

# datos censurados – censored data

Authored by  
**memjavad**

November 13, 2025

## RECOMMENDED CITATION

memjavad (2025). *datos censurados – censored data*. Spanish Psychological Databases.  
Retrieved from <https://spanish.arabpsychology.com/?p=4158>

## Datos Censurados

**Primary Disciplinary Field(s):** Estadística, Bioestadística, Ingeniería de Fiabilidad, Econometría

### 1. Definición Central y Tipologías

El concepto de **datos censurados** se refiere a aquellas observaciones en un conjunto de datos en las que la magnitud de una variable de interés es solo parcialmente conocida. Esta situación surge comúnmente en estudios que analizan el tiempo transcurrido hasta que ocurre un evento específico, conocido como análisis de duración o análisis de supervivencia. A diferencia de los datos perdidos o faltantes (*missing data*), donde la observación simplemente no se registró o se perdió, en el caso de la censura, se tiene información parcial y valiosa que indica que el evento no ha ocurrido aún al momento de la última observación, o que ocurrió dentro de un intervalo específico. Ignorar esta información parcial invalida los resultados estadísticos, llevando a estimaciones sesgadas de los parámetros poblacionales, especialmente si se intenta calcular la media o la mediana del tiempo total hasta el evento. La censura es, por lo tanto, una característica inherente al diseño experimental o de observación, no un error de recolección de datos, y requiere métodos estadísticos especializados para su correcta incorporación en el modelo.

La necesidad de manejar la censura es crítica en la inferencia estadística, ya que la distribución de probabilidad del tiempo hasta el evento no es completamente observable. Por ejemplo, en un ensayo clínico diseñado para medir el tiempo de supervivencia de pacientes tratados con un nuevo fármaco, si un paciente abandona el estudio antes de que ocurra la muerte o si el estudio finaliza antes de que todos los sujetos experimenten el evento de interés, el tiempo exacto de supervivencia para ese individuo es desconocido, aunque se sabe que es mayor que el tiempo que permaneció en observación. Esta información parcial debe ser incorporada mediante técnicas que utilicen tanto la duración conocida (el tiempo de seguimiento) como la indicación de que el evento aún no ha ocurrido. El tratamiento adecuado de los datos censurados es fundamental para evitar la subestimación de la verdadera duración media de la supervivencia o la fiabilidad del producto analizado, asegurando así la validez y la robustez de las conclusiones extraídas del estudio.

La clasificación de los datos censurados se basa principalmente en la dirección de la inobservabilidad, lo cual genera tres tipologías fundamentales: la censura a la derecha, la censura a la izquierda y la censura intervalar. La **censura a la derecha** es, con mucho, la forma más frecuente y se produce cuando el tiempo de interés es mayor que el tiempo de observación registrado; es decir, el evento no ha ocurrido al final del seguimiento. La censura a la izquierda implica que el evento ya ha ocurrido antes de que comenzara el período de observación, por lo que solo se sabe que el tiempo de duración es menor que el tiempo inicial de registro. Finalmente, la **censura intervalar** ocurre cuando se sabe que el evento ha tenido lugar entre dos puntos de

observación definidos, pero no se conoce el momento exacto dentro de ese intervalo. Cada una de estas formas de censura plantea desafíos analíticos distintos que requieren ajustes específicos en los modelos de probabilidad para asegurar estimaciones consistentes y eficientes de la función de supervivencia o de la tasa de riesgo.

## 2. Etimología y Desarrollo Histórico

El reconocimiento formal y el desarrollo de métodos para manejar los datos censurados están intrínsecamente ligados al surgimiento y la maduración del campo de la **bioestadística** y la **ingeniería de fiabilidad** a mediados del siglo XX. Aunque los problemas de inobservabilidad parcial siempre han existido en la estadística, fue la necesidad de analizar rigurosamente los resultados de los ensayos clínicos a largo plazo y la durabilidad de los componentes industriales complejos (como los utilizados en la aviación y la tecnología militar) lo que impulsó la creación de un marco teórico coherente. Antes de estos desarrollos, los investigadores a menudo optaban por métodos simplistas y sesgados, como simplemente excluir a los sujetos censurados del análisis o tratarlos como si el evento hubiera ocurrido al final del seguimiento, lo cual distorsionaba sistemáticamente las conclusiones sobre la duración real.

Un hito fundamental en la historia del tratamiento de datos censurados fue la publicación en 1958 del trabajo seminal de Edward L. [Kaplan](#) y Paul [Meier](#), quienes introdujeron el estimador de límite de producto, universalmente conocido hoy como el estimador de **Kaplan-Meier**. Este método no paramétrico revolucionó la forma en que se estimaban las funciones de supervivencia a partir de datos que incluían censura a la derecha, al proporcionar una curva escalonada que utiliza toda la información disponible, tanto la de los eventos completos como la de los sujetos censurados. Este desarrollo permitió a los investigadores obtener estimaciones no sesgadas de la probabilidad de supervivencia en cualquier momento, sentando las bases para todo el campo del análisis de supervivencia moderno. La elegancia y robustez de este estimador lo convirtieron rápidamente en la herramienta estándar para la descripción de datos de duración.

Posteriormente, en 1972, el estadístico británico [David Cox](#) introdujo otro avance crucial: el modelo de riesgos proporcionales (*Proportional Hazards Model*). Mientras que Kaplan-Meier se enfoca en la descripción univariada de la supervivencia, el modelo de Cox permitió la inclusión de covariables (variables explicativas) en el análisis, permitiendo a los investigadores evaluar cómo factores como la edad, el tratamiento o las condiciones preexistentes afectan la tasa de riesgo instantánea del evento, incluso en presencia de censura. El modelo de Cox es semi-paramétrico, lo que significa que no requiere la especificación de una distribución particular para el tiempo de supervivencia subyacente, sino que se centra en cómo las covariables modifican el riesgo. Estos dos pilares--Kaplan-Meier para la descripción y Cox para la modelización multivariada--constituyen la espina dorsal metodológica para el manejo de los datos censurados en la actualidad, siendo esenciales para la validez de la investigación en campos tan diversos como la epidemiología, la

economía del trabajo y la ingeniería.

### 3. Tipos Fundamentales de Censura

La clasificación detallada de los tipos de censura es vital para seleccionar el método estadístico apropiado y para entender las implicaciones de las suposiciones subyacentes. Si bien la distinción primaria es direccional (derecha, izquierda, intervalar), existen subtipos importantes, especialmente dentro de la censura a la derecha, que se definen por el mecanismo que determina el momento de la censura, afectando directamente la aleatoriedad y la independencia del proceso de observación. La censura a la derecha, que es la más común en la práctica, se subdivide a menudo en tres categorías basadas en el control que el investigador tiene sobre el momento en que finaliza la observación para un sujeto dado.

Estos mecanismos de censura se distinguen por cómo se relaciona el tiempo de censura con el tiempo real del evento que estamos tratando de observar. Es crucial que el mecanismo de censura sea no informativo (o independiente) para que los métodos estándar como Kaplan-Meier y Cox funcionen correctamente. Si el mecanismo de censura está relacionado con la probabilidad de que ocurra el evento de interés (es decir, censura informativa), se introduce un sesgo grave que requiere modelos mucho más complejos y a menudo estructurales para su corrección. Por ejemplo, si los pacientes más enfermos son los que tienen más probabilidades de abandonar un estudio (censura informativa), las estimaciones de supervivencia basadas en los pacientes restantes estarán sesgadas al alza.

A continuación, se detallan los tipos principales, incluyendo las subdivisiones basadas en el diseño experimental:

**Censura a la Derecha (Right Censoring):** Ocurre cuando se conoce el tiempo de inicio de la observación, pero el evento de interés no ha ocurrido cuando finaliza el seguimiento. El tiempo real del evento es, por lo tanto, mayor que el tiempo de censura registrado. Es la forma más analizada y modelada, y se subdivide en:

**Censura Tipo I (Type I Censoring):** El estudio finaliza en un tiempo predeterminado y fijo ( $T$ ). Todos los sujetos que no han experimentado el evento al llegar a  $T$  son censurados. El número de eventos es aleatorio.

**Censura Tipo II (Type II Censoring):** El estudio finaliza cuando un número predeterminado de eventos ( $k$ ) ha ocurrido. El tiempo total del estudio es aleatorio.

**Censura Aleatoria (Random Censoring):** El tiempo de censura varía de un sujeto a otro y se considera independiente del tiempo real del evento. Esto ocurre cuando los sujetos se pierden, abandonan el estudio o mueren por causas no relacionadas con el evento de interés.

**Censura a la Izquierda (Left Censoring):** Ocurre cuando el evento de interés ya ha ocurrido

antes de que el sujeto sea observado por primera vez. Solo se sabe que el tiempo de duración es menor que el tiempo de inicio del seguimiento. Esto es común en estudios retrospectivos o cuando el momento exacto del inicio del proceso es difícil de determinar.

**Censura Intervalar (Interval Censoring):** Ocurre cuando no se observa continuamente al sujeto, sino solo en puntos discretos. Se sabe que el evento ocurrió en algún momento entre dos observaciones consecutivas, pero el momento exacto es desconocido. Este tipo de censura requiere algoritmos de estimación más complejos, a menudo basados en la maximización de la función de verosimilitud.

#### 4. Mecanismos de Censura y el Supuesto de No Informatividad

La validez de la mayoría de los métodos estadísticos estándar para datos censurados depende de un supuesto fundamental conocido como **Censura No Informativa (CNI)**. Este supuesto postula que el proceso de censura es estocásticamente independiente del proceso de supervivencia. En términos prácticos, esto significa que la probabilidad de que un individuo sea censurado en un momento dado no debe estar relacionada con su probabilidad futura de experimentar el evento de interés en ese mismo momento. Si este supuesto se cumple, la información de los sujetos censurados hasta el momento de la censura puede utilizarse de manera efectiva para hacer inferencias sobre toda la población. La independencia estocástica permite tratar el tiempo de censura como una variable exógena al proceso de interés, simplificando enormemente la formulación de la función de verosimilitud.

El incumplimiento del supuesto de CNI conduce al problema de la **Censura Informativa**. La censura es informativa cuando el tiempo de censura está intrínsecamente ligado al resultado de interés. Un ejemplo clásico se presenta en los estudios de trasplantes de órganos: si los pacientes que están a punto de experimentar un fallo orgánico grave son retirados de una lista de espera para recibir un trasplante (siendo así censurados en la lista), el tiempo hasta el trasplante para los pacientes restantes (que son menos enfermos) aparecerá artificialmente más largo, sesgando la estimación del tiempo medio de espera. Cuando la censura es informativa, los métodos estándar producen estimaciones sesgadas y las inferencias son inválidas. La detección y el tratamiento de la censura informativa son temas de investigación avanzada, a menudo requiriendo modelos que especifican conjuntamente la distribución del tiempo hasta el evento y la distribución del tiempo hasta la censura, utilizando variables latentes o modelos de riesgo competitivo.

Distinguir entre la censura aleatoria no informativa y la censura informativa es a menudo una tarea difícil, ya que la naturaleza de la censura se basa en mecanismos no observados. Los investigadores deben basarse en el conocimiento sustantivo del campo y en el diseño del estudio para justificar la suposición de CNI. Por ejemplo, en un ensayo clínico con censura Tipo I (finalización fija del estudio), la censura es inherentemente no informativa, ya que la finalización del estudio es una decisión administrativa no relacionada con el estado de salud individual de los

participantes al final. Sin embargo, en estudios observacionales donde la pérdida de seguimiento es común, la censura es frecuentemente sospechosa de ser informativa, obligando a los analistas a realizar análisis de sensibilidad o a emplear modelos más sofisticados que intenten corregir el posible sesgo mediante la modelización explícita de la dependencia entre los dos procesos de tiempo.

## 5. Aplicaciones Prácticas y Campos de Estudio

Los datos censurados son omnipresentes en cualquier disciplina que involucre el análisis de duraciones. La aplicación más destacada se encuentra en el **Análisis de Supervivencia** dentro de la [Bioestadística](#) y la Epidemiología. En este contexto, el evento de interés puede ser la muerte, la recurrencia de una enfermedad, la falla de un tratamiento o la remisión. Los ensayos clínicos, por su propia naturaleza y duración limitada, generan una gran cantidad de datos censurados a la derecha, ya que muchos pacientes seguirán vivos o libres de enfermedad cuando el estudio concluya. La correcta aplicación del estimador de Kaplan-Meier y del modelo de Cox permite a los médicos e investigadores evaluar la eficacia de los tratamientos y pronosticar la evolución de las enfermedades de manera precisa, haciendo inferencias cruciales sobre la esperanza de vida o el tiempo libre de progresión de la enfermedad.

Otro campo fundamental es la **Ingeniería de Fiabilidad** (Reliability Engineering). En este ámbito, el evento de interés es el fallo de un componente mecánico o electrónico (por ejemplo, el tiempo de vida útil de un motor, un chip de computadora o un neumático). Las pruebas de fiabilidad aceleradas o de larga duración a menudo deben ser detenidas antes de que todos los componentes fallen, resultando en datos censurados a la derecha. El análisis de estos datos permite a los fabricantes estimar la distribución de la vida útil de sus productos, establecer garantías adecuadas y optimizar los cronogramas de mantenimiento preventivo. Modelos paramétricos, como la distribución de Weibull, son frecuentemente utilizados en este campo, ajustados para incorporar la censura y proporcionar predicciones robustas sobre la tasa de fallos a lo largo del tiempo.

Finalmente, la **Econometría** también aborda activamente los datos censurados y truncados mediante los llamados modelos de duración o modelos limitados. En economía, la variable de interés puede ser el tiempo de desempleo, la duración de una huelga, o el tiempo que tarda un consumidor en adoptar una nueva tecnología. Los modelos Tobit, por ejemplo, se utilizan para variables dependientes que están censuradas en un límite (por ejemplo, el gasto familiar en un bien que no puede ser negativo, donde la observación cero es una censura a la izquierda). El manejo riguroso de la censura es esencial en estos campos para evitar conclusiones erróneas sobre el comportamiento económico y para informar la política pública, como en la evaluación de la efectividad de los programas de capacitación laboral en la reducción del tiempo de desempleo.

## 6. Métodos Estadísticos para el Tratamiento

El manejo estadístico de los datos censurados se centra en la construcción de funciones de verosimilitud que incorporen la información parcial proporcionada por las observaciones censuradas. Para una observación no censurada (un evento completo), la contribución a la verosimilitud es simplemente la función de densidad de probabilidad del tiempo hasta el evento. Sin embargo, para una observación censurada a la derecha en el tiempo  $c$ , la contribución a la verosimilitud no es la densidad, sino la probabilidad de que el tiempo del evento  $T$  sea mayor que  $c$ , lo cual es la función de supervivencia en ese punto,  $S(c)$ . Los métodos estadísticos combinan estas dos formas de contribución para obtener estimaciones que utilizan toda la información de la muestra.

El método más básico y descriptivo es el ya mencionado **Estimador de Kaplan-Meier**. Este estimador no paramétrico calcula la probabilidad de supervivencia en puntos de tiempo discretos, ajustando la base de riesgo en cada momento en que ocurre un evento, utilizando solo el número de sujetos "en riesgo" en ese momento (es decir, aquellos que aún no han experimentado el evento ni han sido censurados antes de ese tiempo). La curva resultante, el límite de producto de las probabilidades condicionales de supervivencia, proporciona una estimación visual e inferencial de la función de supervivencia  $S(t)$ , siendo el estándar de oro para la descripción de la duración en presencia de censura a la derecha.

Cuando se requiere la inclusión de covariables, el **Modelo de Riesgos Proporcionales de Cox** se convierte en la herramienta principal. Este modelo se centra en la función de riesgo (*hazard function*), que es la tasa instantánea de ocurrencia del evento. La formulación de Cox es:  $h(t | X) = h_0(t) \exp(\beta'X)$ , donde  $h_0(t)$  es la función de riesgo base (no especificada paramétricamente) y  $\exp(\beta'X)$  es la parte que modela el efecto de las covariables ( $X$ ) sobre el riesgo. Este modelo permite estimar los coeficientes de regresión ( $\beta$ ) que indican la magnitud y dirección del efecto de cada covariable sobre el riesgo, utilizando la función de verosimilitud parcial, que elimina la necesidad de especificar la forma de  $h_0(t)$ , haciendo el modelo robusto ante la forma de la distribución del tiempo.

## 7. Desafíos y Limitaciones Metodológicas

A pesar de la sofisticación de los modelos de supervivencia, el análisis de datos censurados presenta varios desafíos metodológicos. El más grave es la ya discutida **Censura Informativa**, la cual, si no se identifica y corrige, invalida las conclusiones. Abordar la censura informativa a menudo requiere el uso de modelos de riesgos competitivos (si la censura es un evento que impide la observación del evento principal) o modelos de fragilidad, que introducen variables aleatorias no observadas (fragilidades) para modelar la heterogeneidad no medida que podría estar correlacionada tanto con el tiempo de supervivencia como con el tiempo de censura.

Otro desafío importante es la **validación del supuesto de riesgos proporcionales** en el modelo de Cox. Este supuesto establece que el efecto relativo de las covariables sobre el riesgo es constante a lo largo del tiempo. Si este supuesto es violado (es decir, si el efecto de un factor de riesgo cambia con el tiempo), el modelo de Cox estándar proporciona estimaciones sesgadas. Los métodos para abordar esta limitación incluyen la introducción de términos de interacción tiempo-covariable o el uso de modelos de regresión de supervivencia acelerada, que modelan directamente el logaritmo del tiempo de supervivencia en lugar del riesgo.

Finalmente, la gestión de la **censura intervalar** plantea dificultades computacionales significativas. Dado que el momento exacto del evento es desconocido, los métodos de estimación suelen basarse en técnicas iterativas complejas, como el algoritmo de Expectation-Maximization (EM), para estimar los parámetros de la distribución subyacente. Además, al igual que con cualquier método de inferencia que maneja datos incompletos, la precisión de las estimaciones de los datos censurados es inherentemente menor que la de los datos completos, lo que requiere tamaños muestrales considerablemente mayores para alcanzar el mismo nivel de potencia estadística y precisión.

## 8. Conclusión y Significado

Los datos censurados representan una manifestación inevitable de las limitaciones prácticas en la observación de fenómenos de duración. Lejos de ser un simple obstáculo, el desarrollo de metodologías estadísticas robustas para su manejo (principalmente el estimador de Kaplan-Meier y el modelo de Cox) ha permitido a los investigadores extraer inferencias válidas y no sesgadas en campos críticos como la medicina, la ingeniería y la economía. La capacidad de utilizar la información parcial proporcionada por las observaciones incompletas es lo que distingue al análisis de supervivencia moderno de enfoques estadísticos más simplistas que descartarían los datos censurados, produciendo resultados erróneos.

El significado de los datos censurados reside en su papel como catalizador para la innovación metodológica. La necesidad de modelar la inobservabilidad ha impulsado avances en la teoría de la verosimilitud parcial y los modelos de riesgos, permitiendo una comprensión más profunda de la función de riesgo y los factores que influyen en la duración de los eventos. En última instancia, el tratamiento riguroso de la censura garantiza que las decisiones críticas, ya sean médicas (elección de un tratamiento) o industriales (establecimiento de garantías), se basen en estimaciones de duración que reflejan fielmente la realidad subyacente del proceso.

## 9. Lectura Adicional

[Análisis de supervivencia \(Wikipedia\)](#)

[Edward L. Kaplan \(Wikipedia\)](#)

[David Cox \(Wikipedia\)](#)

[Bioestadística \(Wikipedia\)](#)

ARABPSYCHOLOGY.COM