مجله دانشگاه علوم پزشکی و خدمات بهداشتی- درمانی قزوین
سال دوم - شماره ۶ و ۷

# مقاله مروری

# REVIEW ARTICLE

## معرفی تحلیل بقاء

دکتر نصراله بشردوست*

**چکیده :**

*تحلیل بقاء روشی برای جمع‌آوری نتایج یک مطالعه برحسب زمان است. در این روش در هر یک از فاصله‌های زمانی شمار افراد مورد مطالعه‌ای که از آغاز آن فاصله هنوز در مطالعه باقی مانده‌اند و تعداد کسانی که در آن فاصله پیامد مورد نظر را داشته‌اند و شمار گروهی از افراد که از مطالعه خارج شده و یا مشارکت خود را در مطالعه در آن فاصله قطع کرده‌اند ثبت می‌شود. از این اعداد ، احتمال بروز پیامد در هر یک از فاصله‌های زمانی و در پایان مطالعه بدست می‌آید. این روش استفادهٔ بهینه از داده‌های گردآوری شده را برای محاسبات بروز ممکن می‌سازد.*

**کلید واژه‌ها :** تحلیل بقاء

***Keywords :*** *Survival analysis*

* دانشیار دانشکده بهداشت دانشگاه علوم پزشکی اصفهان

## مقدمه :

در مطالعات تحلیلی که در علوم پزشکی انجام می‌شود معمولاً اثر عامل یا عواملی را بر یک نتیجه مشخص و معین مورد بررسی قرار می‌دهند. عامل مورد نظر که به شکل تعریف شده‌ای اندازه‌گیری می‌شود به متغیر مستقل معروف است و نتیجهٔ کار نیز که می‌تواند یک متغیر کمی یا کیفی باشد متغیر وابسته نامیده می‌شود. مثلاً اگر دو داروی $A$ و $B$ در درمان بیماری $X$ مورد مقایسه قرار گیرند نوع دارو ($A$ یا $B$) متغیر مستقل و طول مدت درمان متغیر وابسته نامیده می‌شود.

در بسیاری از مطالعات پزشکی متغیر وابسته مورد نظر یا نتیجهٔ کار مدت زمانی است که باید به انتظار بمانیم تا یک حادثه اتفاق بیفتد. منظور از مدت زمان انتظار سالها ، ماهها ، هفته‌ها و یا روزهایی است که از پیگیری کردن یک فرد مورد مطالعه می‌گذرد تا حادثهٔ مورد نظر پیش آید و منظور از حادثه یا اتفاق ، حوادثی چون مرگ ، بروز بیماری ، عود ، بهبودی و یا هر حادثه دیگری است که ممکن است بوجود آید.

گرچه ممکن است وقوع بیش از یک حادثه مورد نظر باشد که در آن صورت مسئله عوامل خطر رقیب مطرح می‌شود لکن ما در این مختصر فرض می‌کنیم فقط یک اتفاق مورد توجه است. برای روشن شدن بهتر به ذکر چند مثال می‌پردازیم :

۱- مطالعه‌ای که بیماران لوسمی را در حال معالجه پیگیری می‌کند تا به مرگ برسند. در اینجا حادثه مرگ و زمان بقا مدت زمان زنده ماندن است.

۲- در یک مطالعه که بیماران سالم را سالها پیگیری می‌کند تا به بیماری قلبی مبتلا شوند. در اینجا حادثه بیماری قلبی و زمان بقاء مدت زمانی است که فرد سالم مانده است.

۳- مطالعه‌ای که سیزده سال افراد ۶۰ سال به بالا را پیگیری می‌کند تا مرگ آنها فرا رسد. در اینجا حادثه مرگ و زمان بقاء مدت زمان زنده ماندن است.

۴- مطالعه‌ای که در مورد زندانیان تازه آزاد شده انجام می‌شود تا دوباره مرتکب خطا شده و دستگیر شوند. در اینجا حادثه دستگیری مجدد و زمان بقاء مدت زمان آزادی است.

۵- در تحقیقی که برآورد مدت زمان زنده ماندن بیمار بعد از پیوند قلب مورد نظر است ، حادثه مورد نظر مرگ و زمان بقاء مدت زمان زنده ماندن بعد از پیوند قلب است.

در تمام مثال‌های بالا متغیر وابسته مدت زمانی است که محقق باید منتظر بماند تا حادثه‌ای اتفاق بیفتد. تجزیه و تحلیل اطلاعاتی از این قبیل را تحلیل بقاء می‌گویند.

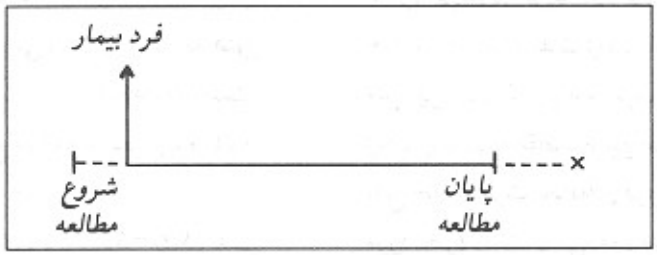
## اطلاعات سانسور شده :

در اغلب مطالعات تحلیل بقاء یک مشکل کلیدی بنام داده‌های سانسور شده (دم بریده) وجود دارد که جزء مشخصه‌های اصلی این روش تجزیه و تحلیل است. در واقع سانسور شدن وقتی اتفاق می‌افتد که ما قسمتی از اطلاعات بقاء را داریم لکن دقیقاً مدت زمان بقاء را نمی‌دانیم. بعنوان مثال ، بیماران لوسمی را در نظر بگیرید که پیگیری می‌شوند تا از حالت بیماری خارج شوند. اگر این حادثه را با $X$ نشان دهیم و حادثه وقتی اتفاق بیفتد که مطالعه پایان یافته در این صورت مدت زمان بقاء سانسور شده است.

در این مثال می‌دانیم فرد مورد نظر حداقل در دوره‌ای که پیگیری می‌شده زنده مانده است اما دقیقاً نمی‌دانیم که زمان حادثه چه وقت بوده است (شکل شماره ۱).

شکل ۱ :

نمایشی از سانسور شدن




بطور کلی سه دلیل برای سانسور شدن اطلاعات وجود دارد :

۱- وقتی که حادثه مورد نظر قبل از زمان از پیش تعیین شدهٔ پایان مطالعه اتفاق نمی‌افتد.

۲- فرد مورد مطالعه را گم می‌کنیم و نمی‌دانیم چه بر سر او آمده است.

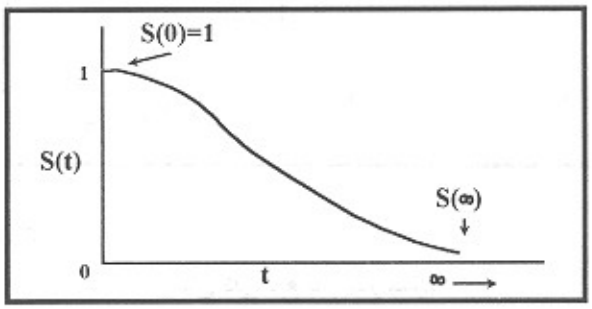
۳- فرد مورد مطالعه به دلایلی چون مرگ (اگر مرگ حادثه مورد نظر باشد) تصادف و یا حوادث غیرمترقبه دیگری از مطالعه خارج می‌شود و ما به اتفاق مورد نظر نمی‌رسیم.

## ◙ هدف و روش مطالعاتی :

اصلی‌ترین و در واقع هدف نهائی این نوع مطالعات مدت زمان بقاء است که با $T$ نشان داده می‌شود و این با فرض آن است که $T\geq 0$ و یک متغیر تصادفی باشد و هر مقدار معین برای $T$ را با $t$ نشان می‌دهند. به عنوان مثال می‌خواهیم بدانیم که بعد از پنج سال پیوند قلب ، شانس زنده ماندن برای یک فرد چقدر است $t=5$ و $T\geq 5$ می‌باشد. بر این اساس تابع بقاء (Survival function) را با $S(t)$ نشان می‌دهند. اصطلاح دیگری که در این رابطه مطرح می‌شود تابع خطر (hazard function) است که درک مفهوم این اصطلاح برای کسی که تحلیل بقاء را انجام می‌دهد حیاتی به نظر می‌رسد. لذا به منظور روشن شدن بهتر مفاهیم به شرح و تفسیر آنها می‌پردازیم.

تابع بقاء یا $S(t)$ در واقع احتمال زنده ماندن فرد بعد از زمان $t$ را نشان می‌دهد. به عبارت دیگر $S(t)$ احتمال آنکه متغیر تصادفی $T$ از زمان $t$ بیشتر شود را محاسبه می‌نماید که فرم ریاضی آن به شکل $S(t)=P(T>t)$ نشان داده می‌شود.

به طور تئوری وقتی $t$ از صفر تا بینهایت تغییر می‌کند ، $S(t)$ از یک تا صفر تقلیل می‌یابد (شکل شماره ۲).




شکل ۲ :

شمای کلی تابع بقاء (بطور تئوری)

تمام توابع بقاء دارای خصوصیات زیر هستند :

۱- توابع نزولی هستند ، بدین معنی که وقتی $t$ افزایش پیدا می‌کند $S(t)$ کاهش می‌یابد.

۲- در زمان صفر $(t=0)$ ، یعنی در شروع مطالعه هنوز حادثه‌ای اتفاق نیفتاده پس $S(t)=S(0)=1$ است. بعبارت دیگر در زمان شروع مطالعه شانس زنده ماندن افراد بعد از آن لحظه صد در صد است.

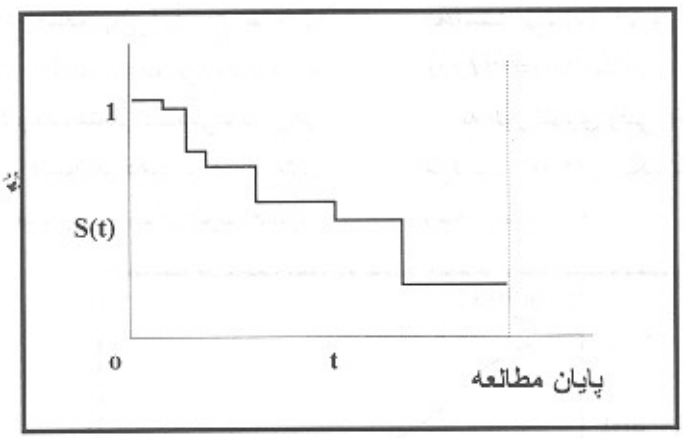
۳- در زمان خیلی دور $(t=\infty)$ ، $S(t)=S(\infty)=0$ ، بدین معنی که اگر مطالعه تا زمان خیلی زیادی طول بکشد بالاخره همه می‌میرند. بنابر این زمانی حادثه اتفاق می‌افتد و شانس زنده ماندن افراد تابینهایت صفراست.

در اینجا لازم به ذکر است که در عمل و در زمان‌های محدود شکل منحنی بقاء پله‌ای خواهد بود (شکل شماره ۳).

تابع خطر (Hazard function) معمولاً با $h(t)$ خوانده شده و با زبان ریاضی به شکل $h(t)=\lim_{\Delta t\to 0} \frac{P(t\le T<t+\Delta t \mid T\ge t)}{\Delta t}$ نشان داده می‌شود. تفسیر این مفهوم ریاضی خیلی ساده نیست لکن برای تفهیم بهتر می‌توان آن را به نوعی تشریح کرد. تابع خطر در واقع پتانسیل فرد برای رسیدن به حادثه مورد نظر را اندازه می‌گیرد. هرچه این توانایی بیشتر باشد زمان حادثه نزدیکتر و مدت زمان بقاء کوتاهتر است. درست مثل این است که هرچه سرعت ماشین بیشتر باشد زمان رسیدن به مقصد معین کوتاهتر است. در حقیقت تابع خطر قدرت لحظه‌ای فرد در نیل به هدف (حادثه مورد نظر) را نشان می‌دهد. پس روشن است که تابع خطر و تابع بقاء برعکس یکدیگر عمل می‌کنند. به همین دلیل است که براساس روابط ریاضی موجود ، ما می‌توانیم با داشتن یکی از آنها دیگری را بدست آوریم. فرمول کلی که ارتباط این دو مشخصه را نشان می‌هد به شکل زیر است.

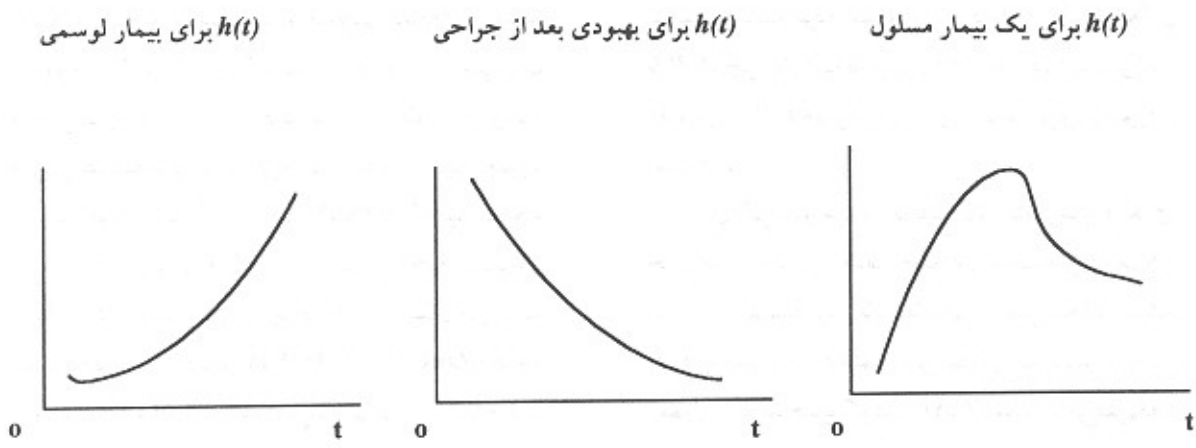
$$h(t) = -\left[\frac{ds(t)/dt}{s(t)}\right] \quad \text{و} \quad S(t) = \exp[-\int_0^t h(u)du]$$

لازم به ذکر است که اگر تابع بقاء یک احتمال است و از صفر تا یک تغییر می‌کند تابع خطر یک میزان است و از صفر تا بینهایت تغییر می‌نماید. نمونه‌هایی از تابع خطر در شکل شماره ۴ آورده شده است.




شکل ۳ :

شمای کلی تابع بقاء (در عمل)




**شکل ۴ :**

**نمونه‌هایی از تابع خطر در شمای متفاوت**

نکته قابل ذکر این است که در تحلیل بقاء ، یک متغیر تصادفی دیگر که معمولاً از حرف یونانی $\partial$ برای نشان دادن آن استفاده می‌کنند برای سانسور شدن بکار برده می‌شود. بدین معنی که در زمان $t$ ، $\partial(t)=1$ ، اگر حادثه اتفاق افتاده باشد و $\partial(t)=0$ اگر سانسور شده باشد.

بطور خلاصه می‌توان اهداف تحلیل بقاء در یک مطالعه تحقیقاتی را به یکی از سه شکل زیر خلاصه کرد :

۱- برآورد و تفسیر تابع بقاء و تابع خطر از حادثهٔ مورد نظر با استفاده از اطلاعات موجود در مطالعه

۲- مقایسهٔ تابع خطر یا تابع بقاء در شرایط متفاوت

۳- تعیین رابطهٔ مدت زمان بقاء با متغیرهای مورد نظر

به منظور آشنایی با چگونگی آماده کردن داده‌های یک مطالعه برای ورود به کامپیوتر به مثال زیر توجه فرمایید.

۴۲ نفر که به بیماری لوسمی مبتلا بوده‌اند بطور تصادفی به دوگروه ۲۱ نفری تقسیم شده‌اند یک گروه تحت درمان معینی قرار گرفته و به گروه دیگر پلاسیبو داده شده است. مطالعه ۳۵ هفته طول کشیده و اطلاعات بدست آمده در جدول شماره ۱ آورده شده است. لازم به ذکر است که علامت + در کنار اعداد به معنی سانسور شدن اطلاعات فرد است.

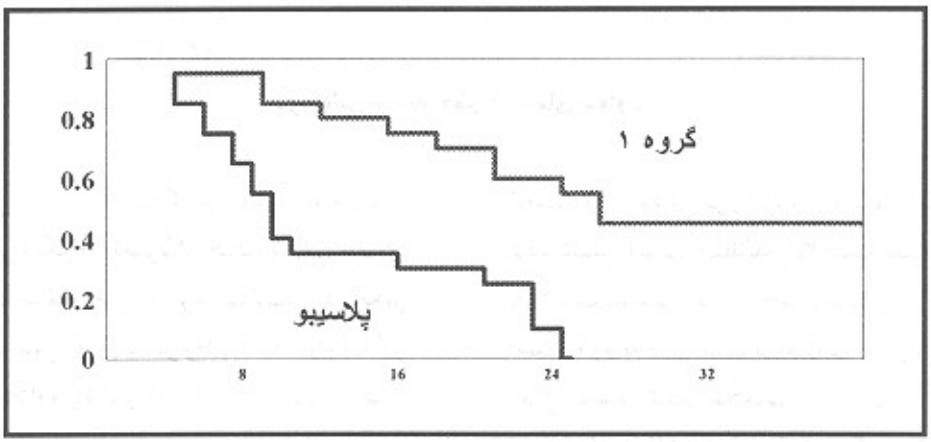
براساس اطلاعات خام فوق در گروه ۱ برای ۹ نفر حادثه اتفاق افتاده و اطلاعات ۱۲ نفر هم سانسور شده است. در گروه دیگر کسی سانسور نشده و برای همه حادثه مورد نظر اتفاق افتاده است. در این صورت اطلاعاتی که برای ورود به کامپیوتر آماده می‌شود باید همانند جدول شماره ۲ باشد. لازم به ذکر است که در جدول فوق متغیر $G$ نشان‌دهندهٔ آن است که فرد در گروه درمان $(G=1)$ و یا در گروه پلاسیبو $(G=0)$ قرار گرفته است.

اگر محقق اطلاعات خود را به شکل یاد شده آماده نماید بسیاری از نرم‌افزارهای کامپیوتری می‌توانند اطلاعات را تجزیه و تحلیل نمایند.

یکی از روش‌های معمول و ابتدایی تجزیه و تحلیل

اطلاعات بقاء استفاده از منحنی کاپلان - مایر (Kaplan-Meier) است که با استفاده از آن می‌توان منحنی‌های بقاء دو گروه ذکر شده در مثال بالا را به خوبی مقایسه کرد (شکل شماره ۵). در این مورد آزمونی موسوم به لاگ رنک (Log-Rank) نیز نتیجهٔ مقایسه را به زبان آماری و با بیان درجهٔ اطمینان (P-value) روشن می‌سازد. در مورد مثال فوق با استفاده از نرم‌افزار SAS ۱۶/۷۹۳ = Log-Rank و P<۰/۰۰۱ بیان شده است. پس می‌توان استنتاج کرد که زمان بقاء گروه مورد درمان بطور معنی‌داری از گروه پلاسیبو بیشتر بوده است. تست دیگری تقریباً معادل با L.R وجود دارد که به آزمون Peto معروف است و کمی دقیق‌تر از L.R عمل می‌کند (از توان بالاتری برخوردار است).

حال اگر بخواهیم در تحلیل بقاء متغیرهای دیگری چون سن، جنس، طبقهٔ اجتماعی و غیره را که ممکن است در مقایسهٔ دو زمان بقاء مورد نظر محقق باشد مورد بررسی قرار دهیم و اثر آنها را بدانیم از روش معمول و نسبتاً جدید کوکس (Cox) استفاده می‌کنیم که به روش خطرنسبی (Proportional Hazard) معروف است.




شکل ۵ :

منحنی (KM) برای مقایسهٔ دو گروه

جدول ۱ :

اطلاعات مربوط به بیماران لوسمی

| گروه ۲ پلاسیبو | گروه ۱ تحت درمان |
|---|---|
| ۵، ۵، ۴، ۳، ۲، ۲، ۱، ۱، | ۲۳، ۲۲، ۱۶، ۱۳، ۱۵، ۷، ۶، ۶، ۶ |
| ۱۵، ۱۲، ۱۲، ۱۱، ۱۱، ۸، ۸، ۸، ۸، | ۲۵+، ۲۰+، ۱۹، ۱۷+، ۱۱+، ۱۰+، ۹+، ۶+ |
| ۲۳، ۲۲، ۱۷ | ۳۵+، ۳۴+، ۳۲+، ۳۲+ |

جدول ۲ :

اطلاعات آماده شده برای ورود به کامپیوتر در تحلیل بقاء

| H | t | $\partial$ | G |
|---|---|---|---|
| 1 | 6 | 1 | 1 |
| 2 | 6 | 1 | 1 |
| 3 | 6 | 1 | 1 |
| 4 | 7 | 1 | 1 |
| 5 | 10 | 1 | 1 |
| 6 | 13 | 1 | 1 |
| 7 | 16 | 1 | 1 |
| 8 | 21 | 1 | 1 |
| 9 | 23 | 1 | 1 |
| 10 | 6 | 0 | 1 |
| 11 | 9 | 0 | 1 |
| 12 | 10 | 0 | 1 |
| 13 | 11 | 0 | 1 |
| 14 | 17 | 0 | 1 |
| 15 | 19 | 0 | 1 |
| 16 | 20 | 0 | 1 |
| 17 | 25 | 0 | 1 |
| 18 | 32 | 0 | 1 |
| 19 | 32 | 0 | 1 |
| 20 | 34 | 0 | 1 |
| 21 | 35 | 0 | 1 |
| 22 | 1 | 1 | 0 |
| 23 | 1 | 1 | 0 |
| 24 | 2 | 1 | 0 |
| 25 | 2 | 1 | 0 |
| 26 | 3 | 1 | 0 |
| 27 | 4 | 1 | 0 |
| 28 | 4 | 1 | 0 |
| 29 | 5 | 1 | 0 |
| 30 | 5 | 1 | 0 |
| 31 | 8 | 1 | 0 |
| 32 | 8 | 1 | 0 |
| 33 | 8 | 1 | 0 |
| 34 | 8 | 1 | 0 |
| 35 | 11 | 1 | 0 |
| 36 | 11 | 1 | 0 |
| 37 | 12 | 1 | 0 |
| 38 | 12 | 1 | 0 |
| 34 | 15 | 1 | 0 |
| 41 | 17 | 1 | 0 |

## ▣ منابع :

*1. Cox , D.R. and Oakes, D. Analysis of Survival Data. Chapman and Hall , 1994.*

*2. Kalbfeisch , J. and Prentice , R. Stat. Analysis of Failure time. J. Wiley and Sons , 1990.*

*3. Kleinbaum , D.G. Survival Analysis. Springer, 1996.*

*4. Lawless , J. Statistifcal Models methods for lifetime Data. J. Wiley and Sons, 1992.*

*5. Miller , R. Survival Analysis. J. Willey and Sons , 1990.*