معرفی و خواص سانسور نوع دوم پیشرو تعمیم‌یافته

آمنه‌سادات میرنیام*[1]؛ زهرا شناوری[2]؛ عبدالرسول برهانی حقیقی[3]

دریافت: 1392/10/09
پذیرش: 1394/03/03

چکیده

این مقاله ابتدا به مفهوم زمان شکست پرداخته و سپس سانسور و انواع آن، به خصوص طرح سانسور نوع دوم پیشرو تعمیم‌یافته را به همراه خواص آن در قالب قضایایی معرفی نموده است. همچنین مثالی برای روشن‌تر شدن این نوع طرح بیان گردیده است.

واژگان کلیدی: آنالیز بقا، زمان شکست، سانسور، آزمون عمر.

---

1. دانشجوی دکتری، آمار دانشگاه شیراز (***نویسنده مسئول**) bahareh104@gmail.com
3. دانشجوی دکتری، آمار دانشگاه شیراز zshenavari@yahoo.com
3. استادیار، بخش آمار دانشگاه شیراز borhani@susc.ac.ir

مقدمه

در آنالیز بقاء کلاسیک، به مطالعه گروهی از افراد پرداخته می‌شود که برای تک تک آنها یک رویداد لحظه‌ای، به نام شکست، تعریف می‌شود. مدت زمانی که طول می‌کشد تا این شکست رخ دهد، زمان شکست نامیده می‌شود. روی هر فرد شکست حداکثر یک بار می‌تواند رخ دهد. داده‌های بقاء در حقیقت زمان‌های شکست را از یک لحظه ورود تعریف شده تا وقوع یک رویداد نهایی اندازه‌گیری می‌کنند. تعیین دقیق زمان شکست مستلزم آن است که:

الف- معنی و مفهوم شکست کاملاً واضح باشد.

ب- یک مبدأ زمانی برای هر یک از افراد مورد مطالعه تعریف شده باشد.

ج- یک مقیاس برای اندازه‌گیری زمان تعیین شده باشد. (حیدری، 1377)

دانستن زمان شکست در آنالیز بقا برای ما حائز اهمیت است. گاهی این زمان‌ها را نمی‌توان مشاهده کرد چون به دلایلی سانسور می‌شوند. محققان بسیاری در مورد داده‌های سانسور شده تحقیق و بررسی کرده‌اند. از جمله سینها[1] (1986)، دیویس و فلدستین[2] (1979)، سن[3] (1979)، هالپرین و همکاران[4] (1989)، ویورس و بالاکریشنان[5] (1994) باندیوپا-دهیای و چاتوپادهیای[6] (1995)، ین وتسه[7] (1996) آگاروالا و بالاکریشنان[8] (1998) بالاکریشنان و همکاران[9] (2001)، گیلباد[10] (2001). در ادامه مفهوم سانسور و انواع آن را بررسی کرده و خواص آنها را نیز در قالب قضایایی بیان خواهیم کرد.

سانسور[11]

همان‌گونه که قبلاً اشاره شد در آنالیز بقاء کلاسیک گروهی از افراد را از یک لحظه ورود تعریف شده تا وقوع یک رویداد نهایی (شکست) مورد مطالعه قرار می‌دهند.

هدف از مطالعه داده‌های بقاء مشاهده و بررسی زمان‌های شکست برای افراد تحت مطالعه می‌باشد غالباً امکان‌پذیر نیست که صبر کنیم تا اینکه این رویداد برای همهٔ افراد اتفاق بیفتد. برای بعضی، فقط می‌دانیم که این رویداد هنوز در یک زمان خاص رخ نداده است، یعنی زمان شکست هنوز مشاهده نشده است. در این حالت مشاهده زمان شکست سانسور شده است. به عبارت دیگر زمانی را که در آن هنوز شکست برای افراد مشاهده نشده، زمان سانسور آن فرد می‌نامند. مثلاً در آزمون عمر در قابلیت اعتماد صنعتی، بعضی از قطعات (مولفه‌ها) ممکن است از کار نیفتند، یعنی شکست برای آنها مشاهده نشود.

اساساً داده‌های زمانی را سانسور شده گوییم هرگاه مولفه‌هایی در نمونه وجود داشته باشند که تنها یک کران پائین (یا بالا) روی طول عمر آنها در دسترس باشد.

دانستن علت سانسور در مسائل، برای تحلیل داده‌های بقاء دارای اهمیت است. یادآور می‌شویم که همانند شکست، سانسور نیز یک رویداد لحظه‌ای است و زمان مشاهده برای افراد سانسور شده (زمان سانسور) بایستی ثبت گردد.

انواع داده‌های سانسور شده طول عمر

برای جمع‌آوری داده‌های طول عمر دو روش وجود دارد:

1- **روش اوّل** به این صورت است که قبل از شروع آزمایش زمان t_0 را مشخص کنیم و فقط طول عمر مولفه‌هایی را که تا زمان t_0 شکست خورده‌اند یادداشت کنیم این روش را زمان- سانسور می‌نامند، زیرا زمان قطعی آزمایش از ابتدا مشخص شده است[12].

1. Sinha
2. Davis & Feldstain
3. Sen
4. Halperin et al.
5. Viveros & Balakrishnan
6. Bandyopadhyay & Chatopadhyay
7. Yuen & Tse
8. Aggarwala & Balakrishnan
9. Balakrishnan et al.
10. Guilbaud
11. Censoring
12. Time Censoring

2- **روش دوم** به این صورت است که قبل از شروع آزمایش یک عدد صحیح کوچک‌تر از n (n تعداد کل مولفه‌ها) را معین کنیم و آن را r بنامیم و آزمایش را تا زمانی که طول عمر r تا مولفه اول به پایان برسد ادامه دهیم و دقیقاً در زمان $T_{(r)}$ یعنی طول عمر r امین مولفه‌ای که شکست خورده، آزمایش را قطع می‌کنیم. این روش را شکست سانسور می‌نامیم زیرا تعداد شکست‌های ثبت شده از ابتدا مشخص شده است[1].

همچنین اگر همه مولفه‌ها آزمایش را با هم شروع کنند، داده‌ها را سانسور منفرد[2] و اگر مولفه‌ها در زمان‌های مختلفی از بازه زمانی شروع به فعالیت کنند، و آزمایش در زمان t_0 قطع شود داده‌ها را سانسور چندگانه[3] می‌نامند.

سانسور نوع اول[4]

به داده‌های منفرد زمان- سانسور[5]، داده‌های سانسور شده نوع اول می‌گویند.

سانسور نوع دوم[6]

به داده‌های منفرد شکست سانسور[7] داده‌های سانسور شده نوع دوم می‌گویند.

سانسور نوع دوم پیشرو[8]

سانسور نوع دوم پیشرو، طرح مفید و عمومی‌تری است که در آن به آزمایشگر اجازه می‌دهد تا تعداد مشخصی از مولفه‌ها را در هر یک از زمان‌های متعدد شکست از مطالعه خارج کند (یعنی سانسور کند). حذف کردن‌های متوالی از این نوع به دلایل مختلفی طراحی می‌شوند. صرفه‌جویی در زمان و هزینه می‌تواند نتیجه چنین طرح نمونه‌گیری باشد. این نوع نمونه‌های سانسور شده توسط دیویس و فلدستین (1979)، سن (1979)، هالپرین و همکاران (1989)، ویورس و بالاکریشنان (1994) باندیوپادهیای و چاتوپادهیای (1995)، ین وتسه (1996) آگاروالاو بالاکریشنان (1998) بالاکریشنان و همکاران (2001) وگیلباد (2001) در نظر گرفته شده است.

این طرح به‌صورت زیر است:

1- یک گروه n تایی از واحدهای آزمایشی در زمان صفر در یک آزمایش ***آزمون عمر*** قرار داده می‌شوند.

2- به‌طور ناگهانی پس از مشاهده اولین شکست (مرگ)، یک تعداد از پیش تعیین شده $R_1 \geq 0$ از $n-1$ واحد باقی‌مانده به تصادف انتخاب و از آزمایش خارج می‌شوند. بنابراین تنها $n-1-R_1$ واحد قابل مشاهده در آزمایش باقی می‌ماند.

3- به‌طور ناگهانی پس از مشاهده دومین شکست (مرگ)، یک تعداد از پیش تعیین‌شده $R_2 \geq 0$ از $n-1-R_1-1$ واحد باقی‌مانده به تصادف از آزمایش خارج می‌شوند. بنابراین تنها $n-2-(R_1+R_2) = n-1-R_1-1-R_2$ واحد قابل مشاهده در آزمایش باقی می‌مانند و به همین ترتیب ...

بنابراین واحدهای خارج شده، در لحظه شکست (مرگ) واحدهای دیگر، از راست سانسور می‌شوند، یعنی طول عمر آنها کران بالا ندارد. مقادیر R_1 و R_2 و ... و R_m، یعنی همان واحدهای خارج شده، طرح سانسور را تعیین می‌کنند.

سانسور نوع دوم پیشرو تعمیم‌یافته[9]

بالاکریشنان و سند هو[10] (1995)، علاوه بر مدل سانسور پیشرو، مدل سانسور از سمت چپ[11] داده‌ها را در نظر گرفتند که سانسور نوع دوم پیشرو تعمیم‌یافته نامیده می‌شود. آنها فرض کردند که:

1- در زمان $x_0 = 0$، n واحد در آزمون قرار داده می‌شوند.

1. Failure Censoring
2. Single Censoring
3. Multiple Censoring
4. TypeI Censoring
5. Single Time Censoring
6. Type-II Censoring (Right Censoring)
7. Single Failure – Censoring
8. Progressive Type-II Censoring
9. General Progressive Type-II Censoring
10. Balakrishnan & Sandhu,
11. Left Censoring

2- اگر X_i زمان i امین شکست مرتب‌شده باشد، $X_1 , \dots , X_r$ مشاهده نمی‌شوند.

3- در زمان X_i، $(i = r + 1, \dots , m - 1)$، R_i واحد به‌طور تصادفی از آزمون خارج می‌شوند. بنابراین پیش از $(i + 1)$ امین شکست، $n_i = n - i - \sum_{j=r+1}^{i} R_j$ واحد در آزمون وجود دارند.

4- در زمان m امین شکست، X_m، آزمایش خاتمه می‌یابد که R_m واحد باقیمانده از آزمون خارج می‌شوند.

توضیح بیشتر و مثال:

فرض کنیم $X_0 = 0$ و $r = 3$ یعنی زمان‌های شکست X_1, X_2, X_3 مشاهده نمی‌شوند. در زمان X_4 که $X_1 \le X_2 \le X_3 \le X_4$، R_4 واحد به‌طور تصادفی از آزمون خارج می‌شوند بنابراین پیش از مشاهده پنجمین شکست، تعداد واحدهای موجود در آزمون عبارتند از:

$$n_4 = n - 3 - 1 - \sum_{j=4}^{4} R_j = n - 4 - R_4$$

اگر $m = 7$ باشد با مشاهده زمان X_6، R_6واحد به‌طور تصادفی از آزمون خارج می‌شوند و تعداد باقیمانده پیش از مشاهده شکست هفتم عبارت است از:

$$n_6 = n - 6 - (R_4 + R_5 + R_6).$$

با مشاهده X_7، R_7 واحد باقیمانده از آزمون خارج شده و آزمایش خاتمه می‌یابد.

شرایط:

R_i ها، r , m اعداد صحیح از پیش تعیین‌شده‌ای هستند که باید در شرایط زیر صدق کنند:

1- $0 \le r < m \le n$

2- $R_m = n_{m-1} - 1$ و $i = r + 1 , \dots , m - 1$ و $n_r = n - r$ و $0 \le R_i < n_{i-1} - 1$

برای راحتی نوشتاری فرض می‌کنیم $\vec{x} = (x_{r+1}, \dots , x_m)$ و $\vec{R} = (R_{r+1}, \dots , R_m)$. در فصل بعدی یک طرح سانسور نوع دوم پیشرو تعمیم‌یافته و زمان‌های شکستی که دارای توزیع نمایی هستند ارائه می‌گردد.

خواص آماره‌های ترتیبی سانسور شده پیشرو تعمیم‌یافته

قضیه 1: تحت شرایط بالا داریم:

الف) توزیع X_{r+i} مستقل از $R_{r+i}, R_{r+i+1}, \dots , R_m$ است. $(i = 1, \dots , m - r)$

ب) $X_{r+1}, X_{r+2}, \dots , X_{r+i}$ یک نمونه سانسور شده نوع دوم پیشرو تعمیم‌یافته، به اندازه i از n واحد موجود در آزمون با طرح سانسور r و $(R_{r+1}, \dots , R_{r+i-1})$ و $n - (R_{r+1}, \dots , R_{r+i-1}) - (r + i)$

اثبات: در منبع (2).

نکته 1: این قضیه در اثبات برخی روابط بازگشتی برای گشتاورهای آماره‌های ترتیبی سانسور شده نوع دوم پیشرو مفید خواهد بود.

قضیه 2: آماره‌های ترتیبی سانسور شده نوع دوم پیشرو تعمیم‌یافته از یک توزیع پیوسته دلخواه، یک زنجیره مارکف تشکیل می‌دهند.

اثبات: هدف ما این است که نشان دهیم:

$$f_{X_{r+i-1}|X_{r+i},X_{r+i-1},\dots,X_{r+1}}(x_{r+i-1}|x_{r+i}, x_{r+i-1}, \dots , x_{r+1}) = f_{X_{r+i-1}|X_{r+i}}(x_{r+i-1}|x_{r+i}) \qquad (1)$$

سمت چپ تساوی (1)، می‌شود:

$$(n - r - R_{r+1} - \dots - R_{r+i} - i) \left[\frac{1-F(x_{r+i-1})}{1-F(x_{r+i})}\right]^{n-r-R_{i+1}-\dots-R_{r+i}-i-1} \frac{f(x_{r+i})}{[1-F(x_{r+i})]} = f_{X_{r+i-1}|X_{r+i}}(x_{r+i-1}|x_{r+i})$$

بنابراین آماره‌های ترتیبی سانسور شده نوع دوم پیشرو تعمیم‌یافته، یک زنجیره مارکف را تشکیل می‌دهند.

قضیه 3: $X_{r+i+1}, \dots , X_m$، $(i \ge 1)$، به شرط $X_{r+i} = X_{r+j}$، توأماً بصورت یک نمونه سانسور شده از راست، نوع دوم پیشرو به اندازه $m - r - i$ از $n - R_{r+1} - \dots - R_{r+i} - r - i$ متغیر تصادفی هم‌توزیع از چگالی $f(.)$، که از سمت چپ در X_{r+j} بریده شده[1] توزیع می‌شوند یعنی با چگالی

1. left – truncated

$f(x)/(1-F(x_{r+i}))$ و با طرح سانسور پیشرو $R_{r+i+1},\dots,R_m$،

اثبات: در منبع (2).

نتیجه‌گیری

در این مقاله ابتدا داده‌های سانسور شده طول عمر معرفی گردیده ولی تاکید اصلی بر روی سانسور نوع دوم پیشرو تعمیم‌یافته می‌باشد. همچنین خواص این نوع سانسور در قالب قضایایی همراه یک مثال مورد مطالعه قرار گرفته است.

به‌عنوان مثال نشان داده شده که این داده‌ها از هر توزیعی که باشند دارای خاصیت مارکف هستند و این خاصیت ما را در یافتن چگالی توأم شرطی بخشی از داده‌های سانسور شده یاری می‌دهد.

منابع

حیدری، فریبرز. (1377). تابع راست‌نمایی برای مدل‌های سانسور معمولی در آنالیز بقاء، *گلچین ریاضی*، جلد پنجم، شماره 1.

Aggarwala, R. and Balakrishnan, N. (1998). Some properties of progressive censored order statistics from arbitrary and uniform distribiutions with applications to inference and simulation. Journal of statistical planning and inference. 70, 35-49.

Balakrishnan, N. and Sandhu, R.A. (1995). A simple simulational algorithm for generating progressive type II censored samples. The American Statistician, Vol. 49, No.2. pp.229-230.

Balakrishnan, N., Cramer, E., kamps, U. and sckenk, N. (2001). Progressive type II censored order statistics from exponential distributions. Statistics 35, 537-556.

Bandyopadhyay, U. and Chattopadhyay, G. (1995). Progressive censoring under inverse sampling for nonparametric two-sample problems. Sequential Anal. 14, 1-28.

Davis, H.T. and Feldstein, M.L. (1979). The generalized pareto law as a model for progressively censored survival data. Biometrica 66, 299-306.

Guilbaud, O. (2001). Exact non-parametric confidence intervals for quantiles with progressive type II censoring. Scand. J. statis. 28, 699-713.

Halperin, M., Hamdy, M.I. and Thall, P.F. (1989). Distribution - free confiedence interval for a parameter of Wilcoxon - Mann - Whitney type for ordered categories and progressive censoring. Biometrics 45, 509-521.

Sen, P.K. (1979). Weak convergence of some quantile processes arising in progressively censored tests. Ann. Statist. 7, 414-431.

Sinha, S.K. (1986). Reliability and life testing. Wiley Eatern Ltd, India :Dehli.

Viveros, R. and Balakrishnan, N. (1994). Interval estimation of life characteristics from progressively censored samples. Technometrics 36, 84-91.

Yuen, H.K. and Tse, S.K. (1996). Parametersestimation for Weibull distributed lifetimes under progressive censoring with random removals. J. Statistics. Comput. Simulation 55, 57-71.