

UNIVERSIDAD DE ZARAGOZA
FACULTAD DE CIENCIAS



Universidad
Zaragoza



Facultad de Ciencias
Universidad Zaragoza

TRABAJO DE FIN DE GRADO

Análisis de fiabilidad de máquinas automáticas de cobro de efectivo.

Reliability analysis of automatic cash payment machines.

Directores: Germán Badía Blasco y María Dolores Berrade Ursúa.

Mikel Balduz Roldán

Grado en Matemáticas

Curso académico 2022/2023

Índice

1. Resumen	2
2. Abstract	2
3. Introducción a la fiabilidad de sistemas	2
3.1. Medidas de la fiabilidad	3
3.2. Datos censurados	5
3.3. Estimación no paramétrica de la fiabilidad. Kaplan-Meier	6
3.4. Estimación paramétrica de la fiabilidad. Distribución Weibull	9
3.5. Contrastes de hipótesis	12
3.6. Prueba de bondad de ajuste. Test de Anderson-Darling	12
3.7. Estimación paramétrica por máxima verosimilitud	13
3.8. Intervalos de confianza basados en la verosimilitud	14
3.9. Intervalos de confianza simultáneos. Desigualdad de Bonferroni	15
3.10. Modelo de regresión Weibull	15
3.11. Residuos de Cox-Snell	16
4. Descripción de la máquina. Funcionamiento y tipos de fallo	17
4.1. Parte de billetes	19
4.2. Parte de monedas	19
5. Tratamiento inicial del fichero de datos	21
5.1. Depuración de datos y análisis descriptivo de las variables	22
5.2. Estudio cruzado: tipo de fallo por establecimiento	23
5.3. Estudio de la variable respuesta	24
6. Estudio de las diferencias entre tipos de fallo	25
6.1. Análisis no paramétrico	25
6.2. Análisis paramétrico	27
6.3. Modelos de regresión Weibull en Cashlogy	28
6.4. Funciones de riesgo asociadas a cada tipo de fallo	33
7. Mantenimiento preventivo en el periodo de garantía	35
7.1. Función de coste de los fallos frente al número de transacciones.	36
7.2. Análisis de políticas óptimas de mantenimiento preventivo	37
7.3. Política de mantenimiento óptima para el reciclador	40
7.4. Número esperado de incidencias por tipo de fallos.	40
8. Conclusiones	42
9. Bibliografía	44

1. Resumen

En este trabajo se exponen los resultados sobre el estudio de fiabilidad realizados en la máquina Cashloghy, dedicada a la gestión automatizada del cobro mediante dinero en efectivo.

Para el análisis de fiabilidad utilizaremos los datos de los fallos registrados por la empresa fabricante mediante la monitorización de 48 máquinas entre febrero y julio de 2022. La información recoge todas las transacciones de las máquinas hasta la fecha de su mantenimiento preventivo lo cual conlleva un esquema de censura de tipo I.

2. Abstract

This paper presents the results concerning a reliability study of an automatic cash payment machine (Cashloghy). The failure data were collected by the Original Equipment Manufacturer by monitoring 48 machines between February and March 2022. The number of transactions until failure is considered the response variable in this analysis. In addition a type I censoring scheme is assumed with a prefixed duration of the test determined by the date of the preventive maintenance.

3. Introducción a la fiabilidad de sistemas

En la revisión de las bases de la fiabilidad que se presenta en esta sección, el término sistema se refiere a cualquier elemento o conjunto de ellos que forman parte de la tecnología o de los mecanismos desarrollados por el hombre. Por tanto puede aludir a una lavadora, a un coche, a un aparato de diagnóstico médico o a alguna de sus partes o componentes. En el problema objeto de estudio en este TFG, define, específicamente, una máquina de cobro electrónico de efectivo. Los sistemas se diseñan para cumplir una función determinada y hacerlo, además, en entornos y bajo condiciones que pueden variar. Así, por ejemplo, los kilómetros anuales que recorre el coche de un usuario particular están muy lejos de los que realiza un vehículo de servicio público como un taxi. Todos los sistemas terminan por fallar debido tanto al deterioro que se produce al usarlos como al envejecimiento natural. En otros casos los fallos ocurren de modo temprano como consecuencia de un diseño deficiente que los hace estar mal dimensionados para su uso real o por problemas en la producción. Esto es lo que habitualmente se conoce como "defectos de fábrica".

Así, cuando hablamos de la fiabilidad de una componente, de una máquina, de un sistema en

general, nos referimos a su funcionamiento sin fallo. La característica de los fallos es que ocurren de modo aleatorio pudiendo ocasionar elevados costes cuando, por ejemplo, se producen interrupciones en la fabricación, producción defectuosa, se paraliza un servicio, etc.

Se precisan, por tanto, medidas que cuantifiquen la fiabilidad de un sistema de modo que sea posible establecer previsiones sobre su duración o el riesgo de que se vaya a producir un fallo. A partir de ellas se podrán establecer políticas de mantenimiento preventivo que eviten la ocurrencia de fallos especialmente en momentos cuando el sistema es particularmente necesario. Así se podrá tener una elevada seguridad de que una máquina podrá completar un ciclo de trabajo o una determinada misión.

Cuando se trata de sistemas complejos formados por varias componentes, es crucial identificar los puntos débiles, es decir, las partes que fallan más tempranamente o con mayor frecuencia. Este punto es especialmente relevante porque la información puede hacer volver al fabricante a la fase de diseño para mejorar la fiabilidad del sistema sustituyendo los elementos más deficientes por otros más robustos.

3.1. Medidas de la fiabilidad

En lo que sigue, T representa el "tiempo que transcurre hasta el fallo de un sistema". T es una variable aleatoria positiva que puede ser discreta o continua aunque, mayoritariamente se suele asumir este segundo caso. Denotaremos $f(t)$ la función de densidad asociada. Es importante señalar que el tiempo puede estar medido en una escala arbitraria. En el problema objeto de estudio de este TFG la respuesta está dada por el "número de transacciones de la máquina hasta el fallo". Dado que las máquinas están instaladas en lugares en los que la afluencia de público varía mucho, es lógico que el tiempo hasta el fallo sea más corto cuando hay muchos usuarios. Al considerar el número de transacciones se tiene en cuenta el uso que se haga independientemente del tiempo hasta alcanzarlo. En la vida real encontramos ejemplos similares. Por ejemplo los mantenimientos de un coche suelen estar determinados por los kilómetros recorridos alternativamente a un periodo de tiempo.

A continuación se presentan las medidas más habituales de la fiabilidad. Su valor puede depender de las diferencias entre fabricantes de un sistema particular, o de las condiciones en las que tiene que funcionar el sistema. Esas diferencias se modelan mediante un vector de parámetros θ .

■ **Función de fiabilidad o supervivencia:**

$$S(t, \theta) = P(T \geq t) = 1 - F(t, \theta), \quad t > 0.$$

Con $F(t, \theta)$ la función de distribución acumulada en t .

La función de supervivencia proporciona la probabilidad de que la variable T alcance sin fallo un determinado instante t , es decir que el sistema al llegar t no haya fallado todavía.

■ **Función de riesgo o tasa de fallo:**

$$h(t, \theta) = \frac{f(t, \theta)}{S(t, \theta)}, \quad t > 0.$$

Representa la probabilidad de fallo inmediato en t dado que éste no había ocurrido todavía. La forma de la función de riesgo refleja igualmente la fiabilidad del sistema y puede servir como indicador del instante para reemplazar un sistema en uso por otro nuevo.

- Funciones de riesgo decrecientes surgen cuando al principio del funcionamiento la probabilidad de fallo es alta debido a posibles defectos de fabricación. Tras el reemplazamiento de las partes débiles por otras más resistentes, la fiabilidad mejora. Estos procedimientos para el cribado de los elementos defectuosos se conocen como "burn-in".
- Funciones de riesgo constantes ocurren durante la vida útil cuando los fallos son debidos a un fenómeno aleatorio como sobrecargas o accidentes. En este periodo ya no hay defectos ocultos y el sistema aún no ha empezado a envejecer.
- Funciones de riesgo crecientes reflejan el envejecimiento o deterioro de la componente por efecto del uso y la edad.

Estos tres posibles escenarios se reflejan en la llamada curva de bañera, usada con frecuencia en fiabilidad que se representa en la Figura 1.

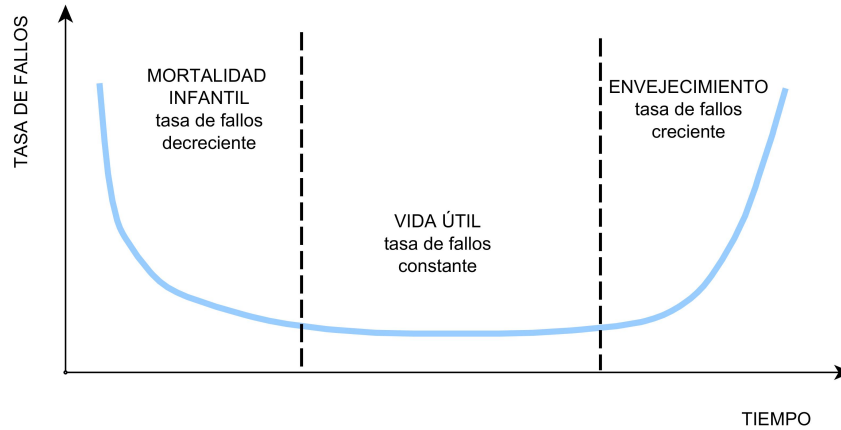


Figura 1: Curva de bañera

- **Cuantil p de T** : $t_p := \min \{t | F(t; \theta) \geq p\}$.
- **Tiempo medio hasta el fallo**: $MTTF \rightarrow E[T] = \int_0^\infty t f(t; \theta) dt$.

3.2. Datos censurados

Una práctica habitual para evaluar la fiabilidad de un sistema es hacer test de resistencia, es decir, se ponen a funcionar unidades en prueba hasta que fallan. Los datos de este trabajo corresponden a un experimento de estas características.

Supongamos que se tienen n unidades en funcionamiento. Como no se hace un seguimiento continuo, es frecuente que al finalizar el experimento se desconozca el valor exacto del tiempo hasta el fallo de algunas de estas unidades, disponiéndose únicamente de cotas de ese valor. Esa información se denomina dato censurado. En el análisis de supervivencia los datos censurados aparecen al estudiar, por ejemplo, el tiempo que un grupo de pacientes permanecen en estado de remisión de una enfermedad. Alguno de ellos puede desaparecer del estudio porque se traslada a otro lugar o porque fallece por una causa diferente, un accidente por ejemplo. En estos casos se sabe que en el momento del último control el paciente permanecía vivo pero se ignora el tiempo final de vida. En lugar de eliminar la observación, el tratamiento como dato censurado permite utilizar la información del individuo mientras permanece en el estudio.

Una observación puede presentar tres tipos de censura:

Censura a derecha si el dato es una cota inferior del tiempo hasta el fallo, es decir, se sabe que la duración está por encima de del valor de censura.

Censura a izquierda ocurre cuando el instante de fallo es anterior al valor de censura.

Censura en intervalo tiene lugar cuando no se hace un seguimiento continuado de las unidades en prueba, sino que se verifica su estado en instantes de observación t_i . Al detectar un fallo sólo se dispone de la información de que éste ha ocurrido entre dos observaciones consecutivas, t_{j-1} y t_j . Los datos así obtenidos resultan censurados por intervalos, esto es, conocemos una cota inferior y una cota superior de su tiempo de fallo. Cuando en el momento de la última inspección una unidad no ha fallado entonces se trata de un dato censurado por la derecha. Así mismo, si una unidad ha fallado antes de la primera inspección entonces se trata de un dato censurado por la izquierda.

En el caso de estudios de fiabilidad de sistemas en la tecnología y la industria los datos censurados pueden tener lugar como consecuencia de las ciertas prácticas para que los experimentos no se prolonguen excesivamente en el tiempo. Esto se conoce como censura de tipo I y censura de tipo II.

- CENSURA DE TIPO I: se fija una duración máxima de la duración de las pruebas. El número de unidades que habrán fallado al finalizar éste es aleatorio mientras que otras seguirán funcionando sin fallo.
- CENSURA DE TIPO II: El test finaliza cuando se hayan observado un número predeterminado, r de fallos. La duración del experimento es aleatoria y cuando concluye quedan $n - r$ unidades que permanecen con vida.

La censura de tipo I y de tipo II proporcionan datos con censura a derecha que resulta ser así la más habitual en estudios de fiabilidad en ingeniería.

3.3. Estimación no paramétrica de la fiabilidad. Kaplan-Meier

En este apartado se presenta la estimación de la función de fiabilidad $S(t)$ de la variable tiempo hasta el fallo T a partir una muestra de tiempos hasta el fallo.

Supongamos que se dispone de una muestra de n individuos y que existen k ($k \leq n$) instantes diferentes en los que ocurren los fallos $t_1 < \dots < t_k$.

Se admite la posibilidad de coincidencias; es decir, más de un fallo en t_j , denotando d_j el número de fallos habidos en t_j .

Una estimación natural se obtiene a partir de la función de supervivencia empírica, $\hat{S}(t)$:

$$\hat{S}(t) = \frac{\text{número de observaciones} \geq t}{n}, \quad t > 0 \quad (1)$$

Supongamos que existen instantes de censura, C_i , correspondientes a aquellas unidades cuyos tiempos de vida no son observados. En este caso es preciso modificar la expresión anterior $\hat{S}(t)$ ya que no se conoce exactamente el número de observaciones mayores o iguales a t . La modificación que se presenta seguidamente se conoce como estimador de Kaplan-Meier.

Sea n_j el número de individuos en riesgo en t_j ; es decir, aquellos que no han fallado ni han sido censurados justo antes de t_j . En el caso de que un instante de censura C_i y uno de fallo t_j coincidan, se adopta la convención de considerar C_i superior a t_j en una cantidad infinitesimal. Por este motivo, cualquier unidad de cuyo instante de censura sea igual a t_j se incluye en el conjunto de las n_j unidades en riesgo justo antes de t_j .

Los tiempos de fallo exactos se obtienen a partir de procesos de inspección continuos, o bien cuando se realizan inspecciones muy frecuentes de forma que los intervalos tienen una amplitud tan pequeña que se puede asignar un valor al tiempo de fallo.

En lo que sigue, se utilizará la siguiente notación:

$$\hat{S}(t_j + 0) = \lim_{x \rightarrow 0} \hat{S}(t_j + x)$$

El estimador de Kaplan-Meier se utiliza para estimar probabilidades de supervivencia a partir de un conjunto de datos correspondientes a tiempos de fallo y censuras. Éste estimador se calcula recursivamente:

$$\hat{S}(t_1 + 0) = \frac{n_1 - d_1}{n_1}$$

$$\hat{S}(t_2 + 0) = \hat{S}(t_1 + 0) \frac{n_2 - d_2}{n_2}$$

$$\hat{S}(t_j + 0) = \hat{S}(t_{j-1} + 0) \frac{n_{j-1} - d_{j-1}}{n_{j-1}}$$

Es decir, la probabilidad de sobrevivir pasado el instante t_j se calcula mediante la probabilidad condicionada de no haber fallado justo antes de t_j . Definiéndose, en cada instante t , como:

$$\hat{S}(t) = \prod_{j:t_j < t} \frac{n_j - d_j}{n_j}$$

$\hat{S}(t)$ una función escalón cuyos saltos, de longitud $\frac{n_j - d_j}{n_j}$, se producen inmediatamente después de cada fallo t_j . En los instantes de censura no se produce variación alguna si bien estos afectan al tamaño del conjunto de riesgo y, por consiguiente a las magnitudes de los saltos.

Una de las características importantes del estimador de Kaplan-Meier es que permite utilizar la información asociada a un individuo hasta el instante en que se falla o es censurado. En el caso de que no haya censura, el estimador de Kaplan-Meier coincide con la función de fiabilidad empírica. Otro aspecto a destacar es que Kaplan-Meier es un método no paramétrico de estimación de la fiabilidad. Al no asumirse ninguna distribución sobre la variable puede aplicarse en condiciones muy generales. En la Figura 2 se representa gráficamente el estimador de Kaplan-Meier del número de transacciones hasta que ocurre un atasco en lectura que es uno de los tipos de fallo analizados en este TFG.

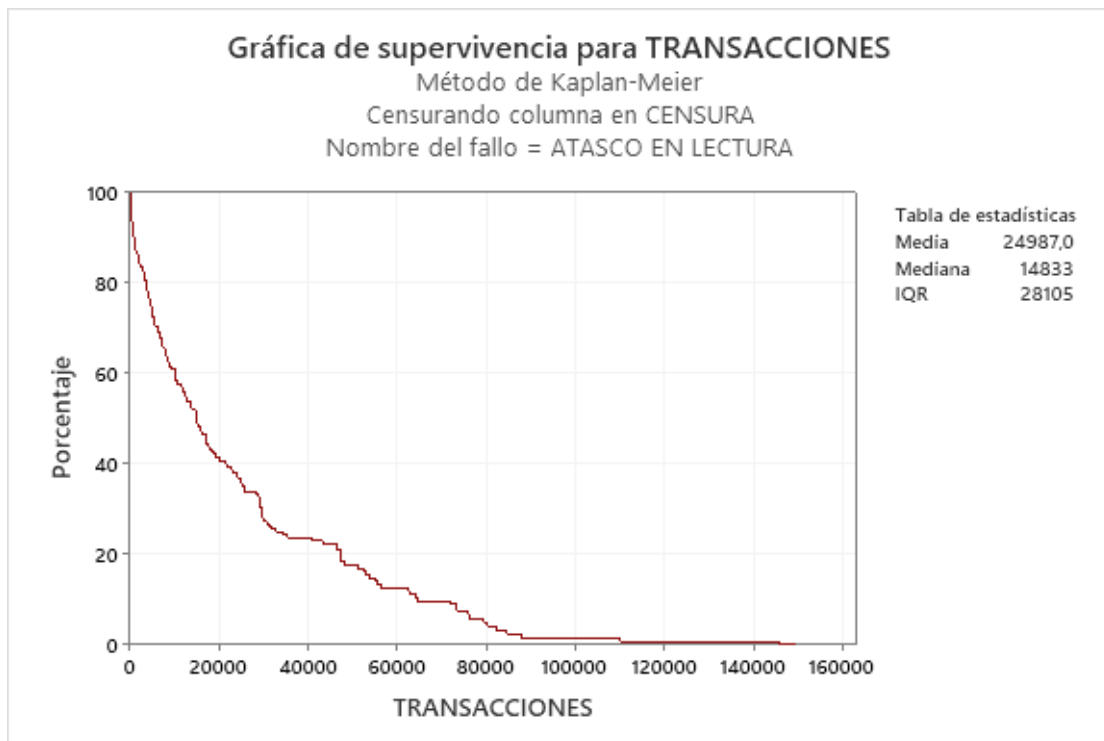


Figura 2: Estimador de Kaplan Meier del número de transacciones hasta el atasco en lectura

La gráfica de supervivencia indica que la probabilidad de superar, por ejemplo, 20000 transacciones sin que ocurra este fallo específico es 0.4. En otras palabras, existe un 40 % de probabilidad de que el atasco en lectura tarde más de 20000 transacciones en aparecer.

3.4. Estimación paramétrica de la fiabilidad. Distribución Weibull

Una variable aleatoria continua T tendrá una distribución Weibull, y se denota por $T \sim Weibull(a, \beta)$, si su función de densidad es

$$f(t) = \lambda\beta(\lambda t)^{\beta-1}e^{-(\lambda t)^\beta}, \quad t > 0$$

Con $\lambda = \frac{1}{a}$, siendo a y β los parámetros de escala y forma respectivamente.

Su función de distribución

$$F(t) = 1 - e^{-(\lambda t)^\beta}$$

su función de fiabilidad

$$S(t) = e^{-(\lambda t)^\beta}$$

y su función de riesgo

$$h(t) = \lambda\beta(\lambda t)^{\beta-1}$$

el p-cuantil

$$t_p = \frac{1}{\lambda} [-\log(1 - p)]^{1/\beta}$$

y el valor medio

$$E(T) = \frac{1}{\lambda} \Gamma\left(1 + \frac{1}{\beta}\right) \quad (2)$$

siendo Γ la función Gamma de Euler, β el parámetro de forma y λ el inverso del parámetro de escala.

La variable aleatoria Weibull tiene especial relevancia. Es una distribución versátil que se puede utilizar para modelar datos de tiempo de fallo en una amplia gama de aplicaciones en ingeniería, investigación médica, control de calidad, finanzas y climatología.

Asimismo existe una extensión de la distribución de Weibull usando 3 parámetros asociados a forma, escala y valor umbral. Este último permite modelar datos en los que el 0 no es el origen de los tiempos lo cual permite su uso con ceros y datos negativos. La distribución Weibull de 2

parámetros resulta ser un caso particular cuando el valor umbral es igual a cero y su uso queda limitado a variables positivas como ocurre en el problema que analizamos en este TFG.

Efecto del parámetro de forma

El parámetro de forma describe la manera en que se distribuyen los datos. Un valor de forma bajo, menor o igual a 1, da una curva con asimetría hacia la derecha. Un valor de forma alto, por ejemplo 10, refleja una curva con asimetría hacia la izquierda. En la Figura 3 se observa el efecto que tienen en la función de densidad las variaciones en el parámetro de forma. Estos cambios afectan también a la función tasa de fallos también llamada de riesgo:

- $0 < \beta < 1$ la función de riesgo es decreciente.
- $\beta > 1$ la función de riesgo es creciente.
- $\beta = 1$ la distribución Weibull coincide con la exponencial.

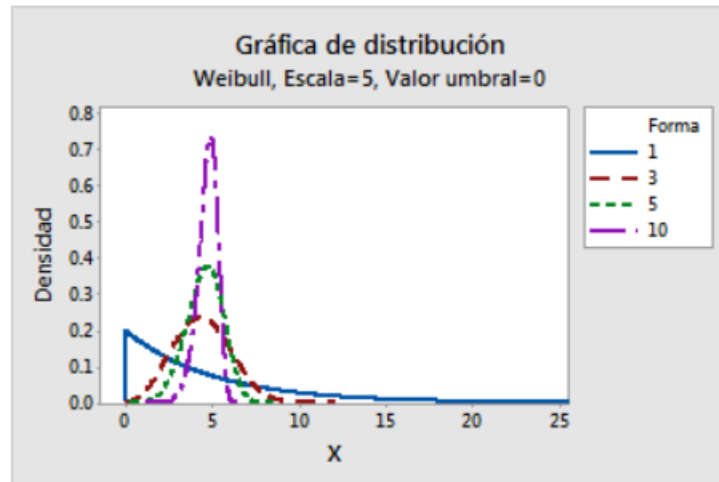


Figura 3: Distribución Weibull. Variaciones en el parámetro de forma

Efecto del parámetro de escala

El parámetro de escala de la distribución Weibull, a , se había definido como $a = \frac{1}{\lambda}$. Una primera propiedad es que que $1 - F(\frac{1}{\lambda}) = 0.368$, es decir, el parámetro de escala corresponde con el percentil 63.2 de los datos y se interpreta como instante temporal en el que han fallado el 63.2% de las unidades. La escala define la posición de la curva de Weibull respecto del valor de umbral. Una escala de 20, por ejemplo, indica que el 63.2% de los equipos fallará en las primeras 20 unidades

de tiempo después del valor umbral. En la Figura 4 se aprecian los desplazamientos en la función de densidad a causa de los cambios en el parámetro de escala.

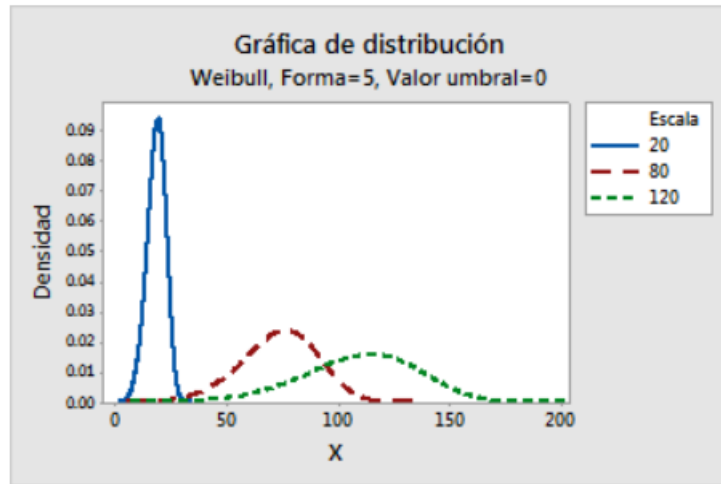


Figura 4: Distribución Weibull. Variaciones en el parámetro de escala

Efecto del valor umbral

El parámetro de valor umbral describe un desplazamiento de la distribución alejándose del 0. Un valor umbral negativo desplaza la distribución hacia la izquierda, mientras que un valor umbral positivo desplaza la distribución hacia la derecha. Todos los datos deben ser mayores que el valor umbral. En la Figura 5 se ve que la variación del valor umbral es una traslación de la variable.

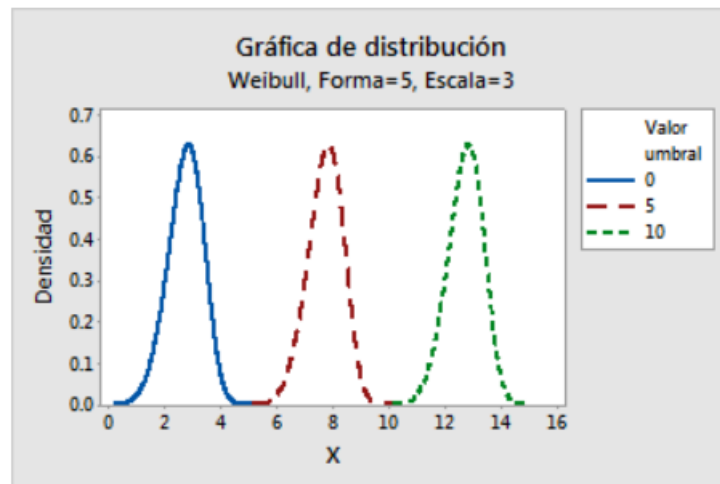


Figura 5: Distribución Weibull. Variaciones en el valor umbral

3.5. Contrastes de hipótesis

Los contrastes de hipótesis tratan de identificar cuál entre dos hipótesis (hipótesis nula H_0 e hipótesis alternativa H_1) resulta mas verosímil con los datos disponibles. Si al final se observa que hay poca evidencia en la muestra que apoye H_0 , esta hipótesis se rechaza. Con esto, existen dos tipos de errores, rechazar H_0 cuando ésta es cierta (error de tipo I) y aceptar H_0 cuando sea falsa (error de tipo II).

El nivel de significación, denotado como α , es fijado por quien analiza los datos y representa la máxima probabilidad que se admite para el error de tipo I (habitualmente $\alpha = 0.05$).

La potencia del test representa la probabilidad de no cometer error de tipo II, es decir, la capacidad del test para detectar que H_0 es falsa.

Estos contrastes están basados en determinados estadísticos, que resumen las características del conjunto de datos a estudio. El p-valor es la probabilidad calculada al asumir que la hipótesis nula H_0 es cierta, de que el estadístico de contraste tome valores tan o más extremos que los calculados con la muestra sin restricciones. Si esta probabilidad es menor que el valor α fijado, rechazaremos H_0 .

3.6. Prueba de bondad de ajuste. Test de Anderson-Darling

Una prueba de bondad de ajuste se refiere al siguiente contraste de hipótesis

H_0 : los datos siguen una distribución F_0

H_1 : los datos no siguen una distribución F_0

Se pueden distinguir dos casos:

- El parámetro (o parámetros) de la distribución es conocido.
- El parámetro (o parámetros) no es conocido y es necesario obtener su estimación.

El contraste de Anderson-Darling está basado en las diferencias entre F_0 y la función de distribución empírica dada en (1).

El estadístico **Anderson-Darling** mide cómo de bien siguen los datos F_0 . Para un conjunto de datos y distribución en particular, mientras mejor se ajuste la distribución a los datos, menor será este estadístico.

En este caso un p-valor inferior al nivel de significación elegido (por lo general 0.05 o 0.10), llevará a rechazar que los datos provienen de esa distribución particular.

También puede utilizar el estadístico de Anderson-Darling para comparar el ajuste de varias distribuciones con el fin de determinar cuál es la mejor. Para concluir que una distribución es la mejor, el estadístico de Anderson-Darling debe ser sustancialmente menor que los demás.

3.7. Estimación paramétrica por máxima verosimilitud

La función de verosimilitud de las observaciones, $L(\theta)$, está dada por:

$$L(\theta) = \prod_{i=1}^n f_{\theta}(t_i) \quad (3)$$

Donde $\theta = (\theta_1, \dots, \theta_k)$ es un vector de parámetros desconocidos que toman valores en un conjunto Ω . Si $\hat{\theta}$ es un elemento de Ω en el cual la función de verosimilitud (3) alcanza un máximo, entonces $\hat{\theta}$ se denomina estimador de máxima verosimilitud del parámetro θ , que en muchos modelos de distribución existe y es único.

Muchas veces para obtener $\hat{\theta}$ que maximiza la función de verosimilitud es más cómodo trabajar con la función log-verosimilitud que también alcanza un máximo en $\hat{\theta}$. La log-verosimilitud tiene la siguiente expresión:

$$l(\theta) = \log(L(\theta)) = \sum_{i=1}^n \log f_{\theta}(T_i) \quad (4)$$

En la mayoría de ocasiones, el estimador de máxima verosimilitud $\hat{\theta}$ se obtiene resolviendo las k ecuaciones

$$\frac{\partial l(\theta)}{\partial \theta_i} = 0, \quad i = 1, \dots, k \quad (5)$$

donde θ_i es la i -ésima componente del vector de parámetros $\theta = (\theta_1, \dots, \theta_k)$.

Este procedimiento es muy utilizado en la práctica debido a que proporciona estimadores consistentes, asintóticamente eficientes, insesgados y normalmente distribuidos en muestras grandes. Es el que se usa habitualmente en fiabilidad.

Dado que los datos censurados son cruciales en fiabilidad, vamos a ver como implementarlos en la función de verosimilitud. Solo vamos a exponer los que sean del tipo de censura I, que son los que se corresponden con los tratados en este estudio.

Con datos censurados del tipo I, existe un tiempo C en el que algunas máquinas no habrán fallado todavía. En esos casos sabemos que los tiempos hasta el fallo, T_i verifican $T_i > C$.

Estos datos pueden ser convenientemente representados por n pares de variables aleatorias (Z_i, δ_i) :

$$Z_i = \min(T_i, C)$$

y

$$\delta_i = \begin{cases} 1 & \text{si } T_i \leq C \\ 0 & \text{si } T_i > C \end{cases}$$

es el indicador de censura que valdrá 1 si el dato i no está censurado y 0 si el dato está censurado.

La contribución de este par (Z_i, δ_i) a la función de verosimilitud es

$$f(Z_i)^{\delta_i} S(Z_i)^{1-\delta_i}$$

donde f es la función de densidad y S la de fiabilidad. Por lo tanto, la función de verosimilitud en el caso de censura del tipo I queda

$$\prod_{i=1}^n f(Z_i)^{\delta_i} S(Z_i)^{1-\delta_i} = \prod_{i=1}^n f(T_i)^{\delta_i} S(C)^{1-\delta_i}$$

En el caso de la distribución Weibull no se pueden calcular analíticamente las soluciones de (5), sino que se debe recurrir a métodos numéricos para obtener dichas estimaciones como por ejemplo Newton-Raphson.

3.8. Intervalos de confianza basados en la verosimilitud

Una característica importante de los estimadores de máxima verosimilitud es que son asintóticamente normales, es decir, que si $\hat{\beta}$ y \hat{a} son los estimadores de máxima verosimilitud de los parámetros de forma y escala de una distribución Weibull entonces

$$\hat{\beta} \sim N(\beta, \sqrt{Var(\hat{\beta})}); \quad \hat{a} \sim N(a, \sqrt{Var(\hat{a})}).$$

Por lo tanto, si estandarizamos nos quedan

$$\frac{\hat{\beta} - \beta}{\sqrt{Var(\hat{\beta})}} \sim N(0, 1); \quad \frac{\hat{a} - a}{\sqrt{Var(\hat{a})}} \sim N(0, 1).$$

Así, los intervalos de confianza $(1 - \alpha) \%$ bilaterales son de la siguiente forma

$$\left[\hat{\beta} - z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\beta})}, \hat{\beta} + z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\beta})} \right]; \quad \left[\hat{a} - z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{a})}, \hat{a} + z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{a})} \right]$$

donde $z_{1-\frac{\alpha}{2}}$ es el cuantil de la distribución normal estándar.

Un aspecto a tener en cuenta es que si cada uno de los dos intervalos anteriores no cubre a su parámetro con una probabilidad α , la confianza global de que ambos contengan simultáneamente al que le corresponde, es inferior a $1 - \alpha$. Esta circunstancia da lugar a la construcción de intervalos de confianza simultáneos cuya confianza global es igual o superior a la que se desea, $1 - \alpha$. La idea en la que se basan, conocida como desigualdad de Bonferroni, se presenta a continuación.

3.9. Intervalos de confianza simultáneos. Desigualdad de Bonferroni

Supongamos que se estiman k parámetros diferentes mediante intervalos de confianza. Sea E_i el suceso correspondiente a que el parámetro $i, i = 1, \dots, k$ esté cubierto por su intervalo cuya probabilidad es $1 - \alpha$:

$$P(E_i) = 1 - \alpha$$

Denotaremos por \bar{E}_i al complementario de E_i , es decir, \bar{E}_i es el suceso que indica que el intervalo de confianza no contiene al parámetro. La probabilidad de la intersección de que todos y cada uno de los parámetros estén cubiertos por su correspondiente intervalo verifica:

$$P(\cap_{i=1}^k E_i) = 1 - P(\overline{\cap_{i=1}^k E_i}) = 1 - P(\cup_{i=1}^k \bar{E}_i) \geq 1 - k\alpha \quad (6)$$

La expresión (6) se conoce como *desigualdad de Bonferroni*. Se convierte en una igualdad si los \bar{E}_i son excluyentes.

La confianza simultánea deseada se alcanza tomando confianzas individuales más elevadas. Si tenemos k parámetros para estimar es habitual considerar intervalos individuales de nivel $1 - \alpha/k$. Esto significa que la probabilidad de que el intervalo no cubra al parámetro es α/k . Aplicando la desigualdad en (4) se obtiene una confianza global igual o superior a $1 - \alpha$.

3.10. Modelo de regresión Weibull

Para analizar el posible efecto de un factor o variable bajo control en una variable respuesta, X ; se utilizan modelos de regresión. Dado que la distribución normal no es un habitual en tiempos hasta el fallo, en el contexto de la fiabilidad se utilizan modelos de regresión Weibull. En este caso la hipótesis de varianza constante de la regresión lineal se transforma en el supuesto de que

el parámetro de forma no cambia en los distintos valores de la variable explicativa. El modelo de regresión está dado sobre el logaritmo del parámetro de escala, de modo que, cuando $X = x$, se tiene el siguiente modelo log-lineal:

$$\ln a_x = b_0 + b_1 x$$

Es decir, la función de fiabilidad en el modelo de regresión Weibull es:

$$S(t, x) = e^{(-t/a_x)^\beta}$$

El estudio de los residuos en el modelo de regresión lineal se transforma, en el caso de la regresión Weibull en el análisis de los residuos de Cox-Snell.

3.11. Residuos de Cox-Snell

Una vez que se ha estimado el modelo, es indispensable juzgar su adecuación a los datos de la muestra.

Los residuos que se consideran en modelos de regresión es supervivencia tienen en cuenta también la presencia de la censura.

Sea T el tiempo de fallo, con función de supervivencia $S(t)$. Entonces, la variable aleatoria $S(T)$, se distribuye como una uniforme $[0, 1]$. Para su demostración basta probar que $U = F(T)$ se distribuye como una uniforme $[0, 1]$.

$$P(U \leq u) = P(F(T) \leq u) = P(T \leq F^{-1}(u)) = F(F^{-1}(u)) = u$$

Por lo tanto

$$S(T) = 1 - U \Rightarrow S(T) \equiv U[0, 1]$$

Lema: Al aplicar la transformación $-\ln$ sobre una distribución uniforme se obtiene una exponencial de parámetro 1.

Demostración: Sea U una variable aleatoria uniforme en $(0, 1)$ y sea $W = -\ln(U)$. La función de distribución de W verifica:

$$P(W \leq w) = P(-\ln(U) \leq w) = P(\ln(U) \geq -w) = P(U \geq e^{-w}) = 1 - e^{-w}$$

En la última igualdad se ha utilizado la expresión de la función de supervivencia de una variable aleatoria uniforme.

Además, como el rango de U es $(0, 1)$ entonces el rango de $W = -\ln(U)$ es de $(0, \infty)$.

Por consiguiente $W = -\ln(U)$ tiene una distribución exponencial de parámetro 1.

QED

Por lo que

$$W = -\ln(U) \equiv \exp(1)$$

Sean t_1, t_2, \dots, t_n los instantes de fallo o censura en la muestra. Si el modelo es correcto, la función de riesgo acumulada estimada de cada individuo, en t_1, t_2, \dots, t_n , debe ser una muestra censurada de una distribución exponencial con parámetro uno. Esta cantidad es conocida como el residuo Cox-Snell.

$$r_{c.s_i} = -\ln(S(t_i, x_i, \hat{\alpha}_{x_i}\hat{\beta}))$$

donde $\hat{\alpha}_{x_i}, \hat{\beta}$ son los parámetros de escala y forma estimados en el modelo respectivamente.

4. Descripción de la máquina. Funcionamiento y tipos de fallo

Cashlogy es una máquina diseñada para el comercio cuya misión es gestionar los pagos en efectivo en el punto de venta, automatizando todos los procesos.

La determinación de la fiabilidad de la máquina se ha realizado analizando los registros de incidencias y reparaciones efectuadas sobre una muestra de **48** máquinas monitorizadas desde su puesta en funcionamiento hasta la fecha del mantenimiento preventivo.

Además del momento en el que se produce la incidencia, la empresa ha facilitado una descripción de la máquina, de su funcionamiento y del tipo de fallo. Esta información adicional ha servido para realizar comparativas relativas a la fiabilidad de la máquina atendiendo a características del negocio o seleccionando determinado tipo de incidencias.

Los fallos registrados no tienen carácter catastrófico, es decir, no requieren del cambio completo de la máquina, sino que, bien mediante una atención telemática o mediante el envío de un técnico especializado, el fallo es reparado devolviendo la máquina al estado inicial.

La estimación de la tasa de fallos a lo largo del tiempo resulta esencial para el establecimiento de

políticas de mantenimiento preventivo de la máquina.

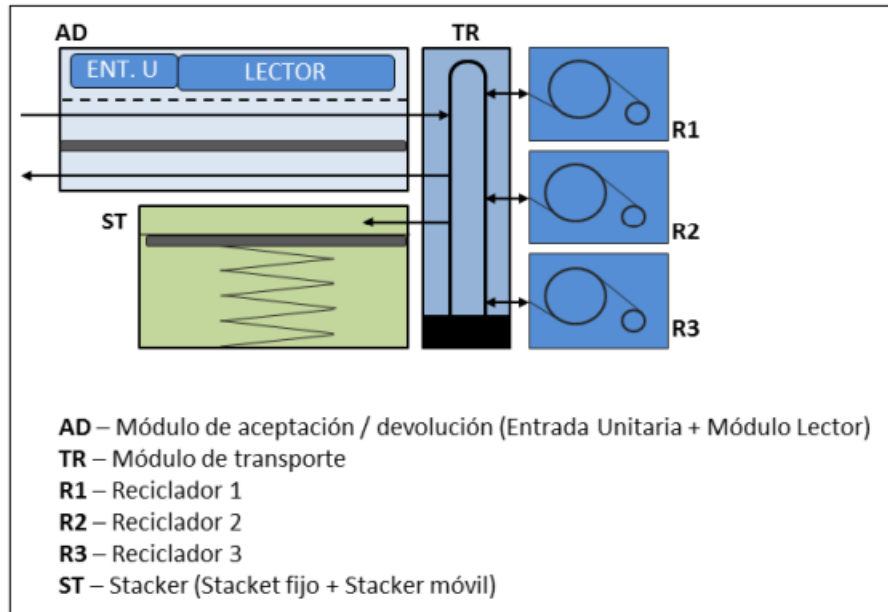


Figura 6: Diagrama de flujo de la parte de billetes

El principal objetivo del análisis de datos que se presenta a continuación es estimar la evolución temporal de la tasa a la que se producen estos incidentes. Para ello, utilizaremos modelos y técnicas estadísticas para el análisis de fiabilidad, tanto de tipo paramétrico como no paramétrico.

Estos análisis se han llevado a cabo utilizando el software Minitab 19, que tiene implementados todos los procedimientos de fiabilidad necesarios para realizar los análisis.

Por último cabe señalar que, si bien a lo largo de la memoria (y en general en fiabilidad) se habla de tiempo hasta el fallo, este tiempo no corresponde a una medida temporal expresada en términos de unidades de calendario, como horas o días, sino al número de transacciones ya realizadas por la máquina, desde su instalación, hasta que se registra dicho fallo.

Es previsible que las máquinas en establecimientos con un elevado número de transacciones fallen antes que las instaladas en locales de actividad media o baja. Así puede producirse con anterioridad el fallo de una máquina altamente fiable con una elevada carga de trabajo que el de otra de menor fiabilidad pero que esté sometida a un menor nivel de esfuerzo. Por este motivo el número de transacciones hasta el fallo resulta ser la variable respuesta adecuada en lugar del tiempo que transcurre hasta que aquél tiene lugar.

La máquina se compone de dos módulos claramente diferenciados: parte de billetes y parte de

monedas.

4.1. Parte de billetes

La Figura 6 presenta la disposición de los componentes en el módulo de billetes y el flujo que estos recorren dependiendo su destino final.

Entrada unitaria: Módulo a través del cual se introduce el billete realizándose una validación previa del mismo, mediante sensores infrarrojos, así como un ajuste de su posición tras lo cual el billete queda perfectamente alineado para ser admitido.

Módulo lector: Se encarga de medir el billete y validarlo. Compuesto por dos zonas, aceptación y rechazo, dónde la admisión es la parte superior por la cual entra el billete y el rechazo la parte inferior por la cual sale el billete en caso de devolver o ser rechazado. El módulo lector dispone de una serie de sensores para determinar el tipo de billete y aceptarlo o rechazarlo. En caso de atasco, gracias a estos sensores, se puede determinar la posición del billete.

Módulo de transporte: Situado verticalmente entre los recicladores y el módulo de admisión, es el encargado de transportar los billetes desde cada uno de los módulos a su destino.

Reciclador: Lugar de almacenamiento con capacidad de hasta 47 billetes y con capacidad de devolverlos. Cada reciclador está asignado a una denominación monetaria concreta. Como hay tres recicladores, por defecto, éstas serán de 5, 10 y 20€.

Stacker: es el componente de la unidad de billetes que se encarga de almacenar los billetes. Para ello consta de dos piezas, el stacker fijo y el móvil. El primero se encarga de introducir los billetes dentro de la parte móvil.

4.2. Parte de monedas

La Figura 7 muestra el recorrido que hace una moneda y su correspondiente clasificación.

Bulk de admisión (K1): Alimentador de monedas capaz de trabajar con todas las denominaciones monetarias. El bulk tiene una capacidad de unas 40 monedas.

Módulo de admisión de monedas (H2): Tras el bulk, las monedas atraviesan a continuación el módulo H2. Este módulo es el que controla electrónicamente el proceso de validación y clasificación de monedas.

Guia monedas: Como su nombre indica, lleva las monedas a la zona de elevación si han sido

aceptadas por el H2 o a la salida de monedas si han sido rechazadas.

Elevación: Introduce las monedas validadas al sistema de clasificación.

Clasificación: Las monedas recorren un clasificador pasivo compuesto por una galga metálica que diferencia las monedas por diámetro y un conjunto de pisadores unidos a una tarjeta electrónica que verificará que cada moneda sea correctamente clasificada.

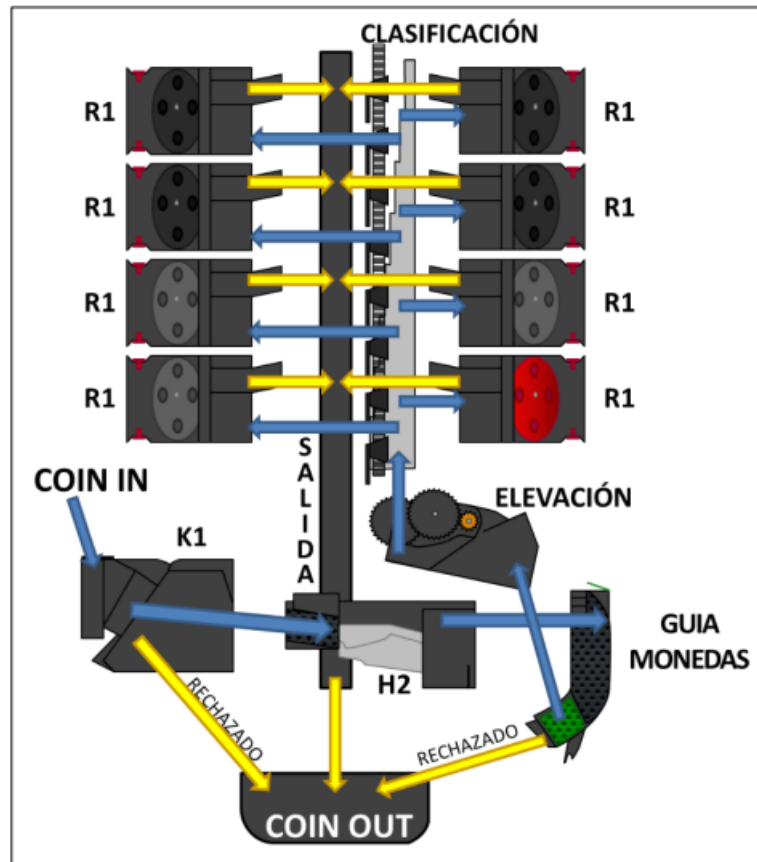


Figura 7: Diagrama de flujo de monedas

En este trabajo se van a estudiar los 6 tipos de fallo de la máquina que ocurren más frecuentemente y que son responsables de la mayor parte de las pérdidas de disponibilidad para los clientes:

- Atasco en lectura
- Atasco en monedas
- Entrada unitaria y lector
- Reciclador

- Stacker
- Transporte

5. Tratamiento inicial del fichero de datos

El fichero de datos contine una fila por cada uno de los fallos registrados, hasta el día 31 de Julio que es el último día de observación. Esta fecha se considera como fecha de censura para cada una de las máquinas monitorizadas, esto es, el siguiente fallo en cada una de ellas ocurrirá con un número de transacciones superior al registrado en dicha fecha.

El fichero de datos contiene información de las siguientes variables

- **ID:** Es el número de identificación de la máquina.
- **Nombre:** El nombre del establecimiento donde está instalada la máquina.
- **Tipología de negocio:** Tipo de negocio donde está instalada la máquina.
 - **Bar/Cafetería Horeca:** Horeca es una abreviación de hoteles, restaurantes y cafeterías.
 - **Carnicería/Pescadería/Frutería**
 - **Estanco.**
 - **Gasolinera.**
 - **Heladería**
 - **Panadería/Pastelería.**
 - **Restaurante Horeca.**
- **Atendida/Semiatendida:** Variable binaria que indica si la máquina esta atendida o semi-atendida. Entendemos como atendida a una máquina que solo la usa el usuario que tiene una formación (camareros, dependientes..) y semiatendida si está expuesta al público en general.
- **Censura:** Variable binaria en la que ponemos un 1 si ha habido incidencias y un 0 en la máquina una vez terminado el proceso de observación.
- **Transacciones:** Número de transacciones que ha hecho una cierta máquina desde una incidencia a otra.

- **Días:** Número de días que han pasado desde una incidencia a otra (con respecto a la misma máquina).
- **Fecha:** Fecha en la que se produce la incidencia.
- **Hora:** Hora en el que se produce la incidencia.
- **Tipo de fallo:** Indicador del componente que ha fallado (entre los 6 estudiados).
- **Excluir:** Variable binaria en la que aparece un SI, si la empresa opta por prescindir de esa incidencia para el estudio o un NO si la empresa cree oportuno incluir esa incidencia en el estudio.

5.1. Depuración de datos y análisis descriptivo de las variables

En primer lugar se realiza un análisis de las 57 máquinas que forman el estudio.

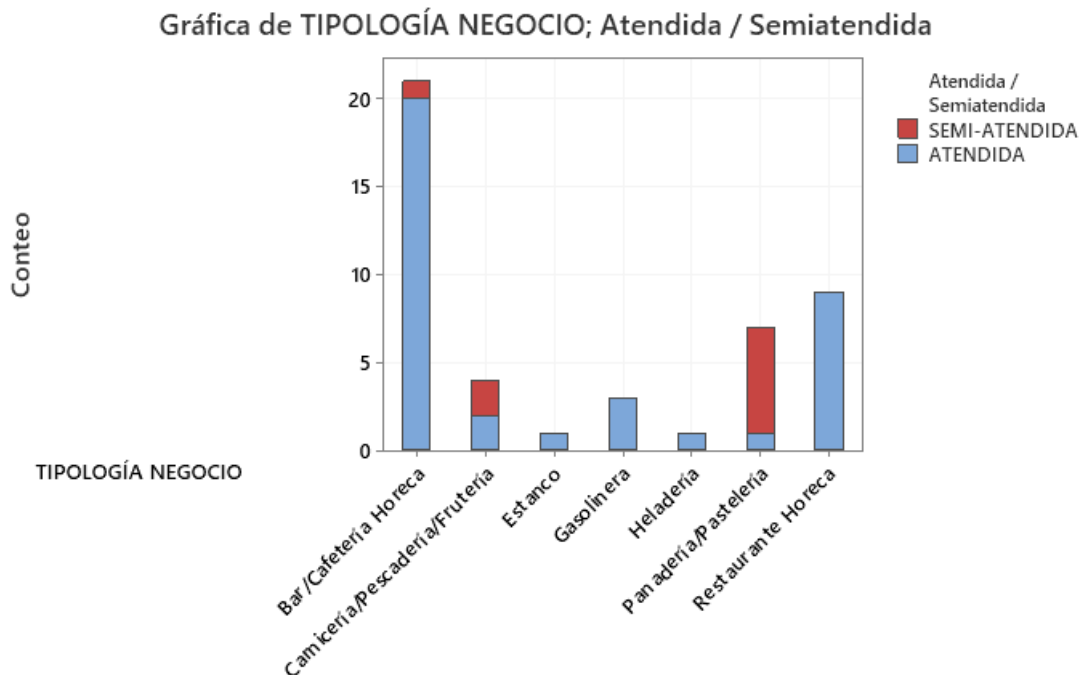


Figura 8: Análisis descriptivo de las máquinas

En la Figura 8 se advierte que salvo las panaderías y pastelerías, las máquinas están mayoritariamente atendidas. Esto significa que no están expuestas al público en general y solo son

manipuladas por el dueño o el trabajador del establecimiento. En el caso de los estancos, gasolineras y restaurantes el 100 % están atendidas.

Una hipótesis de partida es que la materia prima utilizada en heladerías puede ser especialmente adversa para el funcionamiento de la máquina. La primera cuestión que surgió durante el proceso de depuración de los datos es que tan sólo una de las máquinas de prueba estaba situada en heladerías. No es suficiente para asegurar que los fallos observados se deban a los materiales de uso habitual en heladerías y no a este establecimiento en particular. Así optamos por eliminar del estudio los datos referentes a la heladería. Un caso similar se ha producido con el único estanco que había en la muestra. Su falta de representatividad ha forzado su salida de la muestra,

5.2. Estudio cruzado: tipo de fallo por establecimiento

Estadísticas tabuladas: Nombre del fallo; Tipología de negocio

Filas: Nombre del fallo Columnas: Tipología de negocio

	Bar/Cafetería				Restaurante	
	Horeca	Carnicería/Pescadería/Frutería	Gasolinera	Panadería/Pastelería	Horeca	Todo
ATASCO EN LECTURA	108	2	5	14	48	177
	91,998	5,597	4,547	31,132	43,725	
ATASCO MONEDAS	70	12	5	62	45	194
	100,834	6,134	4,984	34,123	47,925	
ENTRADA UNITARIA & LECTOR	12	2	1	7	8	30
	15,593	0,949	0,771	5,277	7,411	
RECICLADOR	14	0	1	3	5	23
	11,955	0,727	0,591	4,045	5,682	
STACKER	35	0	0	2	10	47
	24,429	1,486	1,208	8,267	11,611	
TRANSPORTE	24	0	1	1	9	35
	18,192	1,107	0,899	6,156	8,646	
Todo	263	16	13	89	125	506

Contenido de la
celda
Conteo
Conteo
esperado

Figura 9: Tabla cruzada tipo de fallo por tipo de negocio

En la Figura 9 se presenta la distribución de los diferentes fallos en los distintos de negocio. Se advierte que, por ejemplo, el fallo atasco en lectura es el más usual en bares y cafeterías, mientras que el atasco de monedas es el más frecuente en panaderías. Cabe pensar que en los primeros se haga un mayor uso de billetes.

La Figura 9 contiene asimismo la frecuencia esperada de fallos si las dos variables, tipo de fallo y tipo de negocio, fueran independientes. Sin embargo el elevado número de combinaciones con frecuencia esperada inferior a 5 invalida la correspondiente prueba de independencia a partir del test de la chi-cuadrado.

5.3. Estudio de la variable respuesta

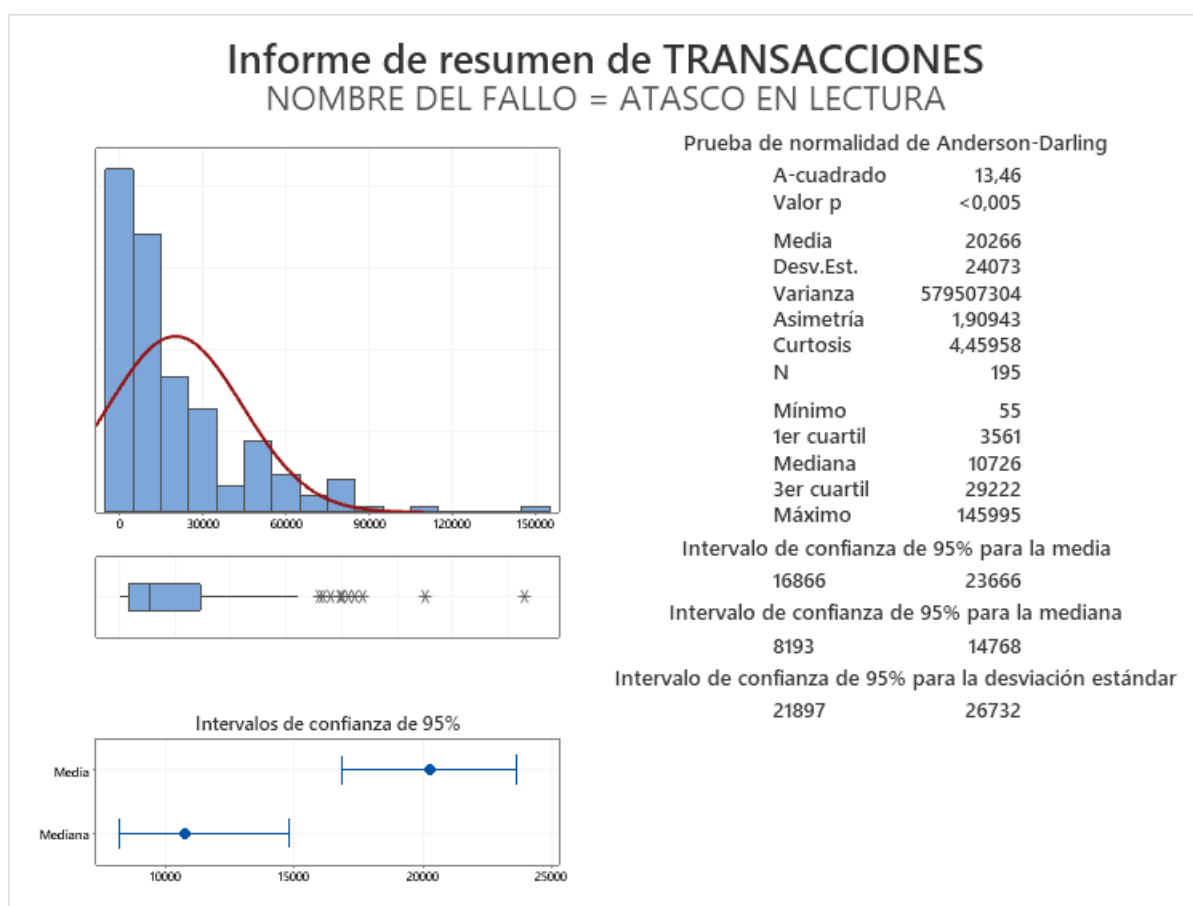


Figura 10: Informe gráfico del atasco de monedas

En la Figura 10 se realiza el análisis estadístico descriptivo del número de transacciones hasta el fallo de atasco en lectura. Se trata de una distribución asimétrica positiva y por tanto no puede tratarse de una distribución normal tal y como se concluye a partir del p-valor (< 0.05)

Realizando un informe similar para todos los tipos de fallo, las gráficas de caja correspondientes presentan en todos los casos varios datos atípicos y que están asociados a las máquinas 13039, 4007, 6582 y 1471.

La elevada frecuencia de fallos llevó a tomar la decisión de cambiar el componente, sustituyéndolo por uno nuevo. Esta circunstancia ha tenido un efecto en el análisis posterior pues se han estimado tasas de fallo decrecientes como las que se describen en la Figura 1 para la tasa de fallo en forma de *curva de bañera*. Como ya se indicó esas tasas de fallo decrecientes surgen como consecuencia de procedimientos de *burn-in* como el que ha realizado los técnicos de la empresa en este caso.

El único fallo que sigue una distribución normal ($p\text{-valor} > 0.05$) es el reciclador. En este caso tampoco se advierte ningún dato atípico y, de hecho, se trata de una componente que no se ha cambiado en ninguna máquina.

6. Estudio de las diferencias entre tipos de fallo

En esta sección analizaremos si el número de transacciones hasta el fallo es diferente de unos tipos de fallo a otros. Si denotamos por $S_1(t), S_2(t) \dots S_6(t)$ a las funciones de supervivencia correspondientes a los fallos de atasco en lectura, atasco en monedas, entrada unitaria y lector, reciclador, stacker y transporte, respectivamente. El estudio se realizará mediante la siguiente prueba de hipótesis.

$$H_0: S_1(t) = S_2(t) = S_3(t) = S_4(t) = S_5(t) = S_6(t).$$

$$H_1: \text{Alguna } S_i(t) \text{ es diferente.}$$

6.1. Análisis no paramétrico

Empezaremos con un análisis no paramétrico, es decir sin hacer ningún supuesto sobre la distribución de los datos. Los contrastes están basados en la comparación de los estimadores de Kaplan-Meier obtenidos para los distintos tipos de fallo. Minitab utiliza dos tipos de tests para evaluar si las funciones de supervivencia difieren significativamente: la prueba de log-rango y la de Wilcoxon.

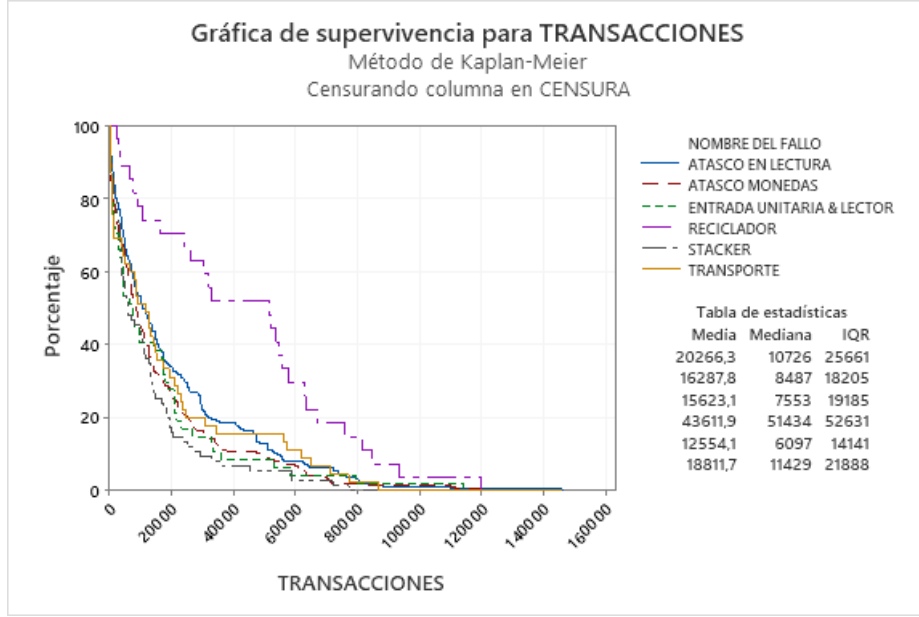


Figura 11: Gráfica de supervivencia por tipo de fallo

En la Figura 11 se presentan los estimadores de Kaplan-Meier correspondientes a los 6 tipos de fallo bajo estudio. Se advierte que el correspondiente del reciclador refleja supervivencias más elevadas. Los contrastes revelarán si la diferencia con el resto de fallos es significativa.

Prueba de log-rango: La prueba de log-rango compara el número real y esperado de fallos entre las curvas de supervivencia en cada tiempo de fallo.

$$\sum_{j=1}^r d_{1j} - e_{1j}$$

con

- d_{1j} es el número de fallos en el grupo 1 en el tiempo j .
- e_{1j} es el número esperado de eventos en el grupo 1 en el tiempo j .
- r es el total de los tiempos únicos en todos los grupos.

Prueba de Wilcoxon: La prueba de Wilcoxon es una prueba de log-rango ponderada por el número de elementos que aún sobreviven en cada instante temporal. Puesto que la prueba de Wilcoxon pondera los tiempos de fallos tempranos con mayor intensidad, la prueba es menos sensible que la del log-rango a las diferencias que ocurren en tiempos de fallo elevados.

Interpretación: El estadístico chi-cuadrado se calcula elevando al cuadrado el estadístico de log-rango y dividiendo entre la varianza estimada. Un valor $p < \alpha$ indica que las curvas de supervivencia

son significativamente diferentes. Como se obtiene $p = 0$, se rechaza H_0 y se acepta H_1 . Es decir los datos proporcionan evidencia de que alguna de las curvas de supervivencia es diferente. En la inspección previa de la Figura 11, se había constatado el diferente comportamiento del reciclador, que de hecho, tiene una fiabilidad más elevada que el resto de componentes.

6.2. Análisis paramétrico

En lo que sigue se hace un análisis basado en la comparación de los parámetros de una distribución. Para ello es preciso encontrar una distribución común que ajuste a los datos de los diferentes tipos de fallos. Después realizaremos contrastes e intervalos de confianza simultáneos que permitan establecer si los parámetros presentan diferencias significativas dependiendo del tipo de fallo.

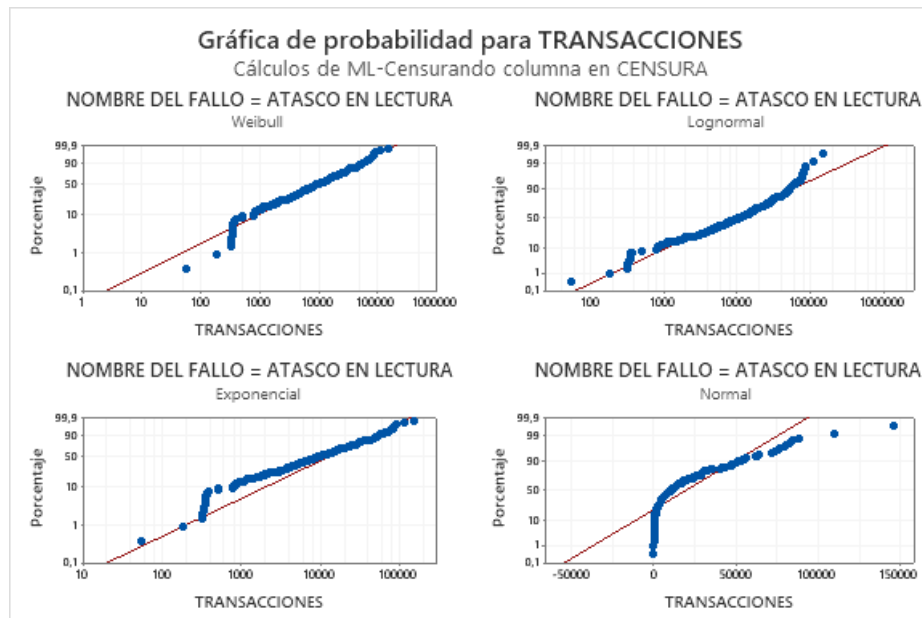


Figura 12: Gráfica de probabilidad para el fallo atasco en lectura

La Figura 12 muestra que el mejor ajuste en el fallo de atasco en lectura lo proporciona la distribución Weibull. La Figura 13 presenta el ajuste Weibull con las estimaciones de los parámetros de forma y escala.

El resultado del estudio es que la distribución Weibull es la que mejor se ajusta a los datos en los distintos tipos de fallo salvo en el caso del reciclador que se ajusta mejor por una normal. La conclusión respecto a la Weibull no es de extrañar si recordamos los efectos de los cambios en los parámetros de forma y escala de la sección 3.4 que dan una gran versatilidad a esta variable. En los ajustes Weibull el parámetro de forma es en todos los casos inferior a 1. Esto quiere decir que

la máquina mejora conforme pasa el tiempo; es decir, tiene menos fallos. Este efecto es el reflejo de la mejora de la fiabilidad que se ha producido al cambiar los componentes en máquinas que presentaban reiteradamente algún tipo de fallo.

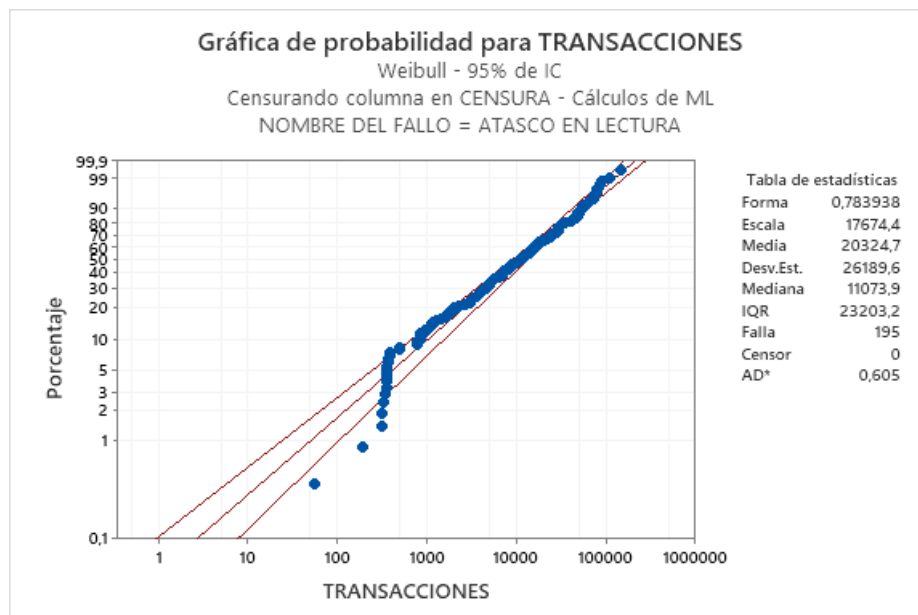


Figura 13: Atasco de monedas siguiendo una distribución Weibull

Nombre del fallo	Parámetro de forma	Parámetro de escala
Atasco en lectura	0.7839	17674
Atasco en monedas	0.7010	13029
Entrada unitaria y lector	0.6930	12204
Stacker	0.7925	10982
Transporte	0.6515	14295

Tabla 1: Parámetros de la distribución Weibull por tipo de fallo

Los parámetros estimados para la distribución Weibull en los distintos tipos de fallo se presentan en la Tabla 1.

Los tiempos de fallo correspondientes al reciclador y cuya distribución, en cambio, se ajusta a una normal presentan tasa de fallos creciente. Este hecho se explica porque no se ha realizado ningún cambio de reciclador en todo el estudio.

6.3. Modelos de regresión Weibull en Cashlogy

Un objetivo central del trabajo era determinar si había diferencia de comportamiento entre los distintas categorías de fallo (F) y diferentes tipo de negocio (N). Este estudio se va a llevar a cabo

mediante un modelo de regresión Weibull que determinará si ambos factores F y N afectan al número de transacciones hasta el fallo.

En la sección 6.2 se ha advertido la diferencia en el comportamiento del reciclador respecto del resto de componentes. Esta circunstancia nos lleva a separar el fallo del reciclador de los demás, analizando dos modelos de regresión. En el primero consideraremos todas las combinaciones de tipo de negocio y fallo excluyendo el reciclador. Así se propone el siguiente modelo sobre el parámetro de escala que puede verse afectado por las categorías de fallo y tipo de negocio:

$$\ln a(F, N) = b_0 + b_1 F + b_2 N$$

Tabla de regresión

Predictor	Coef	Error estándar	Z	P	IC normal de 95,0%	
Intersección	9,86280	0,113207	87,12	0,000	9,64091	10,0847
TIPOLOGÍA NEGOCIO						
Carnicería/Pescadería/Frutería	-0,502412	0,339926	-1,48	0,139	-1,16865	0,163829
Gasolinera	1,14029	0,388669	2,93	0,003	0,378514	1,90207
Panadería/Pastelería	0,163967	0,166267	0,99	0,324	-0,161910	0,489844
Restaurante Horeca	0,0132231	0,145672	0,09	0,928	-0,272289	0,298735
Nombre del fallo						
ATASCO MONEDAS	-0,157999	0,138138	-1,14	0,253	-0,428744	0,112747
ENTRADA UNITARIA & LECTOR	1,36112	0,261607	5,20	0,000	0,848381	1,87386
STACKER	0,844333	0,216520	3,90	0,000	0,419962	1,26870
TRANSPORTE	1,15352	0,244376	4,72	0,000	0,674553	1,63249
Forma	0,763591	0,0280179			0,710605	0,820528

Log-verosimilitud = -5346,540

Figura 14: Modelo de regresión sin reciclador

En la Figura 14 se presentan los resultados del primer modelo de regresión. Es importante destacar que el hecho de que la máquina esté atendida o semiatendida no genera diferencias (p-valor >0.05) por lo cual se ha eliminado este factor de la ecuación. Tomando como referencia bares/caféterías los p-valores indican si el parámetro de escala en el nivel asociado es significativamente diferente del escogido como nivel referencia.

Atendiendo al p-valor se ve que los únicos establecimiento que se comportan de distinto modo con respecto a bares y cafeterías son las gasolineras. Además el coeficiente asociado a las gasolineras es positivo. Este hecho implica que el logaritmo del parámetro de escala es más elevado en este último caso y por tanto, también lo es el propio parámetro de escala. Así, el logaritmo del parámetro de

escala en las gasolineras es $9.86280 + 1.14029$.

Por consiguiente la máquina en las gasolineras presenta una fiabilidad más elevada que en el resto de establecimientos. Una posible explicación es que los empleados suelen ponerse guantes cuando prestan un servicio y posiblemente manejen la máquina con las manos limpias, lo cual incide positivamente en su funcionamiento. En cuanto al tipo de fallo se toma como referencia el atasco en lectura. Los p-valores llevan a la conclusión de que no hay diferencias en el número de transacciones hasta el fallo de atasco en lectura y atasco en monedas. Los otros tres tipos de fallo, entrada unitaria y lector, stacker y transporte sí que son diferentes del tomado como nivel de referencia. Los coeficientes positivos indican que estos tres tipos de fallo tardan más tiempo en ocurrir que los dos tipos de atasco.

Un estudio posterior nos permitirá establecer si el número de transacciones hasta la entrada unitaria y lector, stacker y transporte no son significativamente diferentes lo cual permitirá proponer una política de mantenimiento común para los tres.

El coeficiente de la intersección se corresponde con el logaritmo del parámetro de escala para la combinación de ambos niveles de referencia, es decir, el atasco en lectura en bares/cafeaterias cuyo valor es 9.86280.

La Figura 15 presenta los residuos de Cox-Snell en el primer modelo de regresión, presentando un aceptable ajuste a la distribución exponencial lo cual da validez al modelo propuesto.

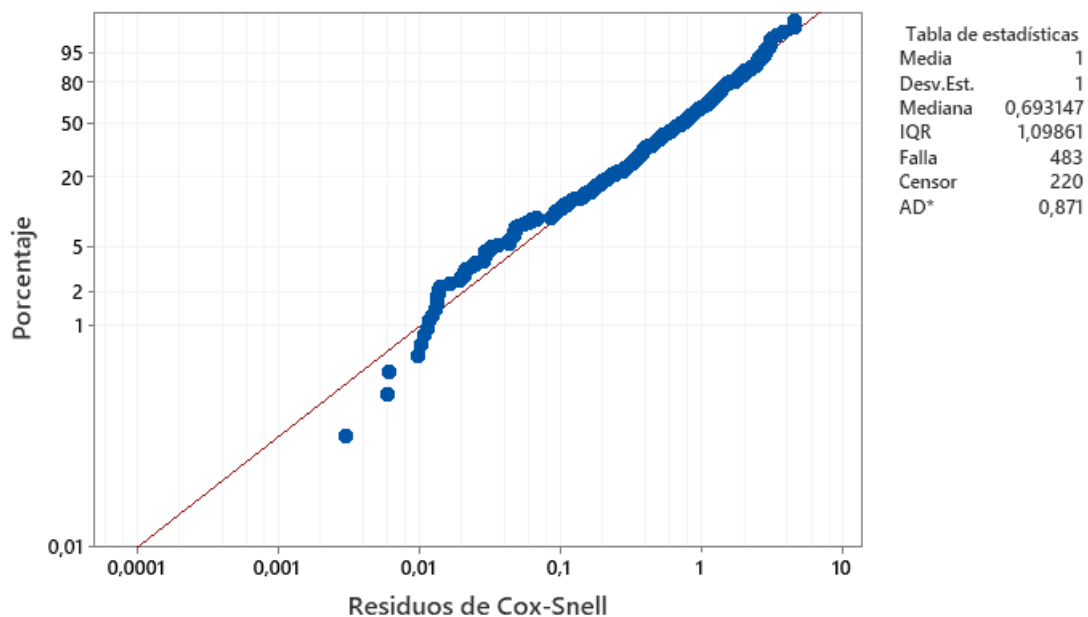


Figura 15: Gráfica de residuos de Cox-Snell. Modelo 1

El segundo modelo tiene por objeto determinar si el comportamiento del reciclador varía en los distintas categorías de negocio (N). En este caso se asume que la normal es la distribución base con una desviación típica común a todos los casos, mientras que la media puede verse afectada por el factor:

$$\mu(N) = c_0 + c_1N$$

Se trata, por tanto de los supuestos del modelo de regresión lineal, sobre el que se añade la presencia de datos censurados.

Tabla de regresión

Predictor	Coef	Error estándar	Z	P	IC normal de 95,0%	
					Inferior	Superior
Intersección	65514,1	7830,60	8,37	0,000	50166,4	80861,8
TIPOLOGÍA NEGOCIO						
Carnicería/Pescadería/Frutería	155023	172817490	0,00	0,999	-338561033	338871079
Gasolinera	26400,2	27321,3	0,97	0,334	-27148,6	79949,1
Panadería/Pastelería	-961,307	15915,6	-0,06	0,952	-32155,3	30232,7
Restaurante Horeca	-6327,80	13244,9	-0,48	0,633	-32287,3	19631,7
Escala	33652,7	4762,48			25501,2	44410,0

Log-verosimilitud = -286,358

Figura 16: Modelo de regresión del reciclador

Con los resultados de la la Figura 16 se concluye que el valor medio del número de transacciones de los fallos del reciclador no depende del tipo de negocio, estimándose en 33652.7 la desviación típica común del número de transacciones hasta el fallo. Esto significaría puede utilizar la misma distribución normal para todos ellos. Sin embargo el análisis detallado de la Figura 16 resulta sorprendente al observar el coeficiente asociado a las carnicerías/pescaderías/fruterías (1555023), que es muy superior al de referencia (65514.1) correspondiente a bares y cafeterías. La causa de este paradójico resultado es que en las carnicerías todos los datos son censurados lo cual lleva a una estimación forzada del coeficiente. Un hecho similar ocurre en el caso de las gasolineras en las que sólo había un único dato de fallo frente a tres correspondientes a censuras. De hecho, no es posible obtener, en ninguno de los dos casos, una estimación aislada de la media y de la desviación típica. Por este motivo la conclusión de que el tiempo hasta el fallo del reciclador no depende del tipo de establecimiento sólo tiene validez en el caso de bares y cafeterías, panaderías y restaurantes. Tanto

sector de la alimentación (carnicerías/pescaderías/fruterías) como las gasolineras quedarían fuera de esta conclusión.

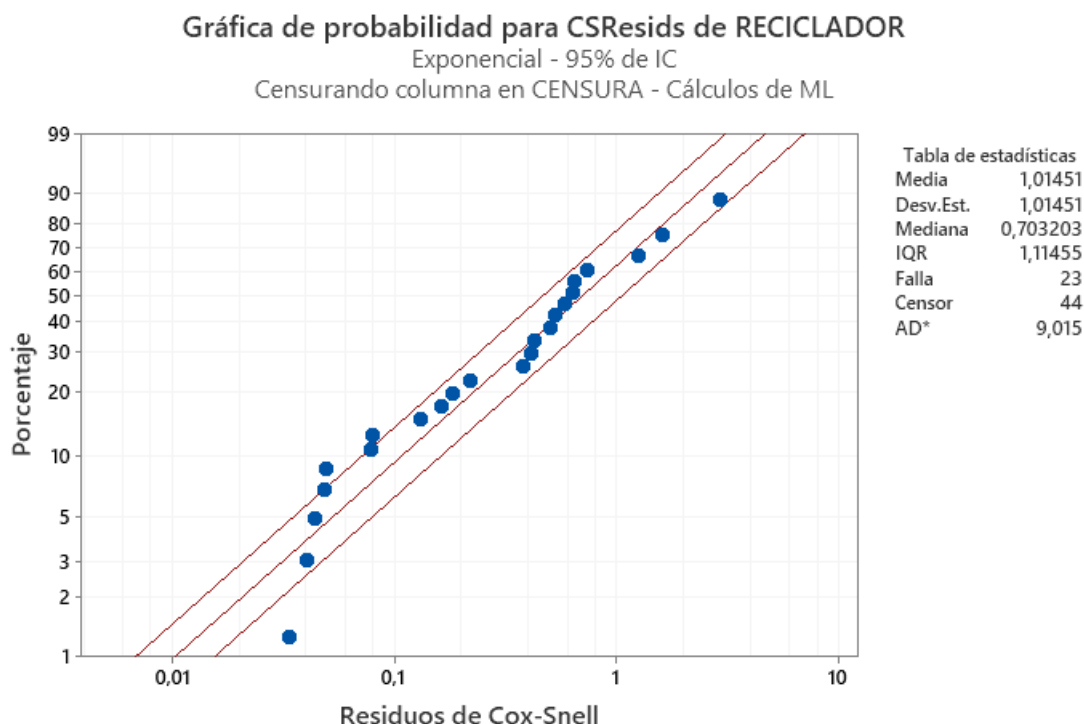


Figura 17: Gráfica de residuos de Cox-Snell. Modelo 2

La gráfica de los residuos de Cox-Snell en la Figura 17, presenta un ajuste razonable a la distribución exponencial con la mayoría de los residuos dentro de las bandas de confianza. Este resultado permite dar por válidos los supuestos del modelo, es decir normalidad con varianza constante. En cualquier caso sería recomendable hacer un estudio adicional considerando los fallos del reciclador en una muestra de establecimientos de mayor tamaño, durante un periodo más prolongado. El análisis estadístico ha revelado que el reciclador tiene una duración más elevada que el resto de componentes y por lo tanto se requieren tiempos más largos de observación para poderla estimar con precisión.

A continuación determinaremos la posible igualdad de los parámetros de escala en los fallos entrada unitaria y lector, stacker y transporte mediante intervalos de confianza simultáneos cuya aplicación se describe en la Sección 3.9. Los resultados correspondientes se muestran en la Tabla 2. En todos los casos el valor 1 está dentro del intervalo, admitiéndose que el parámetro de forma es el mismo en todos los casos.

Respecto al transporte	forma	escala	Bonferroni forma	Bonferroni escala
Stacker	0.289	0.201	(0.79, 1.8)	(0.25, 1.45)
Entrada unitaria y lector	0.57	0.76	(0.69, 1.82)	(0.41, 3.2)

Tabla 2: Intervalos simultáneos por tipo de fallo

Con todos los resultados anteriores se van a distinguir cinco escenarios diferentes que se especifican en la Tabla 3. Un parámetro de escala más grande se corresponde con una duración media más elevada. Los resultados de la Tabla 3 cuantifican la mayor frecuencia de los fallos atasco en lectura y atasco en monedas frente a los fallos en transporte, stacker y entrada unitaria.

Nombre del fallo	Tipo de negocio	Parámetros de la función Weibull
Atasco en lectura/monedas	Todos excepto gasolineras	$\beta = 0,763591 \log(a) = 18265$
Atasco en lectura/monedas	Gasolineras	$\beta = 0,763591 \log(a) = 41087$
Transporte/stacker/E. unitaria	Todos excepto gasolineras	$\beta = 0,763591 \log(a) = 56861$
Transporte/stacker/E. unitaria	Gasolineras	$\beta = 0,763591 \log(a) = 453485$

Tabla 3: Parámetros de la función de distribución de los diferentes fallos

En relación al reciclador, la media y la desviación típica estimadas (en número de transacciones hasta el fallo) para la distribución normal son las siguientes:

$$\hat{\mu} = 65514, \quad \hat{\sigma} = 33652$$

6.4. Funciones de riesgo asociadas a cada tipo de fallo

En esta sección se presentan las gráficas de las funciones de riesgo de los diferentes fallos para analizar su monotonía.

En la Figura 18 se presenta la función tasa de fallos en el caso de atascos (en lectura y monedas). La Figura 19 contiene la relativa a los fallos en transporte, stacker y entrada unitaria, en tanto que la Figura 20 representa gráficamente la del reciclador. Las dos primeras son decrecientes, como consecuencia del procedimiento de *burn-in* que ha llevado a eliminar componentes defectuosas. Además se advierten tasas de fallo más bajas en gasolineras frente al resto de establecimientos. Como ya se indicó, el hecho de que los empleados de las primeras suelen utilizar guantes, puede explicar este comportamiento más favorable.

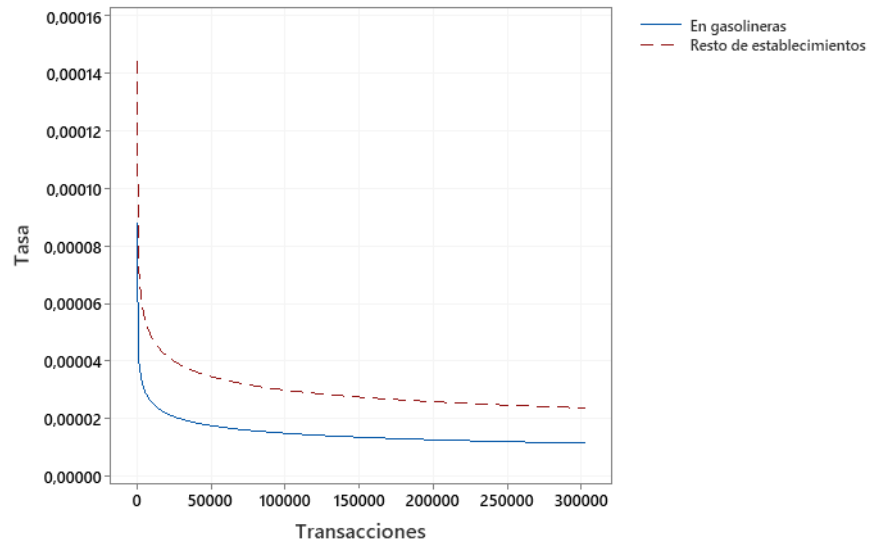


Figura 18: Función de riesgo del fallo por atasco en lectura y monedas

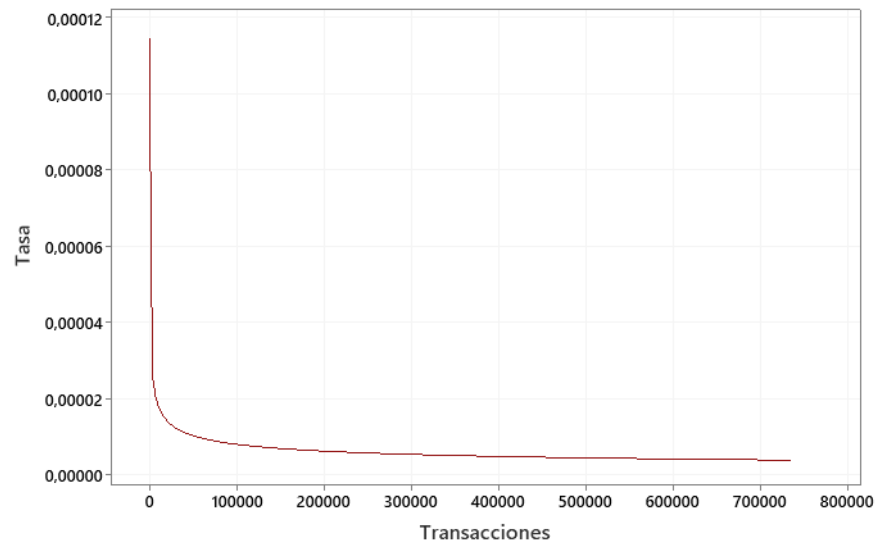


Figura 19: Función de riesgo para los fallos transporte, stacker y E. unitaria

En la Figura 20 se ve la tasa de fallos creciente en el reciclador, que no ha sido sometido a ninguna mejora respecto al que salió inicialmente de fábrica. No obstante hay que recordar que en el análisis no paramétrico se había advertido que sus fallos son más infrecuentes que los de las demás componentes pese a la mejora introducida en estas. El comportamiento creciente refleja el deterioro habitual por uso.

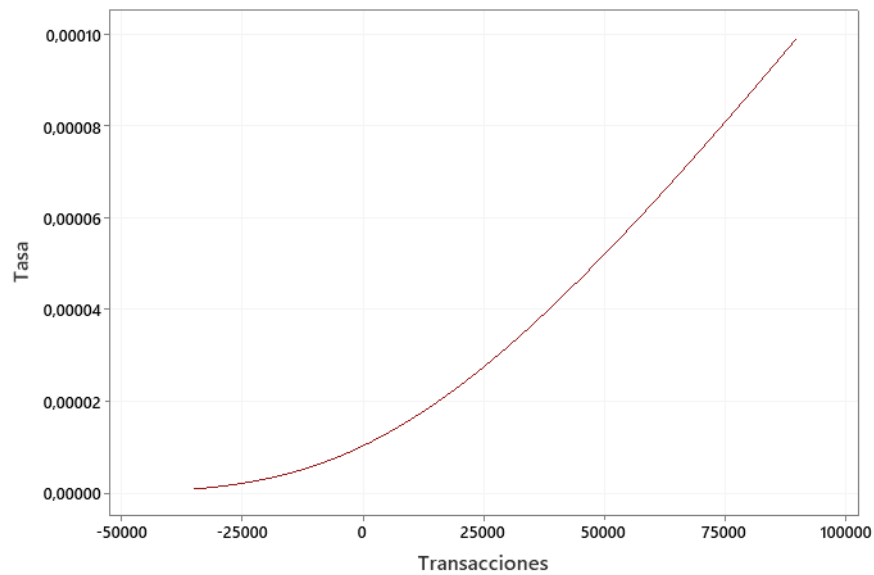


Figura 20: Función de riesgo del reciclador

7. Mantenimiento preventivo en el periodo de garantía

La empresa hace una clasificación de las máquinas en tres tramos dependiendo del número de transacciones diarias:

- 200 transacciones o menos al día
- entre 200 y 300 transacciones al día
- más de 300 transacciones al día.

Además un estudio interno señala que el 85 % de todas estas máquinas rondan las 250 transacciones diarias. Este dato lo tomaremos como referencia para el estudio

La empresa establece un periodo de garantía en dos dimensiones, por edad y uso. Esto significa que realizará, sin coste alguno para el cliente, las reparaciones de todos los fallos que se produzcan durante un periodo de tiempo fijado en 5 años o determinado por un número de transacciones, G , a establecer. Con el dato de 250 transacciones al día, suponiendo que una máquina trabaja 300 días al año y considerando el periodo máximo de garantía de 5 años, determinamos que la segunda dimensión de la garantía se establezca en $G = 375000$ transacciones. Por consiguiente, la empresa realizará, sin cargo al cliente, las reparaciones de todos los fallos durante 5 años o 375000 transacciones, lo que ocurra en primer lugar. Evitar que los fallos ocurran es un aspecto de la

máxima importancia. De una parte, la empresa gana en satisfacción del cliente y, por tanto, en credibilidad frente a sus rivales en el mercado. De otra, evita el gasto que le supone la reparación del fallo por un técnico especializado. Para evitar la ocurrencia de fallos se establecen políticas de mantenimiento preventivo. Estas consisten en realizar algún tipo de procedimiento sobre las componentes (limpieza, reemplazamiento...) antes de que fallen. La ventaja, frente a hacerlo cuando el fallo ya ha tenido lugar, es que el mantenimiento preventivo está programado, se sabe cuándo va a tener lugar mientras que el fallo ocurre aleatoriamente. En este último caso el técnico debe desplazarse exclusivamente para atender un único fallo. Esto supone una pérdida de tiempo para la empresa que tiene ocupado al técnico en traslados pero también para el cliente que puede tener la máquina inutilizada mientras tanto. En consecuencia para ambos, empresa y cliente, se genera un coste. En contraposición, el mantenimiento preventivo se puede realizar en varias máquinas de una misma zona o sector lo cual supone un mejor aprovechamiento del tiempo del técnico. Así la empresa añade un número de mantenimientos preventivos, N , a realizar durante el periodo de garantía. El nuevo objetivo es determinar el valor de N que resulte óptimo, es decir, que minimice el coste asociado para la empresa.

7.1. Función de coste de los fallos frente al número de transacciones.

Sea $N(t)$ el número aleatorio de fallos de una máquina en el intervalo $(0, t)$. En el caso que estamos estudiando, el valor t estará dado por un número de transacciones.

Denotaremos por C_i el coste aleatorio de la reparación del fallo i . En lo que sigue asumiremos que C_i , $i = 1, 2, \dots$ son una colección de variables aleatorias independientes de media c_c . El coste derivado de la reparación de los fallos hasta el instante t , $C(t)$, es una variable aleatoria dada por:

$$C(t) = \sum_{i=1}^{N(t)} C_i$$

Se trata, por tanto, de una suma aleatoria de variables aleatorias. El valor esperado de la función de coste anterior es:

$$E(C(t)) = E\left(\sum_{i=1}^{N(t)} C_i\right) = E(N(t)) E(C_i) = E(N(t)) c_c$$

7.2. Análisis de políticas óptimas de mantenimiento preventivo

En esta sección se analiza la frecuencia óptima de mantenimientos preventivos que minimiza el coste de mantenimiento de una máquina durante el periodo G de garantía.

Es decir, determinaremos el número óptimo N^* de mantenimientos preventivos que deberá realizar la empresa durante el periodo de garantía. Ese número deberá ser tal que evite la ocurrencia de fallos, resultando óptimo desde el punto de vista del coste, es decir, minimizando la función de coste de mantenimiento.

Asumiremos que la máquina después de un mantenimiento preventivo queda, desde el punto de vista de su fiabilidad, en un estado tan bueno como nueva, planteando estrategias de mantenimiento preventivo en intervalos de tiempo idénticos. Esto es, para realizar N mantenimientos preventivos durante el periodo G de garantía estos se realizan cada D unidades de tiempo. La relación entre D , G y N :

$$D = \frac{G}{N} \quad (7)$$

La ventaja de intervalos de tiempo iguales entre mantenimientos, es fundamentalmente que se simplifica su gestión. De hecho es la práctica habitual en las cosas sometidas a mantenimiento. Por ejemplo los cambios de aceite y filtros en un coche.

En cuanto a las reparaciones (mantenimientos correctivos) asumiremos que se realiza una reparación mínima (minimal repair), consistente en hacer que el sistema recupere su funcionalidad sin alterar su fiabilidad. Esto significa que su condición tras el mantenimiento correctivo será justo la que tenía antes de fallar.

El hecho de realizar una reparación mínima implica que los fallos siguen un proceso de Poisson no homogéneo de tasa $h(x)$ con $h(x)$ la tasa de fallo en el instante x , estimada en las secciones anteriores para las distintas categorías de fallo.

El número medio de fallos que tendrán lugar entre mantenimientos preventivos, es decir en intervalos de amplitud D , está dado por:

$$E(N(D)) = H(D) = \int_0^D h(u) du \quad (8)$$

A continuación se presenta la notación necesaria para construir la función de coste.

Notación:

- $c_c \equiv$ coste unitario del mantenimiento correctivo (reparación).
- $C(N) \equiv$ coste aleatorio total. Es decir derivado de los mantenimientos preventivos y correctivos durante el periodo de garantía.
- $G \equiv$ periodo de garantía de la máquina.
- $D \equiv$ tiempo entre mantenimientos preventivos.
- $N \equiv$ número de mantenimientos durante el periodo de garantía.
- $c_p \equiv$ coste unitario del mantenimiento preventivo.

Así, el coste total del mantenimiento durante el periodo de garantía es:

$$C(N) = N(c_c H(D) + c_p) \quad (9)$$

En el caso de la distribución Weibull se verifica: $h(t) = \beta \lambda^{-\beta} t^{\beta-1}$, con $\lambda = \frac{1}{a}$ y a el parámetro de escala.

A partir de (7) y (8) se tiene:

$$H(D) = \int_0^D h(u) du = -\ln(S(D)) = -\ln \int_D^\infty f(x, \hat{\mu}, \hat{\sigma}) dx$$

donde $f(x, \hat{\mu}, \hat{\sigma})$ es la función de densidad de la normal con media $\hat{\mu}$ y desviación típica $\hat{\sigma}$.

Así, la expresión de la función de coste en (9) resulta ser:

$$C_D(N) = N \left(c_c \left(-\ln \int_D^\infty f(x, \hat{\mu}, \hat{\sigma}) dx \right) + c_p \right) \quad (10)$$

El óptimo N^* verifica que

$$N^* = \operatorname{argmin}_N C_D(N)$$

La función $C_D(N)$ no es continua. Al no existir derivada, el proceso de optimización se realizará del modo siguiente:

1. Iniciar en $N = 1$
2. Mientras $C_D(N) > C_D(N + 1)$ pasar al paso 3.
3. Hacer $N = N + 1$

4. Tomar $N^* = N$

El valor N^* así obtenido es un óptimo local. Sin embargo, desde un punto de vista práctico es el que interesa puesto que está en la región de interés para la empresa y para el cliente. Valores mucho más elevados de N no son aplicables considerando el periodo de garantía, G ya que llevarían a un mantenimiento continuo.

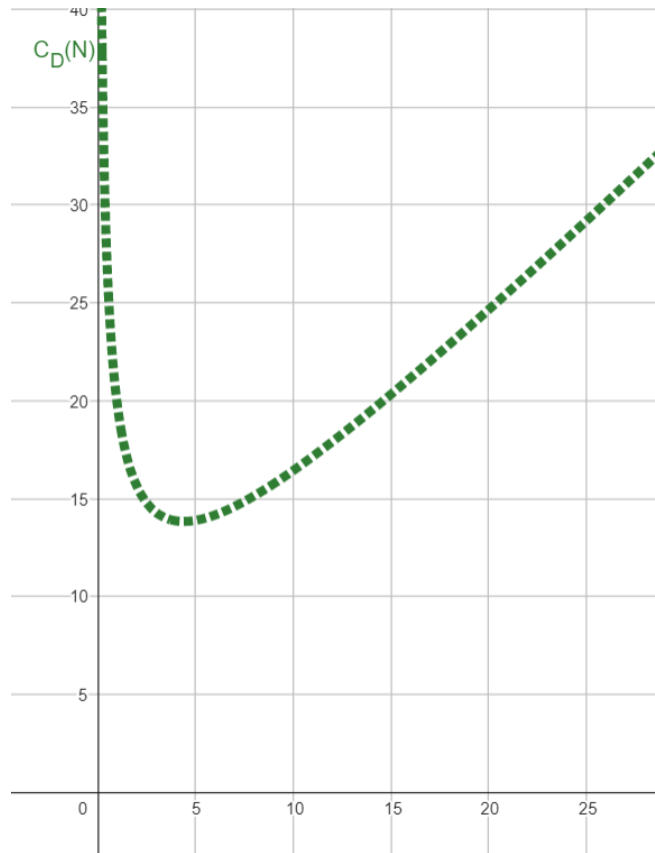


Figura 21: $C_D(N)$

En la figura 21 se representa la función de coste $C_D(N)$ en función de N . El óptimo se obtiene en $N^* = 4$ con un coste óptimo, $C_D(N^*)=13.7$. El incremento del coste que supone no aplicar el número óptimo de mantenimientos se recoge en la Tabla 4.

N^0 de mantenimientos	N^0 trans. entre mant. (D)	Valor función de coste
$N^* = 4$	93750	13.7
$N^* + 1 = 5$	62500	16.3
$N^* - 1 = 3$	125000	14.2

Tabla 4: Número de mantenimientos óptimos

Como se puede observar en la Tabla 4 si el número de mantenimientos es $N^* + 1$, entonces el

coste para la empresa se incrementa un 19 % respecto al número óptimo, N^* . La reducción en un mantenimiento ($N^* - 1$) es también más costosa para la empresa, además de que puede afectar de modo negativo a la satisfacción del cliente.

La función de coste se ha programado en lenguaje Python. El programa evalúa la función en una serie de valores enteros y guarda el valor para el cual se obtiene un mínimo en ese rango. Después se muestra por pantalla tanto el valor del coste como el número óptimo de mantenimientos. Su uso es de utilidad para la empresa ante un posible cambio de los costes.

7.3. Política de mantenimiento óptima para el reciclador

Aunque el mantenimiento preventivo implica el reemplazamiento de la pieza, el bajo coste de los respuestos hace que resulte menos elevado que el mantenimiento correctivo. Como ya se ha mencionado implica una paralización no planificada de la máquina así como la necesidad de que un técnico especializado tenga que ir exclusivamente a resolver ese problema.

Para obtener la política óptima no es preciso conocer los valores exactos de c_p y c_c sino únicamente su razón. Así en el estudio de sensibilidad que vamos a realizar, consideraremos el efecto en N^* cuando dicha razón cambia. Dicho estudio se encuentra en la Tabla 5.

Relación entre costes	Nº mantenimientos	Transacciones entre mantenimientos
$c_c = c_p$	3	125000
$c_c \in (\frac{3}{2}c_p, 2c_p)$	4	93750
$c_c \in (2c_p, 3c_p]$	5	75000
$c_c \in (3c_p, \frac{7}{2}c_p)$	6	62500
$c_c \in (\frac{7}{2}c_p, \frac{9}{2}c_p)$	7	53571
$c_c \in (\frac{9}{2}c_p, \frac{11}{2}c_p)$	8	46875

Tabla 5: Mantenimientos óptimos según la relación entre costes

Se advierte que cuánto mayor sea la razón entre el coste del mantenimiento correctivo y el preventivo, el número de mantenimientos programados aumenta. Es decir, al aumentar el coste como consecuencia de un fallo, compensa hacer más mantenimientos preventivos para evitarlo.

7.4. Número esperado de incidencias por tipo de fallos.

Anteriormente se ha determinado que todos los tipos de fallos se ajustan a una distribución Weibull. Para evitar un exceso de mantenimientos y con ello de interrupciones de las máquinas

se ha realizado un estudio que permite hacer agrupamientos de fallos dependiendo del tipo de establecimiento.

Un aspecto de interés es que no ha habido ningún cambio de reciclador en ninguna de las máquinas de prueba. En el resto de componentes, por el contrario sí se han producido sustituciones. Sin embargo no se tiene la información de en qué incidencias se ha subsanado el error o se ha cambiado directamente el componente. El efecto al cambiar elementos que en la fabricación salieron defectuosos es que las tasas de fallo del resto de componentes son decrecientes (parámetro de forma inferior a 1) y, por tanto, no tiene sentido hacer mantenimiento preventivo sobre una componente nueva.

En la Figura 1, relativa a la tasa de fallos en curva de bañera, se refleja esta situación. La tasa, inicialmente muy elevada como consecuencia de defectos de diseño o fabricación, decrece cuando estos son detectados y eliminados.

Nombre del fallo	MTBF	Nº de mantenimientos
Atasco en lectura/monedas excepto gasolineras	20847	18
Atasco en lectura/monedas gasolineras	47641	8
Transporte/stacker/lector excepto gasolineras	73036	5
Transporte/stacker/lector en gasolineras	155316	2

Tabla 6: MTBF y número de mantenimientos durante el periodo de garantía

Para estas componentes mejoradas propondremos un procedimiento alternativo para lo cual se va a proceder a calcular el número medio de fallos durante el periodo de garantía en cada uno de los escenarios, dividiendo el periodo de garantía entre este valor y así poder tener una aproximación al tiempo en el que realizar el mantenimiento preventivo.

Para obtener el número esperado de incidencias en el periodo de garantía, se calcula el tiempo medio entre fallos (mean time between failures (MTBF)), es decir, $E[T]$ cuya expresión está dada en (2). En la Tabla 6 se resume el MTBF en todos los escenarios con la excepción del reciclador, para el cual ya se ha estudiado la función de coste, junto con su correspondiente número de mantenimientos en el periodo de garantía.

8. Conclusiones

En este TFG se ha realizado un estudio de fiabilidad sobre una máquina automática de cobro de efectivo, a partir de los datos proporcionados por el fabricante. Estos habían sido obtenidos en 48 máquinas durante un periodo de prueba de seis meses. El primer objetivo era analizar 6 tipos diferentes de fallo así como la posible influencia del entorno, considerando 8 establecimientos de diferente ámbito en los que la máquina se localiza en la práctica. Tras un proceso de depuración de datos fue preciso eliminar la información de dos de las máquinas por su escasa representatividad.

Un primer análisis no paramétrico puso de manifiesto que un determinado tipo de fallo, el del reciclador, presentaba tiempos más largos de ocurrencia, siendo los atascos en lectura y monedas, los que ocurren más frecuentemente.

El segundo análisis paramétrico reflejó que las diferencias se extendían incluso a la distribución de ajuste a los datos. Los fallos de atasco en monedas, lectura, entrada unitaria, stacker y transporte se describen bien mediante una variable aleatoria Weibull mientras que el reciclador se modela mejor mediante una distribución normal.

Asimismo se ha analizado el posible efecto de los distintos tipos de establecimiento y de las categorías de fallo en la variable número de transacciones hasta el fallo. Este estudio está basado en modelos de regresión en fiabilidad de los que se concluye que las máquinas presentan una duración más elevada en las gasolineras mientras que el resto de establecimientos no muestran diferencias significativas entre sí. El segundo de los modelos de regresión sobre los datos relativos al reciclador revela que su comportamiento no difiere de unos negocios a otros. La falta de datos de fallo en carnicerías/pescadería/frutería y gasolineras en las que hay presencia mayoritaria de datos censuados, cuestiona la validez de esta conclusión en estos dos tipos de negocio. Sería preciso disponer de más datos.

En este TFG hemos hecho un estudio sobre las políticas de mantenimiento preventivo que debe realizar la empresa durante el periodo de garantía y que resulten óptimas para la empresa desde el punto de vista del coste. Para establecer la función coste nos hemos basado en las estimaciones obtenidas en las secciones previas. Las conclusiones indican que la política de mantenimiento tendría que distinguir los dos casos siguientes:

- El mantenimiento del reciclador se puede hacer con la misma frecuencia en bares/caféterías, panaderías y restaurante. Los resultados no son concluyentes en carnicerías/pescaderías/fruterías

y gasolineras por falta de información suficiente.

- Para la prevención de los demás tipos de fallo se puede realizar un mantenimiento menos frecuente en gasolineras que en el resto de negocios.

El plan propuesto en el periodo de garantía de 375000 transacciones se ha realizado asumiendo que el coste del mantenimiento correctivo sea el doble que el del preventivo.

El atasco en lectura y el atasco en monedas es la incidencia más común ya que se encuentran en el primer nivel de la máquina pues es donde se depositan las monedas y los billetes. La mayoría de las veces esta incidencia solo indica suciedad que se solventa con un sencillo bote de aire comprimido y que por tanto puede hacer el propio propietario de la máquina sin necesidad de esperar al técnico. Tras estas observaciones la política que se propone es la siguiente:

- **Máquinas en gasolineras.** La empresa hará 4 mantenimientos preventivos durante el periodo de garantía cada 75000 transacciones. Por su parte el cliente hará un mantenimiento de los atascos en lectura y en monedas usando aire comprimido cada 45000 transacciones.
- **Resto de establecimientos .** La empresa hará 5 mantenimientos preventivos durante el periodo de garantía cada 70000 transacciones. Los mantenimientos de limpieza con aire comprimido los realizará el cliente cada 20000 transacciones.

El análisis realizado me lleva a sugerir algunas mejoras en el proceso de monitorización si se desea llevar a cabo un estudio similar en el futuro:

- Si se desea estudiar algún tipo de negocio en concreto, se deben colocar varias máquinas en establecimientos de la misma clase. El problema detectado de disponer de una muestra de tamaño 1 en heladerías y estancos ha llevado a tener que prescindir de la información proporcionada por las dos máquinas con la consiguiente pérdida de tiempo y dinero para la empresa.
- Sería importante tener una información más precisa durante los periodos de prueba, indicando qué tipo de incidencia atiende el técnico, es decir, si arregla la componente o la sustituye por una nueva. En el trabajo se ha mencionado que unos cuantos componentes han sido cambiados pero la información de cuáles y cuando se desconoce. Al estar mezcladas reparaciones y sustituciones se pierde la idéntica distribución.

9. Bibliografía

- Crowder, MJ., Kimber, A.C., Smith, R.L. and Sweeting, T.J. Statistical Analysis of Reliability Data. Chapman & Hall
- <https://support.minitab.com/es-mx/minitab/19/>
- Laura Martinez Fernandez, Métodos de inferencia para la distribución Weibull: aplicación en fiabilidad industrial. Trabajo final del máster en técnicas estadísticas. Universidad de Vigo. Julio 2011
- Cristina Pruenza García-Hinojosa, Estudio de análisis de supervivencia. Trabajo final del grado en ingeniería informática y matemáticas. Universidad Autónoma de Madrid. Mayo 2014.
- Klein, John P, and Melvin L Moeschberger. 2006. Survival Analysis: Techniques for Censored and Truncated Data. Second Edition. Springer Science & Business Media.
- W. Nelson (1982). Applied Life Data Analysis. John Wiley & Sons, Inc. 133.
- Elisa T. Lee (1992). Statistical Methods for Survival Data Analysis, Second Edition. John Wiley & Sons, Inc. 73-76.