مربوط به متغیرهای تصادفی مورد نظر است.
فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی باشد، چنان چه آن‌ها را به صورت صعودی مرتب کنیم یعنی $(X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n})$، آنگاه این مقادیر مرتب شده را آماره‌های ترتیبی گوییم که کاربرد وسیعی در آمار، از جمله آزمون‌های طول عمر، آزمایش‌های کلینیکی و آنالیز بقا دارند. فرض کنید در یک آزمون طول عمر با n واحد، به دلایلی باید آزمایش را تا قبل از شکست خوردن همهٔ واحدها پایان دهیم. در این صورت طول عمر تعدادی از مشاهدات به صورت ناقص در اختیار ما قرار می‌گیرند، در این صورت می‌گوییم این مشاهدات سانسور شده‌اند.
پیشینه و منشاء تحلیل بقا کارهایی است که در گذشته در مورد جداول طول عمر انجام شده است. شکل جدید تحلیل بقا از نیم قرن گذشته با کاربردهای مهندسی مورد توجه و مطالعه قرار گرفته است. در جنگ جهانی دوم، علاقه بسیاری به قابلیت اعتماد تجهیزات جنگی بوجود آمد. این علاقه به قابلیت اعتماد، موضوع را به کارهای نظامی و فرآورده‌های تجاری کشاند. قبلاً بیشتر تحقیقات آماری مورد استفاده در مهندسی بر الگوهای پارامتری متمرکز بود. در دو دهه اخیر تحقیقات آزمایشگاهی و پزشکی افزایش یافته است. در این آزمایشها بیشتر به روش‌های ناپارامتری توجه شده است.

۲—۳ داده‌های سانسور شده

در تحلیل داده های بقا این امکان وجود دارد که زمان شکست کامل برخی از واحدها مشاهده نشود. به این معنی که در برخی وضعیت ها، برای برخی از واحدها در طول آزمایش شکست اتفاق نمی افتد و یا از ادامه آزمایش باز نمی مانند و به جای دانستن زمان شکست (t)، تمام آن چه می دانیم این است که این واحدها طول عمری متجاوز از مقداری مانند (y) دارند. در آمار سانسور زمانی رخ می دهد که مقدار یک مشاهده به طور ناقص معلوم باشد و یا هنگامی که یک مقدار خارج از حیطه اندازه گیری روی دهد. به بیان دیگر اگر یک متغیر وابستهٔ دلخواه، زمان پایان رویداد را نشان دهد و طول مدت مطالعه محدود باشد، در زمان پایان دوره مطالعه برخی از واحدها ممکن است هم چنان برقرار باشند. به عنوان مثال، در علوم اجتماعی دوامِ ازدواج، در مقاطع تحصیلی زمان تا افتِ تحصیلی،... را می توان مطالعه کرد. در پایان مطالعه این امکان وجود دارد برخی اشیاء تحت مطالعه هنوز در تأهل باشند و یا افت تحصیلی نکرده باشند و... بنابراین این اشیاء نمایانگر مشاهدات سانسور شده هستند.

فرض کنید n واحد را تا زمان خرابی آن‌ها تحت نظر داریم. این واحدها می‌توانند برخی از سیستم‌ها، مؤلفه‌ها یا تراشه‌های رایانه‌ای در آزمایش‌های مطالعه قابلیت اعتماد، یا بیمارانی باشند که تحت مصرف داروی معینی قرار می‌گیرند. فرض کنید طول عمر این n واحد، متغیرهای تصادفی مستقل و هم توزیع با تابع توزیع پیوسته $F(t)$ و تابع چگالی احتمال $f(t)$ باشند. چون مقادیر نمونه بر حسب افزایش کمیت یادداشت شده‌اند مشاهدات به طور طبیعی به صورت بردار آماره‌های مرتب ظاهر می‌شوند. حال اگر بنا به دلایلی آزمایش را متوقف سازیم، در این صورت یک نمونه سانسور شده خواهیم داشت که در آن آماره‌های مرتب نقش مهمی را ایفا می‌کنند. هر چند با سانسور کردن مشاهدات، مقداری از اطلاعات موجود را از دست خواهیم داد، با وجود این گاهی اوقات به دلایلی مایل (یا مجبور) به سانسور کردن مشاهدات هستیم.

یکی از اولین تلاش‌ها در جهت تحلیل مسائل آماری که شامل داده‌های سانسور شده بود توسط دانیل برنولی (۱۷۶۶) انجام گرفت که در تحلیل شیوع بیماری آبله و اطلاعات مربوط به مرگ و میر، در اثر این بیماری بود.

۲–۴ انواع سانسورها

۲–۴–۱ سانسور نوع اول

وضعیتی را در نظر بگیرید که در آن n واحد به طور تصادفی، از یک جامعه، مورد آزمون قرار گرفته‌اند. هم چنین فرض کنید زمان پایان آزمایش T، پیش از مطالعه تعیین شده است و واحدها در زمان $t = ۰$ فعال شده‌اند. هدف، دستیابی به طول عمر واحدها در طول مدت آزمایش می‌باشد. اگر در این فاصله r شکست را مشاهده کنیم که $۰ \leq r \leq n$ و زمان‌های شکست این r واحد عبارت باشند از: $t_{۱:n}, t_{۲:n}, ..., t_{r:n}$، در این صورت $(n - r)$ واحد وجود دارند که در زمان T هم چنان فعالند (بدون شکست)، پس آزمایش در زمان T پایان می‌پذیرد به طوری که در این زمان $(n - r)$ واحد باقی مانده سانسور شده‌اند.

نکته ۲—۱.۴ : این نوع سانسور را سانسور از راست نیز می‌نامند زیرا زمان‌های شکست سمت راست (بزرگتر از T) نامعلوم می‌باشند.

مثال ۲—۱.۴ : در یک آزمایش در مرکز بین المللی تحقیقات سم شناسی، موش‌ها بوسیلهٔ دوز مشخصی از کارسینوژن تغذیه می‌شوند. هدف از این آزمایش بررسی تأثیر کارسینوژن، بر بقای موش‌ها می‌باشد. موش‌ها از شروع مطالعه تا زمان سانسور از پیش تعیین شده، مورد مطالعه قرار می‌گیرند. چون این احتمال وجود دارد که موش‌ها مدت زمان طولانی‌تراز زمان، از پیش تعیین شده عمر کنند، در این صورت یک نمونه سانسور شده نوع یک خواهیم داشت.

- ایراد وارد بر سانسور نوع اول این است که این احتمال وجود دارد که تا زمان سانسور از پیش تعیین شده هیچ و یا تعداد کمی از واحدها به زمان شکست خود برسند و طول عمر دقیق آنها مشخص شود و این باعث پایین آمدن کارایی در تحلیل داده ها می شود.

۲—۴—۲ سانسور نوع دوم

نوع دیگری از مطالعه، وضعیتی است که قبل از مطالعه بخواهیم دقیقاً r شکست را مشاهده کنیم و آزمون را تا مشاهده r امین شکست ادامه دهیم، در واقع r از پیش تعیین شده و آزمون در زمان $T = t_r$ پایان می‌پذیرد و در لحظه مشاهده r امین شکست $(n - r)$ واحد هم چنان فعالند که سانسور شده‌اند. در این سانسور واضح است که زمان لازم برای رسیدن به r امین شکست، یک متغیر تصادفی است.

مثال ۲—۲.۴ : آزمایشی به منظور تعیین طول عمر لامپ های تولید شده در کارخانه ای انجام گرفته است. n لامپ به تصادف انتخاب و در یک زمان مشخص روشن می شوند. مسلماً انتظار تا سوختن تمام لامپ ها ممکن است طولانی شود. بنابراین مطالعه را با سوختن r لامپ اول خاتمه داده و نتایج گزارش می شود.

- سانسور نوع دوم این مزیت را داراست که از قبل می دانیم آزمون دارای چند شکست است، اما این احتمال وجود دارد که آزمایش تا مشاهدهٔ r امین شکست بسیار طولانی شود که از نقطه نظر مدیریتی ضعف به شمار می آید.

۲—۴—۳ سانسور از راست

در یک آزمایش بر روی تعدادی از واحدها، این امکان وجود دارد که قبل از اینکه تمام واحدها شکست بخورند، آزمایش به پایان برسد. لذا تمام اطلاعی که راجع به بعضی از واحدها داریم این است که می‌دانیم آنها دارای طول عمری هستند که از طول عمر آزمون، تجاوز کرده‌اند. چنین مشاهداتی به داده‌های از راست سانسور شده موسومند و به شکل $X_i > X_c$ هستند. که X_c نشانگر زمان سانسور و X_i ها معرف طول عمر واحدها هستند. در واقع اگر متغیر مورد مطالعه بسیار بزرگ باشد و نخواهیم آن را به طور کامل مشاهده کنیم، از این نوع سانسور استفاده می‌کنیم.
همان طور که گفته شد، سانسور از راست [۲] همان سانسور نوع اول است که گاهی اوقات تحت این عنوان، در مقابل سانسور از چپ قرار می‌گیرد.

مثال ۲—۴.۳ : یک روان پزشک می‌خواهد سن یادگیری گروه خاصی از کودکان آفریقایی را برای انجام یک عمل ویژه بداند. وقتی به روستای مورد نظر می‌رسد، بعضی از کودکان یادگیری را شروع می‌کنند و سن یادگیری آن‌ها را می‌توان ثبت کرد. هنگام بازگشت پژوهش گر، هنوز برخی از کودکان روستایی این آموزش را یاد نگرفته‌اند. آن‌ها مشاهدات سانسور شده راست را نتیجه می‌دهند.

۲—۴—۴ سانسور از چپ

طول عمر X مربوط به یک واحد معین در یک مطالعه، سانسور شده از چپ [۳] گویند، هرگاه کمتر از زمان سانسور C_L باشد، یعنی پیشامد مطلوب برای واحد پیش از اینکه واحد در زمان C_L مشاهده شود، رخ

[۲] Right Censoring
[۳] left Censoring

می‌دهد. این واحدها، پیشامد را قبل از زمان C_L (زمان سانسور از چپ) تجربه کرده‌اند و زمان پیشامد واقعیشان نامعلوم است. طول عمر واقعی X معلوم است اگر و تنها اگر X بزرگتر یا مساوی C_L باشد.

مثال ۲—۴.۴ : در یک مطالعه از ۱۹۱ دانش آموز دبیرستان، پرسیده می‌شود چه وقت برای اولین بار از سیگار استفاده کرده‌اند؟ حال اگر فردی این ماده مخدر را مصرف کرده ولی زمان اولین مصرف را به یاد نیاورد آنگاه زمان اولین مصرف پیش از ورود به مطالعه بوده و زمان دقیق آن مشخص نیست، پس به عنوان مشاهده‌ای سانسور شده از چپ در نظر گرفته می‌شود.

۲—۴—۵ سانسور فاصله ای

سانسور فاصله ای[۴] هنگامی رخ می دهد که زمان شکست (رخداد نهایی) به طور دقیق مشاهده نشده است و این زمان در یک فاصله معلوم شده است. در واقع برای هر واحد i ام به جای مشاهده دقیق زمان بقاء X^i، فاصله $[X_L^i, X_R^i]$ را که شامل X^i است، مشاهده می کنیم. یک مطالعه که بیماران به صورت دوره ای کنترل می شوند و پیشامد مطلوب تنها در زمان مشخصی دست یافتنی است، نمودی از سانسور فاصله ای می باشد.

مثال ۲—۵.۴ : در مطالعات قلبی، سن وقوع آنژین پرینزمتال در بین دو معاینه بالینی، تقریباً به فاصله دو سال از هم معلوم می باشد. این چنین مشاهداتی سانسور فاصله ای هستند. و یا آزمایشی را که در آن یک واحد تحت آزمون قرار می‌گیرد و هر ساعت برای خرابی بازدید می‌شود. اگر واحد مذکور بین زمان‌های دوم و سوم خراب شده باشد (یعنی واحد در زمان بازدید دوم هنوز مشغول به کار بوده ولی در زمان بازدید سوم خراب باشد)، می‌توان دریافت که طول عمر آن واحد بین دو تا سه ساعت است.

[۴] Interval Censoring

نکته ۲—۲.۴ : سانسور راست و چپ حالت خاصی از سانسور فاصله‌ای می باشند. بدین گونه که اگر X^i سانسور راست باشد، مشاهدات X^i در بازه $[X_R^i, +\infty)$ و اگر X^i سانسور چپ باشد، در بازه $(-\infty, X_L^i]$ قرار می گیرند.

یکی از اولین مقالات که در زمینه سانسور فاصله ای ارائه شده است مربوط به پتو[۵] در سال ۱۹۷۳ می باشد.

۲—۴—۶ سانسور تصادفی

در سانسور تصادفی، زمان کل آزمایش ثابت است اما واحدها در زمان‌های مختلف، وارد آزمایش می‌شوند. برخی از واحدها شکست می‌خورند، برخی بنا به دلایلی غیر از پیشامد مطلوب نظیر مرگ تصادفی (یعنی مرگ در اثر عاملی غیر از پیشامد مورد نظر)، عدم میل به شرکت در ادامه مطالعه و ... از ادامه آزمایش باز می‌مانند و تعدادی هم در پایان آزمایش هم چنان برقرارند.

در سانسور تصادفی، هر واحد زمان شکست (X_i) و زمان سانسور (C_i) مخصوص به خود را دارد که این دو از هم مستقلند.

بنابراین می توان بردار تصادفی (Y_i, δ_i) را مشاهده کرد به طوری که

$$\delta_i = \begin{cases} 1, & X_i \le C_i \text{ اگر } \quad \text{(بدون سانسور)} \\ 0, & X_i > C_i \text{ اگر } \quad \text{(با سانسور)} \end{cases}$$

و

$$Y_i = \min(X_i, C_i).$$

که در آن δ_i برای نشان دادن اطلاعات مربوط به سانسور به کار رفته است.

نکته ۲—۳.۴ : هنگامی که $C_i = T$ یعنی C_i ها ثابت باشند، داده های سانسور شده نوع اول را خواهیم داشت.

سانسور تصادفی در مطالعات پزشکی، اپیدمیولوژی و مهندسی کاربرد فراوان دارد.

[۵] Peto

مثال ۲—۶.۴ : در یک بررسی طول عمر، در اثر مصرف دارو در بیماران مبتلا به نارسایی کلیوی بیماران برای معالجه در زمان های مختلف مراجعه می کنند حال عمل سانسور به یکی از صورت های زیر انجام می گیرد.

۱) عدم مراجعه: ممکن است بیمار تصمیم بگیرد به پزشک دیگری مراجعه کند. در نتیجه، از ادامه آزمایش باز می ماند.

۲) قطع معالجه: ممکن است به علت عوارض بد جانبی، درمان بیمار را متوقف کنیم، یا این که ممکن است بیمار هنوز در تماس باشد ولی از ادامه معالجه خودداری کند.

در این حالت فرض کنید سه بیمار داریم، مثلاً بیمار (۱) در زمان $T = \circ$ مورد بررسی قرار گرفته و در زمان $X_۱$ فوت شده است. در نتیجه، یک مشاهده سانسور نشده بدست آمده است. ولی بیمار (۲) مورد بررسی قرار گرفته و در پایان بررسی، هنوز زنده است. پس یک مشاهده سانسور شده $X_۲^+$ به دست آمده است. بالاخره بیمار (۳) مورد بررسی قرار گرفته و قبل از پایان بررسی، معالجه را قطع نموده است در نتیجه، یک مشاهده سانسور شده $X_۳^+$ به دست آمده است.

indent در سانسور تصادفی معقول به نظر می‌رسد که فرض کنیم مطالعه در زمان‌های تصادفی شروع و قطع می‌شوند. به بیان دیگر، در سانسور تصادفی اجزاء در زمان های مختلف وارد مطالعه می شوند برخی از این اجزاء شکست می خورند، برخی از ادامه آزمایش در می مانند و برخی در پایان مطالعه هم چنان برقرارند. در سانسور تصادفی هر جزء زمان شکست T_i و زمان سانسور C_i مخصوص به خود را داراست، که این دو زمان از یکدیگر مستقل می باشند$(T_i \perp C_i)$ یعنی زمان سانسور C_i در بررسی و تحلیل زمان شکستِ T_i فاقد اطلاع است. برای داشتن چنین فرضی باید اطمینان داشته باشیم که اجزاء بنا به دلایلی وابسته به زمان شکست، از ادامه آزمایش باز نمی مانند.

۲—۴—۷ سانسور هیبرید

مخلوطی از طرح‌های سانسور نوع یک و دو به طرح سانسور هیبرید معروف است. این نوع سانسور بر دو نوع است:

الف) سانسور هیبرید نوع یک: در یک مطالعه که در آن n واحد تحت آزمون قرار گرفته‌اند به طوری که در آن، $T \in (\circ, \infty)$ طول آزمایش و $۱ \leq r \leq n$، تعداد واحدهای شکست خورده می‌باشد که از قبل تعیین شده‌اند. آزمایش در زمان تصادفی $T_۱^* = \min(X_{r:n}, T)$ پایان می‌پذیرد. مشاهدات عبارتند از

$$X_{1:n} < X_{2:n} < ... < X_{r:n}, \quad X_{r:n} < T \text{ اگر}$$

$$X_{1:n} < X_{2:n} < ... < X_{D:n}, \quad X_{r:n} > T \text{ اگر}$$

در رابطه فوق $D \leq r$ که D تعداد شکستها تا رسیدن به زمان سانسور می باشد. و علت نامیدن این نوع سانسور به سانسور هیبرید نوع یک، پایان یافتن آزمایش حداکثر در زمان T است. طرح سانسور هیبرید اولین بار توسط اپستین [۶] در سال های ۱۹۵۴ و ۱۹۶۰ معرفی شد. این طرح توسط چایلدز[۷] و همکارانش در سال ۲۰۰۲، طرح سانسور هیبرید نوع یک نام گرفت.

برای جزئیات بیشتر می توانید به عملکرد چن و باتاچاریا[۸] (۱۹۹۸) و دراپر و گاتمن[۹] (۱۹۸۷) و نیز گوپتا و کوندا[۱۰] (۱۹۹۸) مراجعه نمایید.

ب) سانسور هیبرید نوع دو: سانسور هیبرید نوع یک این عیب را داراست که ممکن است تا زمان از پیش تعیین شده T تعداد بسیار کمی شکست رخ دهد. برای رفع این عیب سانسور هیبرید نوع دو مطرح شده است. به این ترتیب در آزمایشی از n واحد که در آن مقادیر T و r از قبل تعیین شده‌اند، آزمایش در زمان تصادفی $T_2^* = \max(X_{r:n}, T)$ پایان می‌پذیرد. مزیت این طرح مشاهده حداقل r شکست است.

$$X_{1:n} < X_{2:n} < ... < X_{r:n}, \quad X_{r:n} > T \text{ اگر}$$

$$X_{1:n} < X_{2:n} < ... < X_{D:n}, \quad X_{r:n} < T \text{ اگر}$$

در رابطه فوق $D \geq r$.

[۶] Epstein
[۷] Childs
[۸] Chen & Bhattacharya
[۹] Draper & Guttman
[۱۰] Kundu & Gupta

۲—۴—۸ سانسور پیشرونده

الف) سانسور پیشرونده نوع یک[11]: فرض کنید در یک آزمایش مربوط به طول عمر، نمونه‌ای به حجم n در زمان $t = ۰$ فعال شود و در زمان‌های $T_۱, T_۲, ..., T_{m-۱}$ که ثابت فرض شده‌اند به ترتیب $R_۱, R_۲, ..., R_{m-۱}$ که (مقدار آنها از قبل معلوم است) به طور تصادفی از بین واحدهای باقیمانده از آزمایش خارج شوند یا به عبارت دیگر سانسور شوند. در این صورت تعداد واحدهایی که تا رسیدن به زمان T_i از کار می‌افتند یک متغیر تصادفی است و آزمایش در زمان T_m با R_m واحد باقیمانده پایان می‌پذیرد.

مثال ۲—۴.۷ : تعداد n موش به تصادف به k گروه به حجم‌های n_i، تقسیم شده‌اند که در آن $n = \sum_{i=۱}^{k} n_i, i = ۱, ۲, ..., k$. تمام گروه ها در یک زمان وارد مطالعه می‌شوند اما زمان سانسور هر گروه C_{ri} متفاوت از دیگر گروه هاست. با سپری شدن زمان سانسور تعیین شده برای هر گروه، مطالعه بر روی آن گروه خاتمه می‌یابد و داده‌های مرگ و سانسور گروه ثبت می‌شود. که به این نوع سانسور، سانسور پیشرونده نوع یک می‌گویند.

ب) سانسور پیشرونده نوع دو[12]: یکی از ایرادات وارد بر سانسور نوع یک و دو و سانسور هیبرید این است که امکان خروج واحدها به جز نقطه نهایی آزمایش وجود ندارد، اما سانسور پیشرونده نوع دوم این مزیت را داراست. طرح سانسور پیشرونده نوع دو را می‌توان به صورت زیر تعریف کرد:
n واحد را که تحت یک مطالعه قرار گرفتند در نظر می‌گیریم و فرض می‌کنیم $m < n$ قبل از مطالعه از پیش تعیین شده‌اند. علاوه بر این، $R_m, ..., R_۲, R_۱$ قبل از آزمایش معلوم شده‌اند به طوری که

$$R_۱ + R_۲ + ... + R_m + m = n$$

در زمان اولین شکست، یعنی $X_{۱:m:n}$، $R_۱$ واحد از واحدهای باقیمانده را به طور تصادفی خارج می‌کنیم. به طور مشابه، در زمان دومین شکست، یعنی $X_{۲:m:n}$، $R_۲$ واحد از واحدهای باقیمانده و به همین ترتیب در لحظه mامین شکست، $X_{m:m:n}$، بقیه R_m واحد را بر می‌داریم.

[11] Progressive Censoring Type-I
[12] Progressive Censoring Type-II

نکته ۲—۴.۴ : در حالت خاص، یعنی $R_۱ = R_۲ = ... = R_{m-۱} = ۰$ طرح سانسور پیشرونده نوع دو همان سانسور نوع دو می‌باشد. یکی از ایرادات وارد بر سانسور پیشرونده نوع دو، امکان طولانی شدن زمان آزمایش است. برای اطلاعات بیشتر در مورد این سانسور به کتاب بالاکریشنان و آگاروالا[۱۳] (۲۰۰۰) رجوع کنید.

۲—۴—۹ سانسور میانی

هرگاه پیشامدی داخل یک بازه تصادفی (که این بازه می‌تواند در طول تحقیق قرار گیرد) رخ دهد، و زمان دقیق آن قابل مشاهده نباشد، در این هنگام سانسور میانی[۱۴] روی می‌دهد. به بیان دیگر، n شیء یکسان را در نظر می‌گیریم که تحت یک آزمون قرار گرفته‌اند و طول عمر این اشیاء $T_۱, T_۲, ..., T_n$ می‌باشند. برای i امین شیء، یک فاصله سانسور شده تصادفی (L_i, R_i) وجود دارد به طوری که T_i قابل مشاهده است تنها اگر $T_i \notin [L_i, R_i]$ در غیر این صورت سانسور می‌شود. بنابراین متغیر نشانگر δ_i را به صورت زیر تعریف می کنیم:

$$\delta_i = \begin{cases} ۱ & \text{اگر} \quad T_i \notin [L_i, R_i] \\ ۰ & \text{سایر نقاط} \end{cases}$$

توجه می شود که هر گاه $\delta_i = ۱$ است، متغیر T_i مشاهده می شود، و فاصله (L_i, R_i) مشاهده نمی شود. از طرفی دیگر، هنگامی که $\delta_i = ۰$، تنها فاصلهٔ سانسور شده $[L_i, R_i]$ را مشاهده می کنیم، که در این صورت سانسور رخ داده است. پس برای واحد i ام خواهیم داشت:

$$(Y_i, \delta_i) = \begin{cases} (T_i, ۱) & \text{اگر} \quad T_i \notin [L_i, R_i] \\ ([L_i, R_i], ۰) & \text{سایر نقاط} \end{cases}$$

[۱۳] Balakrishnan and Aggarwala

[۱۴] Middle Censoring