به زمان دو دستگاه معرفی می‌شود.

در فصل اول، مفاهیم، تعاریف و قضایایی که در طول پایان‌نامه به آن‌ها ارجاع داده می‌شود: یادآوری می‌شوند. مطالعه‌ی خواص $R(t)$ و ذکر برخی کاربردهای آن در فصل دوم انجام می‌شود. در این فصل، رابطه‌ی $R(t)$ را بر حسب توابع توزیع X و Y بیان می‌شود. نشان داده می‌شود که اگر نسبت نرخ شکست X و Y، تابع یکنوا از زمان باشد؛ آنگاه $R(t)$ نیز تابع یکنوا خواهد بود. به ویژه‌این که اثبات می‌شود که ثابت بودن $R(t)$، مدل نرخ‌های شکست متناسب کاکس را مشخص‌سازی می‌کند. مثال‌هایی برای توضیح نموداری این نتیجه آورده و به اهمیت آن اشاره می‌شود. ترتیب احتمال یکی از ترتیب‌های تصادفی است که بر مبنای مدل تنش ـ مقاومت تعریف شده و داری کاربردهای جالب و مفیدی است. با توسعه مفهوم این ترتیب، مفهوم ترتیب احتمال طول عمر باقیمانده را تعریف و نشان داده می‌شود که این مفهوم یک ترتیب تصادفی بین X و Y اعمال می‌کند. بسته بودن مفهوم بیان شده تحت عملگر آمیختگی توزیع‌های احتمال و مینیمم تعداد تصادفی از متغیرها مطالعه شده و ارتباط این ترتیب با برخی ترتیب‌های تصادفی مهم بررسی شده است.

با توجه به کاربردهایی که برای معیار $R(t)$ بیان شد؛ و به طور کلی با توجه به پتانسیل کاربردی بالای آن، ارائه برآوردگری برای این معیار، در شرایطی که هیچ نوع اطلاعاتی در باره ساختار توزیعی X و Y در دست نیست؛ امری لازم و ضروریست. فصل سوم به برآورد ناپارامتری این معیار می‌پردازد. برآوردگر ناپارامتری تجربی و برآوردگر هموار، دو برآوردگر ارائه شده در این فصل هستند. خواص حدی این برآوردگرها، از جمله توزیع حدی آن‌ها در این فصل بررسی می‌شود. در این فصل روش شبیه‌سازی مونت کارلو، برای ارزیابی دقت برآوردگرهای ارائه شده استفاده می‌شود.

در فصل چهار با استفاده از خاصیت مشخص‌سازی $R(t)$، آماره آزمونی برای آزمون فرض متناسب بودن نرخ‌های شکست ارائه می‌گردد. به کمک شبیه‌سازی توان این آزمون بررسی و با برخی از آزمون‌های

موجود مقایسه می‌شود. این آماره برای آزمون فرض متناسب بودن نسبت نرخ‌های شکست یک مجموعه داده‌ی واقعی به کار برده می‌شود.

فصل پنجم مساله‌ی دوگان مقایسه‌ی طول عمر باقیمانده، یعنی مقایسه‌ی مدت از کار افتادگی سیستم‌ها، $X^t = t - X|X \leq t$، را بررسی می‌کند. معیار متناظر که با $L(t)$ نشان داده می‌شود؛ می‌تواند در شناسایی مولفه‌ی از کار افتاده در یک سیستم پیچیده و نیز برای مشخص کردن لحظه‌ی شکست مولفه‌های این گونه سیستم‌ها مفید واقع شود. خواص و نتایجی مشابه قبل برای این معیار نیز مطالعه شده، به کاربردهای ویژه آن در مدل‌های شکست عام و مساله‌ی بازرسی پرداخته شده است.

جمع‌بندی مطالب بحث شده در طول پایان‌نامه، نتیجه‌گیری شبیه‌سازی‌های انجام شده و موضوعات پیشنهادی برای آینده‌ی تحقیق در فصل ششم آورده شده است.

فصل ۱

تعاریف و مفاهیم مقدماتی

۱ـ۱ مقدمه

رویکرد آماری بررسی یک موضوع مورد مطالعه را می‌توان ارائه الگو، مدل‌بندی مناسب مساله و برآورد پارامترها و مشخصه‌های الگو براساس مشاهدات و داده‌های موجود دانست. برای مثال، برازش توزیع‌های آماری، مدل‌های رگرسیونی و مدل‌های ریاضی مختلف برای نرخ مرگ و میر در تبیین و الگوبندی مسائل گوناگون به کار گرفته شده اند. بدیهی است که تنوع و طبیعت مسائل و ویژگی داده‌های موجود، شکل و گوناگونی خاصی به این مدل‌ها بخشیده است. این مدل‌ها در نظریه‌ی قابلیت اعتماد[۱] و تحلیل بقا[۲] با توجه به کاربرد گسترده و نوع داده‌های آن‌ها (سانسور[۳] و برش[۴])، تنوع و گاه شکل متفاوت‌تری پیدا کرده‌اند. مدل تنش ـ مقاومت[۵] از جمله مدل‌هایی است که در سه دهه‌ی اخیر مورد توجه پژوهشگران بوده و برای مدل‌بندی قابلیت اعتماد در مسائل مختلف استفاده شده است. اهمیت و کاربردهای گسترده‌ی این مدل انگیزه اصلی مطالعه صورت گرفته در این پایان نامه بوده است. در این فصل مدل تنش ـ مقاومت معرفی و به کاربردهای مهم آن اشاره می‌شود. مفاهیم و تعاریفی از نظریه‌ی قابلیت اعتماد و ترتیب‌های تصادفی[۶] که در این رساله مطاله شده‌اند؛ یادآوری می‌شوند. بخشی از این رساله به استنباط آماری و برآورد ناپارامتری هموار مدل ارائه شده می پردازد. معرفی ایده‌ها و مفاهیم اصلی این روش ناپارامتری برآورد توابع، بخش دیگر این فصل

Reliability[۱]
Survival Analysis[۲]
cencoring[۳]
Truncation[۴]
Stress - Strength[۵]
Stochastic Orders[۶]

را تشکیل می‌دهد. دربحث برآورد ناپارامتری، از نظریه‌ی آماره‌ی U[۷] استفاده شده است. معرفی مختصر این نظریه نیز در این فصل انجام می‌شود. همچنین، قضایای حدی استفاده شده در رساله را یادآوری می‌شود. پایان این فصل به تعاریف سیستم منسجم و تابع قابلیت اعتماد سیستم‌ها اختصاص داده شده است.

۱ــ۲ مدل تنش ـ مقاومت

انگیزه‌ی اصلی مطالعه و توسعه‌ی این مدل را می‌توان به مساله‌ی آمار کاربردی که در آن موضوع مقاومت یک واحد و تنش وارده بر آن مطرح است؛ نسبت داد. به عبارت ساده در این مساله، هدف، ارزیابی قابلیت اعتماد یک مولفه، براساس متغیرهای تصادفی Y به عنوان تنش وارد شده به مولفه و X به عنوان مقاومت مولفه برای غلبه بر تنش است. بر این اساس، اگر تنش بیش از مقاومت باشد $(Y > X)$؛ مولفه از کار خواهد افتاد. بر عکس، اگر مقاومت بیشتر از تنش باشد مولفه همچنان کار خواهد کرد. بنابراین، در این مساله قابلیت اعتماد، در حقیقت، به صورت احتمال عدم شکست (یا خرابی) تعریف می شود: $P(X > Y)$. اصل این ایده توسط بیرنام[۸](۱۹۵۶) ارائه و توسط بیرنام و مک‌کارتی[۹](۱۹۵۸) توسعه داده شد.

مساله دیگر از موقعیتی ناشی می‌شود که در آن X و Y نشان دهنده‌ی طول عمر دو وسیله بوده و هدف برآورد احتمال از کار افتادن یکی قبل از دیگری است. در بحث قابلیت اعتماد سیستم‌ها، هال[۱۰](۱۹۸۴) مثالی را ارائه کرده است که در آن ولتاژ تخلیه یک خازن، X، بایستی بیش از ولتاژ خروجی یک انتقال‌دهنده، Y، باشد تا یک مولفه به درستی کار کند. لازم به یادآوری است که احتمال $P(X > Y)$ در آمار ناپارامتری و در ارتباط با روش ویلکاکسون[۱۱](۱۹۴۵)، من[۱۲] ـ ویتنی[۱۳](۱۹۴۷) نیز همواره مطرح بوده است.

اگر چه با گذشت زمان تعمیم و کاربردهای بیشتری از مدل تنش ـ مقاومت مطالعه شده است؛ اما حجم زیادی از مقالات، به مساله‌ی احتمالی محض ارزیابی و برآورد کارا و پایای پارامتر $R = P(X > Y)$ براساس مقادیر نمونه و با فرض توزیع‌های مختلف برای X و Y (مانند نرمال، نمایی، لگ‌نرمال، وایبل و ...) اختصاص داده شده است. در حالی که بخش کمی از پژوهش‌ها مساله‌ی تغییر X و Y با گذشت زمان یعنی

---

[۷] U-Statistic
[۸] Birnbaum
[۹] McCarty
[۱۰] Hall
[۱۱] Wilcoxon
[۱۲] Mann
[۱۳] Whitney

مدل تنش ـ مقاومت وابسته به زمان را در نظر گرفته اند. باسو[۱۴] و ابراهیمی (۱۹۸۳) و بیلیکام [15](۱۹۸۵) مدل تنش ـ مقاومتی را در نظر می‌گیرند که از طریق پارامترهای $\theta_1(t)$ و $\theta_2(t)$ به زمان وابسته اند. ابراهیمی و راملینگام[۱۶](۱۹۹۳) حالتی را در نظر می‌گیرند که $X = X(t)$ و $Y = Y(t)$ دو حرکت براونی مستقل هستند. امین‌زاده(۱۹۹۹) مدل تنش ـ مقاومت با ساختار سریهای زمانی را مورد بررسی قرار داده است. برای جزئیات بیشتر و مرور نتایج و تعمیم‌های مدل تنش ـ مقاومت خواننده را به مرجع کاتز[۱۷] و همکاران (۲۰۰۳) ارجاع می‌دهیم.

۱ــ۳ مفاهیم پایه‌ای در قابلیت اعتماد

شاخه‌ی مهمی از علم آمار به مطالعه و تحلیل متغیرها و داده‌های طول عمر[۱۸] می‌پردازد. این شاخه، در بحث طول عمر واحدهای زنده و بقای انسان با نام تحلیل بقا و در مطالعه و مهندسی طول عمر سیستم‌ها و واحدهای صنعتی، با نام قابلیت اعتماد مشهور است. این شاخه از علم آمار، هم به سبب نوع داده‌ها و مشاهدات (ناکامل و سانسور شده) و هم به دلیل نوع متغیرهای (طول عمر و زمان) مورد مطالعه، مفاهیم و روش‌های نسبتاً متفاوت و مخصوص به خود پیدا کرده است. به طور مثال، برای بیان ریاضی رفتار احتمالی یک متغیر تصادفی، عموماً توابع چگالی و توزیع به کار گرفته می‌شوند؛ در حالی که در مطالعه‌ی متغیرهای طول‌عمر به دلیل اهمیت اندازه‌گیری تاثیر گذر زمان، توابع دیگری مانند تابع نرخ شکست و تابع میانگین طول عمرباقیمانده نقش محوری ایفا می‌کنند. معیارهای قابلیت اعتماد در عین حال که توزیع احتمالی متغیرها را مشخص می‌کنند؛ برای دسته‌بندی توزیع‌های طول عمر و برازش مدل‌های مختلف (مانند مدل نرخ‌های شکست متناسب) نیز به کار می‌روند. در این بخش، برخی از این مفاهیم که در رساله استفاده شده‌اند؛ معرفی می‌گردد. برای جزئیات بیشتر، بارلو[۱۹] و پروشان[۲۰](۱۹۷۵) و اسمیت[۲۱](۲۰۰۲) را ملاحظه کنید.

[۱۴] Basu
[۱۵] Bilikam
[۱۶] Ramallingam
[۱۷] Kotz
[۱۸] Lifetime Data
[۱۹] Barlow
[۲۰] Proschan
[۲۱] Smith

۱—۳—۱ توابع و معیارهای قابلیت اعتماد

مقادیری که یک متغیر تصادفی طول عمر اختیار می‌کند؛ در فاصله $[0,\infty)$ قرار می‌گیرد. فرض کنید X یک متغیر طول عمر با تابع توزیع مطلقاً پیوسته‌ی F و تابع چگالی احتمال f باشد. آنگاه تابع بقا یا قابلیت اعتماد X به صورت

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

تعریف می‌شود. واضح است که $\bar{F}(0) = 1$ و $\bar{F}(\infty) = 0$.

تابع بقا شانس خرابی یا از کار افتادگی را در زمانی بعد از زمان داده شده بیان می‌کند. برای بررسی سالخوردگی در هر لحظه از زمان، تابع نرخ خطر[۲۲] استفاده می‌شود. این معیار نقش اساسی در مطالعات قابلیت اعتماد و تحلیل بقا ایفا می‌کند.

تعریف ۱.۱ تابع نرخ خطر، $h(t)$، برای هر t که $\bar{F}(t) > 0$؛ به صورت

$$h(t) = \lim_{\delta \to 0} \frac{P(t < X \leq t + \delta | X > t)}{\delta} = \frac{f(t)}{\bar{F}(t)}$$

تعریف می‌شود.

به تابع $h(t)$، نرخ شکست[۲۳]، نیروی مرگ[۲۴] و نرخ شدت[۲۵] نیز می‌گویند. در ادامه و در طول این پایان‌نامه نام نرخ شکست را برای تابع $h(t)$ به کار خواهیم برد.

تعریف ۲.۱ انتگرال تابع نرخ شکست روی بازه‌ی $(0,t]$ یعنی

$$\Lambda(t) = \int_0^t h(x)dx,$$

را تابع نرخ شکست تجمعی[۲۶] می‌نامند.

معیار دیگری که در مطالعات طول عمر به کار می‌رود، تابع نرخ شکست معکوس[۲۷]، $r(t)$، است. این تابع معمولاً در مطالعه‌ی زمان سپری شده از شکست یک مولفه و در آزمون‌های طول عمر ظاهر می‌شود.

[۲۲] Hazard rate function
[۲۳] Failure rate
[۲۴] Force of mortality
[۲۵] Intensity rate
[۲۶] Cumulative failure rate function
[۲۷] Reversed failure rate function