
دانشگاه رازی

دانشکده علوم

گروه آمار

پایان نامه کارشناسی ارشد آمار ریاضی

تحلیل بقا بر اساس مدل های شکنندگی

توسط

سمیه فخری

استاد راهنما

دکترسید رضا هاشمی

دکتر داود قزوینی نژاد

استاد مشاور

تیر ماه ۱۳۹۰

کلیه حقوق اعم از چاپ و تکثیر، نسخه برداری، ترجمه، اقتباس و ... از این پایان نامه برای دانشگاه رازی محفوظ است. نقل مطالب با ذکر مأخذ بلامانع است.

تقدیم به:

تقدیم به خانواده عزیزم بخصوص به محضر پاک پدر و مادرم که نخستین آموزگارام بودند

تقدیم به همسرم به خاطر همه فداکاری ها و کمک هایش

تقدیم به فرزند عزیزم آیناز کوچولو که بزرگترین عطیه الهی به من است.

قدردانی

سپاس خدایی را که انسان را قدرت اندیشه آفرید.
بی تردید اثر حاضر مدیون راهنمایی و مشاوره اساتید گرانقدر و مساعدت و همکاری انسان های بزرگوار
بسیاری است.

اکنون در این مرحله از زندگی علمی خود، که خود را وام دار تمامی آنان می بینم، بر خود فرض
می دانم که از زحمات ارزنده آنان صمیمانه قدردانی نمایم.
در ابتدا لازم می دانم از زحمات بی شائبه و راهنمایی های ارزنده اساتید محترم جناب آقای دکتر
هاشمی و جناب آقای دکتر قزوینی نژاد سپاس گذاری کنم، که شخصیت علمی و اخلاقی ایشان همواره
برایم الگو بوده است و بی شک بدون همدلی و مساعدت ایشان پژوهش حاضر به انجام نمی رسید.
همچنین از اساتید محترم جناب آقای دکتر مجید صادقی فر و جناب آقای دکتر مهرداد نیاپرست
سپاسگزارم که بزرگوارانه زحمت داوری این پایان نامه را متقبل شده و با دقت عالمانه خود نکات
ظرفی را در زیبا شدن این پایان نامه بیان داشتند.
از خانواده و همسر نیز صمیمانه تشکر و سپاس گزاری می نمایم که سرایت مشکلات تحصیلی اینجانب
در دوران تحصیل در دانشگاه را در زندگی عادی و خانوادگی تحمل نمودند، امیدوارم با چشم پوشی از
کوتاهی های اینجانب در انجام وظایف خانوادگی ام بر من منت گزارند.
در پایان از همه دوستانی که به هر نحو در انجام این کار مرا یاری رساندند، تشکر می کنم.

سمیه فخری

کرمانشاه، تیر ماه ۹۰

پیشگفتار

در آمار استنباطی اغلب اوقات با متغیرهایی روبرو می‌شویم که مستقل فرض می‌شوند و با این فرض استنباط‌هایی در مورد جامعه انجام می‌گیرد. اگر این فرض زیربنایی درست نباشد، آنگاه تمامی استنباط دچار اشکال می‌شود. از این رو برای بررسی درستی یا نادرستی این فرض زیربنایی، انجام آزمون استقلال اهمیت اساسی دارد. آزمون استقلال براساس ضریب همبستگی پیرسن، یک آزمون پارامتری است که در عمل کاربرد قابل ملاحظه‌ای دارد. اما اگر آزمون باعث قبول (رد نشدن) فرض صفر (استقلال) شود، در این صورت نمی‌توان نتیجه گرفت که متغیرهای تصادفی مستقل‌اند، بلکه نتیجه می‌گیریم که متغیرهای تصادفی به طور خطی مستقل‌اند. دو روش معروف دیگر برای انجام آزمون استقلال، آزمون با ضریب همبستگی اسپیرمن و آزمون با ضریب همبستگی کندال است، که دو آزمون ناپارامتری برای استقلال هستند. این دو آزمون بعضی اوقات توان‌تر از آزمون پارامتری با ضریب همبستگی پیرسن هستند. نظر به این توانمندی است که در این رساله هدف خود را روی آزمون‌های ناپارامتری متمرکز کرده و در بین این آزمون‌های ناپارامتری توان‌ترین آزمون موضعی، که یک آزمون رتبه‌ای است را بدست می‌آوریم. روش پیدا کردن آزمون رتبه‌ای موضعاً توان‌ترین، با استفاده از یک قضیه از شیراهاتا^۱ داده شده و چندین مثال در مورد آن حل می‌گردد. ملاحظه خواهیم کرد که برای پیدا کردن آزمون رتبه‌ای موضعاً توان‌ترین و حل مثال‌های مربوطه چگونه از کپولاها استفاده می‌کنیم. سپس این آزمون را با آزمون‌های دیگر موجود مقایسه می‌کنیم، به خصوص مقایسه‌ای براساس کارایی نسبی مجانبی (*ARE*) آزمون رتبه‌ای موضعاً توان‌ترین و آزمون با ضریب همبستگی پیرسن انجام می‌شود. *ARE* برای دو آزمون برابر نسبت حجم نمونه‌ها تا رسیدن به یک توان خاص می‌باشد. مثلاً اگر آزمون ۱ با نمونه‌ای به حجم ۲۰ به توان مطلوب ۰/۹۵ و آزمون ۲ با نمونه‌ای به حجم ۱۰ به توان ۰/۹۵ می‌رسد، در این صورت *ARE* آزمون ۱ نسبت به آزمون ۲ به صورت $0.5 = \frac{1}{2}$ می‌باشد.

از آنجائیکه برای یافتن آزمون رتبه‌ای موضعاً توان‌ترین از کپولاها استفاده می‌کنیم، از این رو فصل اول به معرفی کپولاها می‌پردازد. همچنین چند خانواده معروف از کپولاها را که با پارامتر نامعلوم θ

^۱ Shirahata

پارامتریده شده‌اند معرفی می‌کنیم. در فصل دوم معیارهای هماهنگی به خصوص ضریب همبستگی پیرسن، ضریب همبستگی اسپیرمن و ضریب همبستگی کندال را شرح می‌دهیم. فصل سوم به مدل‌های پایه برای توان آزمون استقلال اختصاص داده شده‌است که در آنجا به معرفی کیپولاهای تجربی پرداخته و توان سه آزمون استقلال برای مدل‌های مختلف از کیپولاها با هم مقایسه شده‌اند. در فصل چهارم، آزمون رتبه‌ای موضعاً تواناترین را یافته و در پایان مقایسه‌ای بین این آزمون و آزمون‌های دیگر خصوصاً آزمون با ضریب همبستگی پیرسن ارائه می‌شود.

تحلیل بقا بر اساس مدل های شکنندگی

چکیده

تحلیل بقا یکی از مهم ترین مباحث آماری است که برای تحلیل داده های طول عمر در زمینه های مختلف به کار برده می شود. در سال های اخیر برای توسعه مدل های بقای کلاسیک به مدل هایی که قابلیت انعطاف بیشتری برای برآزنده شدن به داده های بقا را داشته باشند، مدل های شکنندگی معرفی شده اند. پارامتر شکنندگی می تواند برای هر فرد و یا برای هر خوشه تغییر کند که در حالت اخیر آن را منبع شکنندگی مشترک شده می نامند. اخیراً برای تحلیل داده های بقای خوشه بندی شده مقالات بسیاری به چاپ رسیده است.

تعداد زیادی از مدل های شکنندگی و چندین تکنیک برای برآزش بر این مدل ها در این پایان نامه مورد بحث قرار گرفته است. در این پایان نامه مدل های شکنندگی را به صورت مدل های پارامتری در نظر گرفته ایم. هدف اصلی در این پایان نامه مطالعه روی مدل های شکنندگی و بررسی جزئیات آن ها است.

مدل خطرهای متناسب و مدل زمان های شکست شتابیده از جمله مدل هایی است که در این پایان نامه معرفی شده اند. در این پایان نامه پارامتر مدل های مختلف را به کمک روش ماکزیمم کردن لگاریتم درستنمایی و یا با استفاده از نظریه بیز برآورد می نماییم. بعلاوه توابع چگالی مختلف برای شکنندگی نظیر گاما، پایای مثبت و و نیز واریانس توانی مورد بحث قرار گرفته است و هر کدام از این چگالی ها ساختار همبستگی مختلفی را روی داده ها ایجاد می کنند.

واژه های کلیدی : شکنندگی ، خطرهای متناسب ، ماکزیمم درستنمایی ، زمان های بقا ، کپولا ، تبدیل لاپلاس ، مدل وایبول ، گاما ، پایای مثبت ، تابع واریانس توانی

فهرست مندرجات

۱	تعاريف و مفاهيم اوليه	۱-۱
۲	مقدمه	۱-۱
۳	نمونه هاي بريده	۲-۱
۳	نمونه هاي چپ بريده	۱.۲-۱
۴	نمونه هاي راست بريده	۲.۲-۱
۴	انواع سانسور	۳-۱
۴	سانسور نوع اول	۱.۳-۱
۴	سانسور نوع دوم	۲.۳-۱
۵	سانسور چپ	۳.۳-۱
۵	سانسور راست	۴.۳-۱
۶	تحليل بقا	۴-۱
۶	تابع درستنمايي بقا	۱.۴-۱
۸	مدل خطر هاي متناسب	۲.۴-۱
۱۱	مدل زمان شكست شتابيده	۳.۴-۱
۱۴	مدل لگ خطي	۴.۴-۱
۱۷	مفهوم شكندگي	۵-۱

۲ بررسی مدل خطرهای متناسب تحت شکنندگی ۲۵

۲۶ مقدمه	۱-۲
۲۷ مدل خطرهای متناسب پارامتری با عامل شکنندگی	۲-۲
۲۸ ماکزیمم کردن درستنمایی حاشیه‌ای	۳-۲
۳۴ مدل با اثر ثابت	۴-۲
۳۴ مشخصات مدل	۱.۴-۲
۳۵ کارایی مجانبی برآورد پارامترها در مدل با اثرات ثابت	۲.۴-۲
۳۶ چگالی های پسین و نظریهٔ بیز	۵-۲
۳۶ الگوریتم متروپلیس برای مدل شکنندگی گاما پارامتری	۱.۵-۲

۳ ارتباط بین تابع مفصل و مدل شکنندگی ۴۴

۴۵ مقدمه	۱-۳
۴۵ مدل تابع مفصل	۲-۳
۴۵ نمادگذاری و تعاریف	۱.۲-۳
۴۷ تعریف مدل تابع مفصل	۲.۲-۳
۴۹ تابع مفصل کلایتون	۳.۲-۳
۵۲ تابع مفصل کلایتون در برابر مدل شکنندگی گاما	۴.۲-۳
۵۴ مدل حاشیه‌ای	۳-۳
۵۴ تعریف مدل حاشیه‌ای	۱.۳-۳
۵۴ سازگاری برآوردهای پارامتر در مدل حاشیه‌ای	۲.۳-۳

۵۶	۳-۳-۳ واریانس η
۵۸		۴ بررسی خواص توزیع های شکنندگی مختلف
۵۹	۱-۴ مقدمه
۶۰	۲-۴ مشخصات کلی توزیع های شکنندگی
۶۱	۱.۲-۴ تابع بقای توأم و تبدیل لاپلاس
۶۲	۲.۲-۴ تابع بقای جامعه و تابع مفصل
۶۴	۳.۲-۴ مقیاس های وابستگی
۶۶	۳-۴ توزیع گاما
۶۶	۱.۳-۴ تعاریف و ویژگی های اصلی
۶۷	۲.۳-۴ تابع بقای جامعه و تابع بقای توأم
۷۱	۳.۳-۴ نمایش فرم تابع مفصل
۷۱	۴.۳-۴ مقیاس وابستگی
۷۳	۴-۴ توزیع پایای مثبت
۷۳	۱.۴-۴ تعاریف و ویژگی های اصلی
۷۵	۲.۴-۴ تابع بقای جامعه و تابع بقای توأم
۷۷	۳.۴-۴ نمایش فرم تابع مفصل
۷۹	۴.۴-۴ مقیاس وابستگی
۸۰	۵-۴ توزیع واریانس توانی
۸۰	۱.۵-۴ تعاریف و ویژگی های اصلی
۸۳	۲.۵-۴ تابع بقای جامعه و تابع بقای توأم
۸۵	۳.۵-۴ نمایش فرم تابع مفصل
۸۵	۴.۵-۴ مقیاس وابستگی

نمادها و علائم اختصاری

$P(\cdot)$	اندازه احتمال
w_i	اثر تصادفی خوشه i ام
β	بردار اثرات ثابت (از بعد p)
\mathbf{S}	بردار امتیاز
ξ	بردار پارامترهای مخاطره پایه
\mathbf{x}	بردار شامل متغیرهای کمکی
$ $	به شرط آنکه
iid	به طور مستقل و مشابه توزیع شده
∞	بینهایت
F, f	تابع توزیع و تابع چگالی زمان بقا
G, g	تابع توزیع و تابع چگالی زمان سانسور
$C_\theta(\cdot)$	تابع کپولا با بردار پارامتر θ
$\Gamma(\cdot)$	تابع گاما
$h_o(t)$	تابع مخاطره پایه
$H_o(t)$	تابع مخاطره پایه تجمعی
$h_{ij}(t)$	تابع مخاطره شرطی
$h_{\mathbf{x},p}(\mathbf{t}_n)$	تابع مخاطره توأم برای خوشه های به اندازه n با متغیر کمکی \mathbf{x} بر اساس مدل حاشیه ای
$h_{\mathbf{x},f}(\mathbf{t}_n)$	تابع مخاطره توأم برای خوشه های به اندازه n با متغیر کمکی \mathbf{x} بر اساس مدل شکنندگی
$h_{X,p}(t)$	تابع مخاطره جامعه برای مؤلفه های با متغیر کمکی X بر اساس مدل حاشیه ای
$h_{X,f}(t)$	تابع مخاطره جامعه برای مؤلفه های با متغیر کمکی X بر اساس مدل شکنندگی
$h_{ij,c}(t)$	تابع مخاطره شرطی برای j امین مشاهده از خوشه i ام با شکنندگی u_i
$h_i(\mathbf{t}_{n_i})$	تابع مخاطره شرطی توأم برای خوشه i ام

$L(\cdot)$	تابع درست‌نمایی
$L_{marg}(\cdot)$	تابع درست‌نمایی حاشیه‌ای
$l(\cdot)$	تابع لگاریتم درست‌نمایی
$\mathcal{L}(\cdot)$	تبدیل لاپلاس
d_i	تعداد پیشامدها در خوشه i ام
s	تعداد خوشه‌ها
n_i	تعداد مشاهدات در خوشه i ام
τ	توکندال
$W(\lambda, \rho)$	چگالی وایبول با پارامتر مقیاس λ و پارامتر شکل ρ
t_{ij}	زمان پیشامد
c_{ij}	زمان سانسور
u_i	شکندگی خوشه i ام
T	متغیر زمان پیشامد
C	متغیر زمان سانسور
y_{ij}	مینیمم زمان پیشامد و زمان سانسور
δ_{ij}	نشانهگر سانسور اگر زمان پیشامد مشاهده شود $\delta_{ij} = 1$ و در غیر اینصورت صفر می باشد
θ	واریانس چگالی شکندگی با میانگین یک (به جز در حالتی که توزیع شکندگی استیبل مثبت باشد)
ζ	(ξ, θ, β)
\mathbf{x}	$(X_1^t, \dots, X_n^t)^t$
\mathbf{x}_i	$(X_{i,1}^t, \dots, X_{i,n_i}^t)^t$
z_{ij}	(y_{ij}, δ_{ij})

فصل ۱

تعاریف و مفاهیم اولیه

در این فصل ابتدا بعضی از ویژگی های مهم داده های بقا و مدل بقا را بدون در نظر گرفتن خوشه ها معرفی می کنیم. ویژگی های خاصی که تجزیه و تحلیل داده های بقا را از تجزیه و تحلیل آماری کلاسیک متمایز می کند داده های سانسور شده هستند.

به عنوان نمونه، زمان بقا برای بعضی از مؤلفه ها نامعلوم است و تنها اطلاع موجود آن است که مؤلفه تا یک زمان معلوم زنده مانده است. یکی از منابع مشکل ساز در تحلیل داده های بقا عدم امکان مشاهده افراد در تمام مدت تا وقوع شکست یا پیشامد مورد نظر است. در حالات بسیاری وقتی که طول عمر داده ها بررسی می شود، ممکن است تا زمان مشخص شده تا پایان آزمایش همه واحدها در نمونه خراب نشوند (یعنی پیشامد مورد نظر مشاهده نشود) یا زمان های دقیق خرابی همه واحدها در دسترس نباشد، این نوع داده ها را به طور معمول داده های سانسور شده می نامند. آنچه اصطلاحاً سانسور شدن تصادفی نامیده می شود، در کاربردهای پزشکی با مطالعه روی حیوان، در کاربردهای اپیدمیولوژی با مطالعات انسانی، در آزمایش های بالینی و یا در مطالعات قابلیت اعتماد پیش می آید. در این موارد آزمایش قبل از وقوع پیشامد خاتمه می یابد. مثلاً در یک آزمایش بالینی ممکن است بیماران در زمان های مختلف به مطالعه وارد شوند، می خواهیم طول مدت زندگی آن ها را از زمان ثبت نام مشاهده کنیم، اما ممکن است سانسور شدن به یکی از صورت های زیر رخ دهد:

- گم شدن در پیگیری (بیمار ممکن است تصمیم به مهاجرت بگیرد)،
- خارج شدن (به علت اثرهای نامطلوب روش درمانی مورد استفاده، قطع درمان ضروری می شود)،
- خاتمه مطالعه (برای تحلیل داده ها در زمان از قبل تعیین شده)،
- مرگ به علتی غیر از علت تحت بررسی (مثل خودکشی).

باید توجه کرد که داده های سانسور شده را باید با استفاده از آخرین اطلاع از واحدها دقیقاً ثبت کرد تا در تحلیل داده ها مورد استفاده قرار گیرد.

استنباط بر مبنای داده های سانسور شده در طول ۵۰ سال گذشته بوسیله نویسندگان متعددی برای طیف وسیعی از توزیع های طول عمر از قبیل نمایی، گاما، وایبول، لگ نرمال، گوسین معکوس، لجستیک، لاپلاس و پارتو مورد مطالعه قرار گرفته است. بسته به اینکه داده ها چگونه از آزمایش های طول عمر جمع آوری شوند، انواع مختلف سانسور به وجود می آید که در بخش های بعد توضیح داده خواهند شد.

۲-۱ نمونه های بریده

در آمار با استفاده از نمونه های به دست آمده از فضای نمونه ای به استنباط در مورد جامعه می پردازیم. حال وقتی که یک مشاهده یا منتخبی از مشاهدات روی بخشی از فضای نمونه ای محدود شوند، آنگاه نمونه های به دست آمده با توجه به ماهیت محدودیت به صورت بریده یا سانسور شده معرفی می شوند. به بیان دیگر اگر فقط به قسمتی از فضای نمونه ای دسترسی داشته باشیم آنگاه با نمونه های بریده یا سانسور شده سروکار خواهیم داشت، که در این حالت برآورد پارامترهای جامعه با حالت معمول متفاوت است و در واقع روش های معمول باید اصلاح شوند. اگر چه این نمونه ها اطلاعات کمتری دارند، اما همچنان می توانیم مانند داده هایی که شامل غیر سانسورها هستند برای تحلیل های آماری و نموداری استفاده کنیم.

نمونه های بریده آن هایی هستند که مقادیر مشخصی از جامعه کاملاً حذف شده است. شاید صحیح تر این است که بیان کنیم برش در جامعه اتفاق می افتد نه در نمونه. در حقیقت نمونه بریده، نمونه ای از جامعه بریده است. نمونه های بریده در آزمایش هایی ایجاد می شود که انتخاب نمونه فقط در قسمتی از محدوده متغیر امکان پذیر است. مثال هایی از این نمونه ها مکرراً در کارخانه ها اتفاق می افتد که از تولیدات غربال شده انتخاب می شوند.

نمونه های بریده برطبق اینکه نقطه برش معلوم است یا نه رده بندی می شوند. اگر این نقاط معلوم باشند فقط باید از نمونه به دست آمده، پارامترهای جامعه را برآورد کرد ولی اگر این نقاط نامعلوم باشند، این نقاط نیز پارامترهایی اضافی هستند که باید از روی نمونه برآورد شوند.

۱-۲-۱ نمونه های چپ بریده

در این نمونه ها فرض می کنیم n مشاهده وجود دارد، یک نمونه چپ بریده است اگر برای هر کدام از مشاهدات، $x \geq T$ باشد که T یک نقطه برش معلوم است. حال اگر چندین نقطه برش T_1, \dots, T_k وجود داشته باشد به طوری که $T_1 < \dots < T_k$ آنگاه یک نمونه چپ بریده فزاینده خواهیم داشت.

مثال ۱.۱ خانه سالمندان

در یک مطالعه بقا برای بررسی عملکرد کارکنان خانه سالمندان، سن افراد در زمان مرگ و سن آنها در زمان ورود به خانه سالمندان پالوآلتو^۱ بین سال های ۱۹۶۴ تا ۱۹۷۵ جمع آوری شده است.

^۱ Palo Alto

طول عمر در این مجموعه اطلاعات چپ بریده است، چون یک فرد باید تا زمان ورود به خانه سالمندان زنده بماند و افرادی که زودتر از سن بازنشستگی می میرند، شانس برای بودن در نمونه ندارند. این مثال یک نمونه چپ بریده است که داده های آن در کلاین و موش برگر^۲ (۱۹۹۷) آمده است.

۲.۲-۱ نمونه های راست بریده

در این نمونه ها فرض می کنیم n مشاهده وجود دارد، یک نمونه راست بریده است اگر برای هر کدام از n مشاهده $x \leq T$ باشد که T یک نقطه برش معلوم است. حال اگر چندین نقطه برش T_1, \dots, T_m وجود داشته باشد به طوری که $T_1 < \dots < T_m$ آنگاه یک نمونه راست بریده فزاینده خواهیم داشت.

مثال ۲.۱ وزن قطعات هواپیما

در یک شرکت تولید کننده قطعات بدنه هواپیما فقط قطعاتی به قسمت بسته بندی ارسال می شوند که وزن آن ها کمتر از ۱۵۰ پوند باشد حال اگر از قطعاتی که به قسمت بسته بندی ارسال شده نمونه ای انتخاب شود آنگاه این یک نمونه راست بریده با نقطه برش معلوم است.

۳-۱ انواع سانسور

۱.۳-۱ سانسور نوع اول

n واحد تحت مشاهده در یک آزمایش خاص را در نظر بگیرید. در طرح سانسور نوع اول، آزمایش تا زمان از پیش تعیین شده T ادامه پیدا می کند و در آن شکست بعد از زمان T در نظر گرفته نمی شود یعنی واحدهای سالم باقیمانده در زمان T را از آزمایش کنار گذاشته یا به عبارت دیگر سانسور می کنیم. در اینجا فرض می شود که زمان پایان مطالعه مستقل از زمان های شکست می باشد. پس در این طرح سانسور نقطه پایان آزمایش ثابت و تعداد شکست ها یک متغیر تصادفی است.

۲.۳-۱ سانسور نوع دوم

n واحد تحت مشاهده در یک آزمایش خاص را در نظر بگیرید. در طرح سانسور نوع دوم، آزمایش تا زمان رخداد m امین شکست ادامه می یابد، و پس از وقوع m امین شکست، واحدهای سالم باقی مانده

^۲ Klein and Moeschberger

از آزمایش کنار گذاشته یا به عبارت دیگر سانسور می شوند. در عمل، در این طرح تنها طول عمرهای با کوچکترین زمان مشخص می شوند. پس در طرح سانسور نوع دوم تعداد شکست ها ثابت و نقطه پایان آزمایش تصادفی است .

۳.۳-۱ سانسور چپ

در داده های سانسور شده چپ، یک زمان خرابی (پیشامد مورد نظر) پیش از زمان مشخص T_0 (شروع مشاهده در آزمایش) اتفاق افتاده است که زمان دقیق وقوع آن نامعلوم است. یعنی وقتی که اطلاعات زمان های قبل از شروع دوره مشاهدات در دسترس نباشد سانسور چپ را داریم. مسأله سانسور چپ در بسیاری از مسایل علمی، مثل مطالعات اپیدمیولوژی وقتی که آغاز پروسه به طور خاص مشخص نیست، اتفاق می افتد. برای مثال، در طول مطالعه یک بیماری (مثل سرطان) شروع ناخوشی اغلب مشخص نیست (دست کم امر مسلمی نیست). در نمونه سانسور شده چپ، طول عمر m واحد به طور کامل اندازه گیری شده در حالیکه برای $R = n - m$ مشاهده دیگر فقط می دانیم که $T < T_0$ (بیانگر متغیر زمان شکست است). یعنی $n - m$ عضو قبل از شروع آزمایش خراب شده اند. برای مشاهدات اندازه گیری شده $T \geq T_0$ است.

۴.۳-۱ سانسور راست

یکی از حالت های بسیار معمول سانسور، داده های سانسور شده از راست هستند. در بررسی طول عمر داده ها، این مجموعه از داده ها از واحدهایی که شکست نخورده اند (خراب نشده اند) تشکیل می شوند. در نمونه سانسور شده راست، طول عمر m واحد به طور کامل اندازه گیری شده در حالیکه برای $R = n - m$ مشاهده دیگر فقط می دانیم که $T > T_0$ است، یعنی $n - m$ عضو بعد از زمان T_0 (زمان پایان آزمایش) همچنان سالم هستند و کار می کنند در صورتیکه برای مشاهدات اندازه گیری شده $T \leq T_0$ است .

۴-۱ تحلیل بقا

در حالت کلی، فرض می‌کنیم که زمان سانسور و زمان بقا از نظر آماری متغیرهای تصادفی مستقل هستند. برای داده‌های سانسور شده راست، اطلاعات حقیقی برای مؤلفه i ام، به ازای $i = 1, \dots, n$ شامل زوج (y_i, δ_i) است به طوری که y_i مینیمم زمان پیشامد t_i و زمان سانسور c_i است یعنی $y_i = \min(t_i, c_i)$ و δ_i نشانگر سانسور است و اگر پیشامد مورد نظر اتفاق افتد مقدار یک را می‌گیرد و در غیر این صورت مقدار صفر را خواهد گرفت.

$$\delta_i = \begin{cases} 1 & t_i \leq c_i \\ 0 & t_i > c_i \end{cases}$$

روش‌های بحث شده در این جا فقط برای مدل زمان مرگ مفید نیستند، بلکه برای هر متغیری که زمان یک پیشامد خاص را توصیف می‌کنند نیز مفید است، مخصوصاً اگر زمان پیشامد برای همه مؤلفه‌ها مشاهده نشده باشد. واحد در نظر گرفته شده برای زمان پیشامد را به طور کلی مؤلفه می‌نامند، و واحدهای مورد نظر در خوشه‌ها گروه بندی می‌شوند.

۱-۴-۱ تابع درست‌نمایی بقا

برای داده‌های سانسور شده، درست‌نمایی بقا کاملاً از درست‌نمایی کلاسیک برای داده‌های مستقل بدون سانسور متفاوت است. در اینجا یک سری از نمادها که در نوشتن درست‌نمایی بقا به ما کمک میکند را معرفی می‌کنیم.

فرض کنید که f تابع چگالی احتمال زمان پیشامد T با تابع توزیع تجمعی زیر

$$F(t) = P(T \leq t) = \int_0^t f(u) du$$

و تابع بقای

$$S(t) = 1 - F(t) = P(T > t)$$

تابع مهم دیگر در تحلیل بقا نرخ مخاطره یا نرخ مرگ لحظه‌ای $h(t)$ است. نرخ مخاطره $h(t)$ متناسب با احتمال شرطی وقوع پیشامد در بازه $[t, t + \Delta t)$ به شرط اینکه پیشامد تا قبل از زمان t رخ نداده باشد است. نرخ آن با تقسیم این احتمال شرطی بر بازه زمانی Δt بدست می‌آید. نرخ مخاطره حد این نسبت برای وقتی است که Δt به سمت صفر میل کند

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (1.4.1)$$