دانشگاه پیام نور مرکز مشهد

پایان نامه کارشناسی ارشد آمار (گرایش ریاضی)

استنباط آماری مبتنی بر داده های سانسور هیبرید فزاینده نوع ۲

توسط :

مهدی حسین پوربوری آبادی

استـاد راهنـما :

دکتر باقرمقدس زاده بزاز

استـاد مشاور :

دکتر مجید رضائی

شهریورماه ۹۰

استنباط آماری مبتنی بر داده های سانسور هیبرید فزاینده نوع ۲

چکیده

تحلیل داده های مربوط به زمان بقا و خرابی در بسیاری از شاخه های آمار کاربردی (قابلیت اعتماد ، مطالعات پزشکی) مورد توجه است . ماهیت آزمایش های مربوط به این داده ها اغلب همراه با حذف واحدهایی از آزمایش است که اینگونه حذف ها " سانسور" نامیده می شوند . در واقع داده های سانسور شده ، مرتبط با آن دسته از واحدهای آزمایشی است که ممکن است در طول آزمایش به طور کامل شرکت نداشته باشند یا تا پایان مطالعه خراب نشوند . زیرا در بعضی از آزمایشات بررسی طول عمر، یا زمان مطالعه محدود است و نمی توان تا وقوع آخرین خرابی آزمایش را ادامه داد ، یا بعضی از واحد های آزمایشی ممکن است به هر دلیلی قبل از اتمام آزمایش ، از نمونه خارج شوند (انصراف دهند) . از این رو باید از نمونه گیری همراه با سانسور استفاده کرد . سانسور انواع مختلف دارد که در این پایان نامه از سانسور هیبرید فزاینده نوع ۲ استفاده شده است . مسئله ی مورد بررسی در این تحقیق این است که طرح سانسور باید به چه صورت باشد تا برآورد پارامتر نامعلوم بهینه باشند ، یا برآوردگر بیز و فواصل اطمینان مختلف برای آن چطور و چگونه بدست می آید .

در فصل ۱ مسائل مطرح شده در آنالیز بقا همچون تابع بقا ، تابع مخاطره ، نرخ شکست و مسائلی از این دست مورد بحث قرار گرفته است . در فصل دوم انواع سانسور ها معرفی شده و مثال های مرتبط با آن ارائه شده است در فصل سوم به بررسی سانسور هیبرید نوع ۱ و ۲ روی سه توزیع نمایی ، وایبل و لگ نرمال پرداختیم و متناسب با این توزیع ها برآوردگرهای درستنمایی ماکزیمم و برآوردگرهای مجانبی وتقریبی و بیز و همچنین فواصل اطمینان مرتبط با هر یک از این برآوردگرها ارائه شده است و در پایان برای هریک مثال عددی جهت روشن شدن موضوع ارائه گردیده است برای اینکه بینشی نسبت موضوع داشته باشیم بر روی یک سری داده آزمایشگاهی تحت سانسور هیبرید سه توزیع را پیاده کرده و مقایسه ای انجام گرفته است . در

فصل چهارم رویه بالا را تحت سانسور هیبرید فزاینده نوع ۲ روی توزیع نمایی انجام دادیم در فصل پنجم به بحث یکنوایی تصادفی برآورددرستنمایی ماکزیمم سانسور هیبرید تحت توزیع نمایی پرداختیم.

واژهای کلیدی : سانسور هیبرید ، برآوردگر درستنمایی ماکزیمم ، برآوردگر درستنمایی ماکزیمم تقریبی، برآورگر مجانبی، برآوردگر بیز، فاصله اطمینان

فهرست مطالب

لیست جداول

لیست اشکال

پیشگفتــار

در بازار رقابتی جهان امروز که محصولات بسیار متنوع و با کیفیت های بالا روانه بازارهای جهان شده اند ، تولید کنندگان تلاش دارند، بتوانند در جهت جلب هر بیشتر رضایت مشتریان خود کارکنند. یکی از سیاست هایی که کارخانه های تولید کننده محصولات برای جلب رضایت مشتریان خود اتخاذ می کنند ، فراهم کردن ضمانت برای طول عمر محصولات خود هستند . به طور قطع ضمانت طول عمر محصولات هزینه های اضافی را برای تولید کنندگان خواهد داشت بخش عمده ای از این هزینه ها مربوط به تعمیر و یا تعویض قطعاتی است که قبل از زمان اتمام دوره ضمانت خراب می شوند ، برای اینکه تولید کننده بتواند این هزینه ها را برآورد کند لازم است توزیع زمان شکست محصولات خود را بداند . بنابراین لازم است قبل از آن که محصولات وارد بازار شوند تحت انجام آزمایش های مربوط به طول عمر قرار بگیرند . البته اطلاعاتی که در طول انجام این آزمایش ها بدست می‌آیند در بسیاری از موارد دیگر جهت بهبود کیفیت و همچنین افزایش توان رقابتی محصولات کمک کند .

چگونگی آنالیز داده های طول عمر از دیر باز مورد توجه بوده است چرا که باعث کاهش هزینه های تولید ، رقابتی شدن تولیدات در بازار و ارائه تولیدات با طول عمر بیشتر می شود .

در بسیاری از مطالعات مربوط به داده های طول عمر با مواردی مواجه می شویم که واحدهای آزمایشی قبل از اتمام آزمایش از مطالعه خارج می شوند . این حذف شدن ممکن است به صورت غیر عمدی و یا قبل از توسط آزمایشگر طراحی شده باشد . به عنوان مثال اگر یکی از واحدهای آزمایشی به طور تصادفی خراب شود یا شخص تحت عنوان از ادامه همکاری کناره گیری کند حذف از نوع غیر عمدی رخ داده است و آزمایشگر باید برای آزمایش خود یک طرح سانسور مشخصی تعریف نماید . به داده های که قبل از مشاهده زمان شکست آن ها از آزمایش حذف می شوند داده‌های سانسور شده می گوییم .

فصل اول

۱.۱ مفاهیم وکلیات :

نظریه اعتماد در مهندسی یا تحلیل بقاء در علوم اجتماعی، شاخه ای از آمار است که به مرگ و میر موجودات زنده و شکست سیستم های مکانیکی می‌پردازد بطور کلی تحلیل بقاء شامل مدل سازی زمانی برای رخداد پیشامد هاست خیلی از موضوعات در تحلیل بقاء بوسیله نظریه شمارش توضیح داده می‌شود این نظریه اخیرا مطرح شده و قابلیت انعطاف پذیری فوق العاده آن به فرآیند های شمارشی بدین معناست که به ما امکان مدل سازی پیشامدهای چندگانه را هم می‌دهد به عنوان مثال افراد می‌توانند چندین دفعه به زندان بروند یا یک فرد معتاد چندین بار می‌تواند اعتیاد را خاتمه و مجددا آن را شروع کند یا مردم می‌توانند چندین مرتبه ازدواج و طلاق بگیرند تحلیل بقا تلاش می‌کند به سوالاتی چون :

چند درصد از جمعیت بعد از زمان t زنده اند ؟

آن افرادی از جامعه که زنده اند با چه نرخی می‌میرند ؟

آیا می‌توان تعداد مرگ و میر را محاسبه کرد ؟

برای پاسخ به این سوالات ضروری است تا ابتدا به معرفی طول عمر بپردازیم این پدیده (طول عمر)درمورد ارگانیسم‌های زیستی ناشناخته و مبهم نیست ولی در سیستم های مکانیکی تا حدی مبهم است و خیلی خوب تعریف نشده است. نظریه بقا که در این قسمت آورده شده است با این فرض همراه است که مرگ و میر یا شکست تنها یک با ر برای هر واحد آزمایشی (موجود زنده / سیستم مکانیکی) اتفاق می‌افتد و تکرار پیشامد که به انعطاف پذیری نظریه کمک می‌کند در بسیاری از موارد چون علوم اجتماعی و تحقیقات دارویی کاربرد دارد ، این فصل در وحله اول به فرموله کردن بقای موجود زنده می‌پردازد و برای پرداختن به سیستم های مکانیکی کافیست شکست های مکانیکی جایگزین مرگ و میر شود [۲۰] .

تابع چگالی و تابع توزیع طول عمر پیشامد : که به ترتیب به صورت زیر تعریف می‌شود .

$$f(t) = F'(t) = \frac{d}{dt}F(t)$$

$$F(t) = \Pr(T \le t)$$

تابع بقا [۱] :

آن را با $S(t)$ نشان می‌دهیم که T زمان و متغیر تصادفی است که آن را طول عمر می‌نامیم و $"Pr"$ احتمال متناظر آن احتمال بقا پس از یک زمان خاص است که آزمایشگر به آن علاقمند است .

دراین جا لازم ابتدا به مفهوم زمان تا ازکار افتادگی توجه کنیم بدین منظور فرض کنید یک مولفه غیر قابل تعمیر شدنی داریم منظور از زمان تا از کارافتادگی مدت زمانی است از شروع به کار مولفه تا اولین زمانی که از کار می‌افتد نقطه شروع را $t = 0$در نظر می‌گیریم

تابع بقا : تابع بقای یک مولفه به صورت زیر تعریف می‌شود

$$S(t) = \Pr(T > t) = \int_t^{\infty} f(u)\, du = 1 - F(t) \qquad (۱ . ۱)$$

لذا $S(t)$احتمال از کار نیفتادن سیستم در بازه $[\, 0 \,,\, t \,]$ است تابع بقا همچنین تابع زیستی نیز نامیده می‌شود که این تعریف در مسائل زیستی مطرح است از این تعریف در مبحث سیستم های مکانیکی تحت عنوان تابع قابلیت اعتماد $R(t)$ یاد می‌شود به طور معمول یکی از اولین فرضیه‌ها این است که $S(0) = 1$ است اگرچه می‌تواندکمتراز ۱ هم باشد که در این صورت فورا مرگ یا شکست نتیجه می‌شود تابع بقا یک تابع نا صعودی است $S(u) \le S(t)$ اگر $u > t$ این ویژگی از $1 - F(t) = S(t)$ نتیجه می‌گردد که از جمع بندی تابع نا منفی بدست می‌آید .

این تصور تعبیر بقا پس از یک زمان خاص است این ویژگی تابع توزیع طول عمر و چگالی را می‌دهد در این صورت F و f بخوبی تعریف می‌شوند. به طور مشابه تابع چگالی بقا یک پیشامد بدین صورت تعریف می‌شود

$$s(t) = S'(t) = \frac{d}{dt} S(t) = \frac{d}{dt} \int_t^{\infty} f(u)\, du = \frac{d}{dt} [1 - F(t)] = -f(t) \qquad (۲ . ۱)$$

هرچه زمان بیشتر می‌شود احتمال بقا به صفر نزدیک می‌شود اگرچه حد آن می‌تواند بیشتر از صفر باشد .

نرخ خطر : نرخ مرگ و میر یا نرخ شکست در واحدهای آزمایشی است به این صورت تعریف می‌شود احتمال اینکه مولفه در بازه زمانی $[\, t \,,\, t + \Delta t \,)$ از کار بیفتد به شرط اینکه بدانیم تا زمان t در حال کار کردن بوده است و برابر است با :

$$P\,(t \,,\, t + \Delta t \mid T > t) = (F(t + \Delta t) - F(t)) \frac{1}{S(t)}$$

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t \mid T \ge t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \cdot \frac{1}{S(t)}$$

[۱] Survival function

تابع مخاطره[۱] : عبارت است از نرخ خطر در زمان $T > t$ و برابر است با :

$$\lambda(t)dt = \Pr\left(t \le T \le t + dt \,\middle|\, T \ge t\right) = \frac{f(t)dt}{S(t)} = -\frac{S'(t)dt}{S(t)} \qquad (۳.۱)$$

در واقع تابع نرخ خطر (شکست) به معنای احتمال اینکه شخصی که سن t را دارد بایستی بمیرد قبل از اینکه به زمان $t + \Delta t$ برسد این تابع به طور خاص در آمارگیری نفوس در مطالعات آماری مورد استفاده قرا می‌گیرد و تابعی نا منفی است و انتگرال ان در بازه $[0,\infty]$ بایستی نا متناهی باشد اما در سایر جاها این قید روی آن نیست تابع مخاطره می‌تواند صعودی یا نزولی یا ثابت یا ناپیوسته باشد .

تابع مخاطره تجمعی[۲] : وقتی تابع مخاطره در حوزه تعریف اش جمع بندی کنیم تابع مخاطره تجمعی را خواهیم داشت که رابطه آن مثل رابطه تابع چگالی احتمال و تابع توزیع تجمعی احتمال است که آن را به صورت $\Lambda(t)$ نشان داده و تعریف می‌کنیم.

$$\Lambda(t) = -\log S(t) \qquad (۴.۱)$$

با تبدیل نمایی داریم :

$$S(t) = \exp(-\Lambda(t))$$

یا

$$\frac{d}{dt}\Lambda(t) = -\frac{S'(t)}{S(t)} = \lambda(t)$$

و در حالت کلی می‌توانیم داشته باشیم $\Lambda(t) = \int_0^t \lambda(u)du$ که این فرمها همگی به تابع مخاطره تجمعی دلالت دارد .

چون داریم $S(0) = 1$, $F(0) = 0$, $\lim_{t\to\infty}[F(t)] = 1$ پس می‌توان نوشت $\lim_{t\to\infty}[S(t)] = 0$ و $\lim_{t\to\infty}[\Lambda(t)] = \infty$

۱.۲ تعریف :

یک متغیر نا منفی X با تابع توزیع F و تابع بقای S دارای نرخ شکست صعودی[۳] (IFR) است اگر و تنها اگر $\log S$ - روی $\{t : S(t) > 0\}$ محدب باشد یعنی S مقعر لگاریتمی‌است به طور معادل اگر $\lambda(t)$ تابع مخاطره X باشد آنگاه دارای IFR است اگر و تنها اگر نسبت به t صعودی باشد متناظرا متغییر تصادفی نا منفی X دارای خاصیت شکست نزولی[۴] (DFR) است اگر $\log S$– روی $\{t : S(t) > 0\}$ مقعر باشد یعنی S محدب لگاریتمی‌است به طور معادل اگر

[۱] Hazard rate
[۲] Cumulative Hazard rate
[۳] Increasing Failure Rate (IFR)
[۴] Decreasing Failure Rate (DFR)

$\lambda(t)$ تابع مخاطره X باشد آنگاه DFR است اگر و تنها اگر نسبت به t نزولی باشد به طور مثال توزیع وایبل با پارامتر شکلی α چنانچه $\alpha > 1$ باشد خاصیت IFR را داریم چنانچه $\alpha < 1$ باشد DFR را داریم در واقع بترتیب نسبت به حالت اول تابع مخاطره صعودی ونسبت به حالت بعد نزولی است از این موضوع در فصل‌های بعد استفاده خواهیم کرد .

پیش بینی طول عمر: این موضوع وقتی معنا پیدا می‌کند که واحد آزمایشی لااقل تا زمان t_0 بقا داشته باشد وبه معنای باقی مانده طول عمر تا زمان مرگ است که آن را با $T - t_0$ نشان می‌دهیم از این تعریف برای روشن ساختن نرخ طول عمر استفاده می‌کنیم.

نرخ طول عمر: عبارتست از میزان طول عمر مورد انتظار برای واحد / واحدهای آزمایشی و عبارتست از احتمال اینکه مرگ آزمودنی در زمان $t + t_0$ یا قبل از این زمان است به شرط آن که لااقل t_0 واحد زمانی حیات داشته باشد معنای آینده کاملا در آن واضح است ، داریم :

$$P\left(T < t_0 + t \mid T > t\right) = \frac{P\left(t_0 < T < t_0 + t\right)}{P\left(T > t\right)} = \frac{F\left(t_0 + t\right)}{S\left(t\right)}.$$

بنابراین تابع چگالی طول عمر آینده برابر است با :

$$\frac{d}{dt}\frac{F\left(t_0 + t\right) - F\left(t_0\right)}{S\left(t_0\right)} = \frac{f\left(t_0 + t\right)}{S\left(t_0\right)} \qquad (۵.۱)$$

و نهایت امر طول عمر مورد انتظار برابر است با :

$$\frac{1}{S\left(t_0\right)}\int_0^\infty t\, f\left(t + t_0\right)dt = \frac{1}{S\left(t\right)}\int_{t_0}^\infty S\left(t\right)dt \qquad (۶.۱)$$

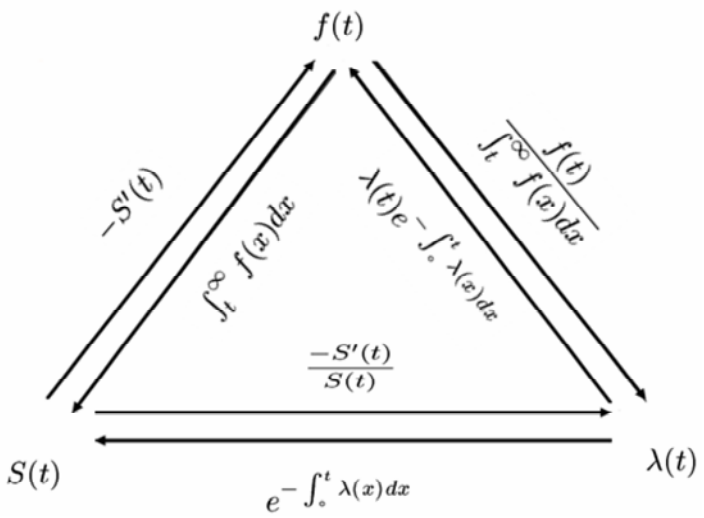
که $t_0 = 0$ به معنای تولد موجود زنده یا شرع به کار یک مولفه مکانیکی است و دراین زمان کمترین طول عمر مورد انتظار را داریم .در مسائل قابلیت اعتماد از طول عمر و طول عمر مورد انتظار به ترتیب تحت عنوان متوسط عمر تا شکست و متوسط طول عمر باقیمانده یاد می‌شود [۱] .

احتمال بقا هر واحد بصورت تکی تا زمان T یا بعد از آن برابر با S(t) است و احتمال بقا n واحد با فرض یکی بودن تابع بقا برای همه واحدها برابر با $n \times S(t)$ است همانند توزیع دوجمله ای که برآمد پیروزی یا شکست را داشتیم اینجا بقا یا شکست را داریم بنابراین چنانچه واریانس نسبت بقا را خواسته باشیم می‌توانیم بنویسیم : $S(t) \times \frac{1 - S(t)}{n}$

اگر بخواهیم سنی را که واحد آزمایشی می‌تواند تا آن زمان به بقای خود ادامه دهد را تعیین کنیم می‌توانیم از حل معادله $S(t)=q\ ;\ \forall t$ که آن را بدست آوریم که q چندک مطرح شده در مساله است در بسیاری از موارد علاقه مند به محاسبه میانه طول عمر هستیم در این صورت $q=\frac{1}{2}$ قرار می‌دهیم چنانچه بخواهیم چندک های $q=0.90$ یا $q=0.99$ را محاسبه کنیم در این صورت می‌توانیم به این کار مبادرت ورزیم .

مدلهای بقا بطور مفیدی می‌توانند در مدل های رگرسیونی معمولی وقتی که متغیر پاسخ (متغیر وابسته) زمان است به ایفای نقش بپردازند با این وجود محاسبه تابع درستنمایی تحت سانسور(نیازمند برازش پارامترها یا گرفتن استنباط به طرق دیگر است)پیچیده است به مبحث سانسورها در فصل بعد بطور کامل خواهیم پرداخت .

رابطه بین تابع چگالی احتمال و تابع بقا و تابع نرخ خطر :



اصولا در برخورد با داده های بقا (طول عمر) از سه روش جداول طول عمر – روش کاپلان ومایر[۱] و مدل های آماری استفاده می‌شود که در دو روش اول برای برآورد میزان بقا فقط از مدت زمان بقا استفاده می‌کند به عبارت دیگر از سایر متغیرها هیچگونه استفاده ای نمی‌شود ولی در روش سوم کلیه متغیرها وارد مدل می‌شوند و تابع مخاطره با توجه به تاثیر کلیه متغیرها برآورد می‌شود اگر حجم نمونه در کلیه گروهها ی تحت مطالعه از ۳۰ بیشتر باشد از روش جداول عمر استفاده می‌شود و چنانچه از ۳۰کمتر باشد روش حاصلضرب کاپلان مایر مناسب تر است چنانچه نتوان تابع نرخ خطر را تعیین کرد مدل رگرسیونی کاکس که یک مدل ناپارامتری است مورد استفاده قرار می‌گیرد ذیلا به روش حاصلضرب کاپلان مایر که یک

[۱] Kaplan and Meier

روش ناپارامتری برای برآورد تابع بقا ست می‌پردازیم در فصل های ۳ و ۴ به روشهای پارامتری تحت توزیع‌های مطرح شده می‌پردازیم .

وقتی زمان های بقا از یک تابع توزیع پیروی می‌کنند ، روشهای ناپارامتری کم کارآمدتر از روشهای پارامتری هستند و زمانیکه هیچ توزیع برای داده‌ها در دسترس نیست ، کارآمدترند بنابراین روشهای ناپارامتری برای آنالیز داده های بقا داده‌ها زمانی پیشنهاد می‌شود که توزیعی برای داده‌ها متصور نباشد در سال (۱۹۵۸) کاپلان ومایر روش ناپارامتری برای برآورد تابع بقا پیشنهاد کردند که با افزایش روز افزون توانایی ابر کامپیوترها برای نمونه های با حجم‌های متفاوت چه بزرگ چه کوچک و چه متعادل مناسبت دارد که تحت عنوان حاصلضرب حدی PL در زیر معرفی می‌شود .

برای نمونه ای با تابع بقای $S(t)$ می‌توان میانه تابع بقا را نمودار تابع $\widehat{S}(t)$ برآورد کرد این روش وقتی داده‌ها رو به زوال باشند می‌تواند بکار گرفته شود و اگر برخی از داده‌ها در پایان مطالعه هنوز باقی بمانند ، از روش برآورد حاصلضرب حدی PL کاپلان- مایر استفاده می‌شود .

۱. ۳ مثال‌هایی ازمبحث تحلیل بقا :

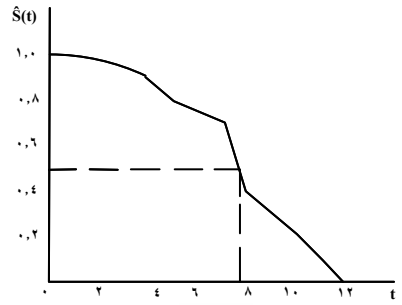
مثال ۱.۱ : فرض کنید همه بیماران در حال مرگند طوریکه زمانهای بقا واقعی معلوم است و $t_1,\ldots,t_n$ زمانهای بقای واقعی n فرد مورد مطالعه باشد. بطور مفهومی، ما این گروه بیماران را بصورت یک نمونه تصادفی از یک جامعه خیلی بزرگتر درنظر می‌گیریم. n زمان بقای $t_1,\ldots,t_n$ را به ترتیب صعودی بصورت $t_1 < t_2 < \ldots < t_n$ مرتب می‌کنیم. از آنجایی که تابع بقا را می‌توان از تابع بقای تجمعی بدست آورد اگر تعداد بیماران پس از زمان t را با A نشان دهیم با علم به اینکه تعداد کل بیماران n است در آن صورت برآورد تابع بقا برابر با $\widehat{S}(t)=\frac{A}{n}$ است که $A=n-i$ است پس برآورد تابع بقا برای زمان i برابر با $\widehat{S}(t_{(i)})=1-\frac{i}{n}$ اگر دو یا بیشتر از دو مشاهده زمانهایشان یکی باشد i مربوط به بزرگترین رخداد ترتیبی لحاظ می‌شود مثلا اگر $t_{(2)}=t_{(3)}=t_{(4)}$ باشند در آن صورت داریم.

$$\widehat{S}(t_{(2)})=\widehat{S}(t_{(3)})=\widehat{S}(t_{(4)})=\frac{n-4}{n}$$

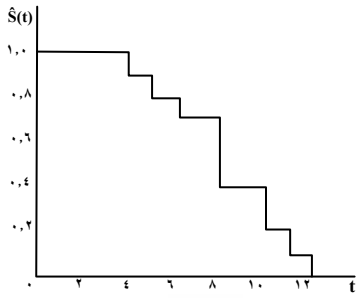
با این برآورد محافظه کارانه برای مشاهدات برابر می‌توانیم ، نتیجه بگیریم که هر واحد در شروع مطالعه (زنده است / کار می‌کند) و هیچ واحدی بیش از زمان $t_{(n)}$ حیات ندارد. $\widehat{S}(t_{(0)})=1\,,\,\widehat{S}(t_{(n)})=0$

برآوردهای تابع بقا چه در نقطه شروع چه در نقطه پایان وچه در زمان i می‌رساند که $\widehat{S}(t)$ تابعی پله ای است که از نقطه یک شروع با گامهای $\frac{1}{n}$ کاهشی در نهایت به صفر منتهی می‌شود .

مثال ۲.۱ : یک آزمایش بالینی را که در آن ۱۰ بیمار سرطان ریه‌ای رو به مرگند در نظر بگیرید.در جدول (۱) برآوردهای بقا را برحسب ماه ذکر شده است تابع $\hat{S}(t)$ از فرمول $\widehat{S}(t_{(i)})=1-\frac{i}{n}$ محاسبه شده و به صورت منحنی در شکل (۱. ۱) و به صورت یک تابع پله‌ای در شکل (۱. ۲) رسم شده است.



شکل(۱.۱) : تابع منحنی شکل بقا برحسب زمان ماه



شکل(۲.۱) : تابع پله ای بقا برحسب زمان ماه

i	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
t	۴	۵	۶	۸	۸	۸	۱۰	۱۰	۱۱	۱۲
$\hat{S}(t)$	۰/۹	۰/۸	۰/۷	۰/۷	۰/۴	۰/۴	۰/۲	۰/۲	۰/۱	۰/۱

جدول (۱.۱) : محاسبه $\widehat{S}(t)$ برای ۱۰ بیمار مبتلا به سرطان ریه

برآورد میانه تابع بقا در شکل (۱.۱) برابر با ۸ ماه و در شکل ۲ برابر با ۷.۶ ماه است برآورد صحیح تر می‌تواند از درون یابی خطی بدین صورت محاسبه گردد.

t	$\hat{S}(t)$
۶	۰/۷
m	۰/۵
۸	۰/۴

$$\Rightarrow \qquad \frac{8-6}{0.4-0.7}=\frac{8-m}{0.4-0.5}\Rightarrow m=7.3$$

از لحاظ تئوری$\hat{S}(t)$ باید بصورت یک تابع پله‌ای رسم شود زیرا بین دو زمان بقای واقعی می‌دانیم این برآورد بایستی ثابت باقی بماند با این حال، همانطور که در مثال مشخص شد وقتی متوسط زمان بقا باید از یک منحنی بقا برآورد شود یک منحنی مانند شکل (۱.۱) ممکن است برآورد بهتری از تابع پله‌ای بدهد این روش فقط اگر همه بیماران رو به فوت باشند ، می‌تواند بکاررود.

اگر برخی از بیماران در پایان مطالعه هنوز زنده باشند، یک روش متفاوت برآورد $\hat{S}(t)$ وجود دارد که برآورد حاصلضرب حدی PL که توسط کاپلان و مایر (۱۹۵۸) ارائه شده لازمست [۳۸] . به مثال زیر توجه کنید .

مثال ۳.۱ : فرض کنید ۱۰ بیمار در شروع سال ۲۰۰۰ به یک مطالعه بالینی ملحق شوند در طول سال ، ۶ بیمار می‌میرند و ۴ نفرزنده می‌مانند. در پایان سال اول ودر شروع سال دوم۲۰نفر دیگر به مطالعه اضافه می‌شوند در سال ۲۰۰۱، سه بیماری که در شروع سال ۲۰۰۰ وارد شده بودند و ۱۵ بیماری که بعدا" وارد شدند می‌میرند. بازماندگان، به ترتیب یک و پنج نفرند فرض کنید که مطالعه در پایان سال ۲۰۰۱ خاتمه یابد و می‌توانیم نسبت بیماران در جامعه را که به مدت دو سال یا بیشتر زنده ماندند بصورت $S(۲)$ برآورد کنیم اولین گروه بیماران برای ۲ سال و گروه دوم برای یک سال تحت مطالعه قرار داشتند .یک برآورد ممکن برابر است با $\hat{S}(2) = 0.10$ که برآوردی تقلیل یافته از نمونه است چرا که ۲۰ بیماری که در ابتدای سال دوم به مطالعه وارد شدند، نادیده گرفته می‌شوند. طبق نظر کاپلان و مایر نمونه دوم که فقط برای یک سال مطالعه شده ، می‌تواند در برآورد $S(۲)$ سهم داشته باشد .

بیمارانی که دو سال جان سالم به دربرده اند ، ممکن است بصورت بازمانده سال اول درنظر گرفته شوند و پس از آن بازمانده یک سال بیشتر. به این ترتیب ، احتمال زنده ماندن برای دو سال یا بیشتر برابر است با احتمال زنده ماندن در سال اول و پس از آن بازمانده یک سال بیشتر ، آنها این مطلب را در برآوردشان لحاظ کردند ، داریم :

$S(۲) = P$ (باقی ماندن یکسال بیشتر و بقای اولین سال)

$S(۲) = P$(باقی ماندن از سال اول) P × (بازمانده دو سال به شرطی که از اولین سال زنده باقی مانده باشند)

$\hat{S}(۲) =$ (نسبت بیماران بازمانده یکسال) × (نسبت بیماران بازمانده ازدوسال به شرطی که برای یکسال زنده بودند)

برای اطلاعات داده شده بالا یکی از ۴ بیماری که از سال اول زنده مانده بود ، ۲ سال جان سالم به دربرد. بنابراین اولین نسبت برای $\hat{S}(2)$ مساوی ۰/۲۵ است.۴ تا از ۱۰ بیماری که در شروع سال ۲۰۰۰ وارد مطالعه شدند و ۵ تا از ۲۰ بیماری که در پایان سال ۲۰۰۰ وارد مطالعه شدند به مدت یک سال زنده باقی‌ماندند پس دومین نسبت در $\hat{S}(2)$ برابر ۰/۳۰ است :

برآورد PL برای S(۲) برابر است با : $\hat{S}(2)=\frac{1}{4}\times\frac{4+5}{10+20}$

این روش بصورت زیر تعمیم داده شود:

احتمال بقای k سال یا بیشتر (k≥۲) از شروع مطالعه برابر با حاصلضرب k میزان بقای مشاهده شده است .

$$\hat{S}(t)=p_1\times\ldots\times p_k$$

p_1 : نشان دهنده نسبت بیماران باقیمانده برای حداقل یک سال است.

p_2 : نسبت بیماران بازمانده دومین سال بعد از اینکه آنها از سال اول زنده ماندند .

p_3 : نسبت بیماران بازمانده سومین سال بعد از اینکه آنها برای دو سال زنده ماندند .

p_k : نسبت بیماران بازمانده k امین سال بعد از اینکه آنها ۱-k سال زنده ماندند .

بنابراین برآورد PL برای احتمال بقای هرچند سال خاص از شروع مطالعه برابر است با حاصلضرب برآورد یکسان تا سال قبل ومیزان بقای مشاهده شده برای سال خاص است این برآوردگرها برآوردهای درستنمایی ماکزیمم نیز هستند.

برسلو وهمکارانش[۱] (۱۹۷۴) و مایر (۱۹۷۵) نشان داده‌اند که تحت شرایط خاص ، برآورد کاپلان و مایر ثابت و نرمال است. با این حال ، چند ویژگی مهم که آنها به آن رسیدند را اینجا ذکر می‌کنیم.

۱- برآوردهای کاپلان - مایر به فاصله‌های زمانی که در آن مشاهدات نزول می‌کنند محدوداند. اگر بزرگترین مشاهده سانسور نشده باشد برآورد PL در آن زمان برابر صفر است اگر بزرگترین مشاهده سانسور شده باشد برآورد PL نمی‌تواند هرگز مساوی صفر باشد و پس از بزرگترین مشاهده چیزی تعریف نشده است چون نسبت به آن اطلاعی نداریم .

[۱] Borsello and et al

معمول ترین آماره تلخیصی استفاده شده در آنالیز بقا، میانه زمان بقاست. یک برآورد ساده میانه از منحنی‌های بقا برآورد شده بوسیله روش PL که در زمان t ، $\hat{S}(t) = ۰/۵$ است .

۲- اگر کمتر از ۵۰٪ مشاهدات سانسورنشده باشد و بزرگترین مشاهده سانسورشده باشد، میانه زمان بقا نمی‌تواند برآورد شود. یک راه عملی برای برخورد با این وضعیت، استفاده از احتمالات بقای یک طول زمانی داده شده است مثلا" ۱ و ۳ یا ۵ سال یا میانه زمان بقای محدود به یک زمان t داده شده است .

۳- در روش PL فرض می‌شود که زمان های سانسور شده مستقل از زمان های بقا هستند. به عبارت دیگر، دلیل اینکه یک مشاهده سانسور شده است به علت مرگ نیست اگر بیمار هنوز در پایان دوره مطالعه زنده باشد. با این حال اگر بیمار دچار عوارض جانبی شدید درمان شده و قبل از مرگ مجبور به ترک بررسی گردد یا در صورتی که بیمار به علتی غیر از مطالعه بمیرد (به عنوان مثال ، مرگ به علت تصادفات خودرو) فرض نقض می‌شود. وقتی یک سانسور نامناسب وجود دارد ، روش PL مناسب نیست .

۴- همانند دیگر برآوردگرها ، خطای استاندارد (S.E) برآوردگر کاپلان - مایر برای $S(t)$ نشانه‌ای از خطای بالقوه $\hat{S}(t)$ می‌دهد یک فاصله اطمینان خوب دقت به مراتب بیشتری نسبت به برآورد نقطه‌ای $\hat{S}(t)$ دارد[۳۴] . بنابراین یک فاصله اطمینان ۹۵٪ برای $S(t)$ برابر است با :

$$\hat{S}(t) \times 1.96\, S.E.\left\{\hat{S}(t)\right\}$$

فصل دوم

سانسور یک پدیده متداول در آزمون‌های طول عمرومطالعات قابلیت اعتماد می‌باشد. داده‌ها و به تبع آن نمونه‌های سانسور شده سبب تمایز تحلیل بقا (تحلیل بقا به روش‌های تحلیلی زمان خاتمه پیشامد اشاره می‌کند) ازسایر تحلیل‌های آماری می‌شوند. این مجموعه داده‌ها، داده‌های مرتبط با واحدهای آزمایشی هستند که ممکن است در طول مطالعه به طور کامل شرکت نداشته باشند و یا تا پایان مطالعه خراب نشوند. باید توجه کرد که داده‌های سانسور شده را باید با استفاده ازآخرین اطلاع از واحد دقیقا ثبت کرد تا در تحلیل داده‌ها مورد استفاده قرار گیرند. نمونه‌های سانسور شده، نمونه‌هایی هستند که در آن هر عضو نمونه در ناحیه‌ای محدود از فضای نمونه قرار می‌گیرند به این معنی که زمان مورد بررسی از قبل مشخص است بنابراین فضای نمونه محدود شده است پس برخی از اعضا شمارش می‌شوند ولی اندازه گیری نمی‌شوند.

نمونه‌های سانسور شده اغلب نتیجه‌ای از آزمایش‌های طول عمر و زمان واکنش است در تالیفات سالهای اخیر نمونه‌های سانسور شده به عنوان نمونه‌های بریده‌ای تلقی می‌شوند که در آنها تعداد مشاهدات گمشده (اندازه‌گیری نشده) معلوم است. کریچ[۱] (۱۹۴۹) برای اولین بار عنوان سانسور شده را برای این نمونه پیشنهاد کرد.

۲. ۱ : انواع مختلف سانسور

۲. ۱. ۱- سانسور از راست: تحت این سانسور، مطالعه تا زمان t_c که از قبل توسط آزمایشگر مشخص شده است ادامه داده می‌شود تحت طرح سانسور از راست مشاهدات عبارتند از:

$$Y_i = \begin{cases} X_i & X_i \le t_c \\ t_c & X_i > t_c \end{cases}$$

[۱] Kerrich