Comparing the Maintenance Quality of Two Groups of Equipment: A Case of Water Chillers

Jeng-Hsiang Lin
Department of Architecture
Hwa Hsia University of Technology
New Taipei City, Taiwan
hsiang@cc.hwh.edu.tw

***Abstract*—In practice, air-conditioner service companies often wish to compare the maintenance quality of two groups of equipment. One of the feasible methods is to compare the survival rates at some specified survival times. However, differences in survival rate between two groups of equipment can change significantly depending on the chosen time. To overcome this problem, this study tried to find out a statistical test method for determining whether there is a statistically significant difference between the survival functions or hazard functions of two groups of water chillers. This study collected two sets of data on the maintenance of water chillers in the stations of the Red Line and the Green Line, Taipei Metro, Taiwan. Then, survival analysis was performed to estimate the survival rates for the two groups of water chillers and identify the probability model of the interarrival time of equipment failure. This study used the Kaplan-Meier method to produce the survival curves for water chillers and thereby comprehend the cycle of maintenance and repairs. Finally, the differences in survival rate between the two groups of water chillers were examined and used to infer the difference in maintenance quality. Research results show that the survival time data of water chillers satisfy the proportional hazard model. The log-rank test can be used to verify the difference in maintenance quality between two groups of water chillers.**

***Index Terms*—Water chillers, maintenance quality, survival rate.**

I. INTRODUCTION

A comfortable temperature and good air quality are important indicators of environmental quality in underground transportation systems, and the air-conditioning system is the most significant factor influencing these elements [1-3]. Water chillers are key components in air-conditioning systems, and consume the most power of any equipment in Mass Rapid Transit (MRT) stations. For these reasons, rapid transit corporations and air-conditioner service companies are keenly interested in the prediction of downtime and maintenance costs for water chillers.

The design and maintenance quality of air-conditioning systems has an enormous influence on the environmental quality of MRT stations [4,5]. The main function of water chillers in air-conditioning systems is to produce chilled water, which is provided to downstream components to facilitate the heat exchange and cooling cycle. When water chillers suffer sudden failure or damage, air-conditioning systems in MRT stations shut down or malfunction, causing significant operational loss and inconvenience to passengers. Therefore, prediction of the probability model for the interarrival time of equipment failure in water chillers is an important issue in the management of air-conditioning systems.

In practice, we often wish to compare the maintenance quality of two groups of equipment. One of the feasible methods is to compare the survival rates at some specified survival times. However, differences in survival rate between two groups of equipment can change significantly depending on the chosen time. This problem can be overcome through the use of statistical test methods (e.g. log-rank test) to determine whether there is a statistically significant difference between the survival functions or hazard functions of two equipment groups. In this study, survival analysis [6,7] was performed to identify the probability model for the interarrival time of equipment failure in water chillers in MRT air-conditioning systems. Statistical test were employed to evaluate and compare the maintenance quality of two groups of water chillers.

II. PROBABILITY MODEL AND SURVIVAL FUNCTION

This study used the Kaplan-Meier method [8] to produce the survival curves for water chillers and thereby comprehend the cycle of maintenance and repairs. Statistical inference was used to estimate differences in the survival rate and maintenance quality of water chillers.

A. Data Collection And Processing

The objective of survival analysis is to analyze the random distribution of event time [6]. Apart from being related to the definition of events, the observations of event time are also associated with the starting point of observation, the time of occurrence, and the duration of data collection (the length of the observation period) for each water chiller. In this study, the event was defined as equipment failure in water chillers in MRT air-conditioning systems. Empirical data of event time was collected from the repair logs of water chillers in the stations of the Red Line and the Green Line, Taipei Metro, Taiwan, and used to estimate the parametric values of the probability model. Equipment failures in water chillers were categorized according to the cause of failure, including: control panel anomalies, high voltage tripping, insufficient cooling water, power source tripping, and insufficient chilled water.

A longer period of data collection and a greater quantity of data produce analysis results of greater reliability. Due to the

IJARE

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

limitations of research deadlines, we were unable to obtain observed values for longer event period; as a result, the information collected by this study may include incomplete data. In the domain of statistics, incomplete data is known as right-censored data. This means that the starting point of observation is known, but the defined event (equipment failure) has not occurred by the end of the observation period. It is therefore not possible to confirm the interarrival time of the event. Another cause of right-censored data is when units of observation are withdrawn from the experiment prior to the research deadline (for example, a change of maintenance providers or the replacement of equipment), making it impossible to determine the interarrival time of the event for a particular unit. The interarrival time between two consecutive failures observed in a single water chiller was termed survival time. For right-censored data, the equivalent term is right-censored time, indicating the interarrival time between the last occurrence of the event and the end of the observation period. Figure 1 shows the definitions of survival time and right-censored time.

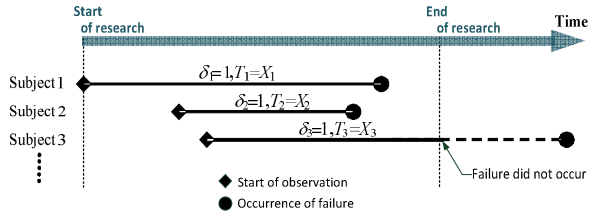


Fig 1. Method of Defining Observational Data.

To enhance the comprehensiveness of data analysis, this study included censored time data in thereby survival analysis, thus considering all observations. In other words, when estimating the likelihood function of the probability model, this study considered the likelihood of both survival time data and right-censored time data.

B. Likelihood Function Of Survival Data

The random variable X represents the interarrival time between two consecutive failures; the random variable C represents censored time. Both variables (X and C) are non-negative, mutually independent continuous random variables. In survival analysis, X is the survival time variable, defined as $[0,\infty)$. We established the probability density function $f(x)$ and cumulative distribution function $F(x)$ of X, and the probability density function g(x) and cumulative distribution function $G(x)$ of C.

Consider n sets of random sample data (X_1, C_1), (X_2, C_2),.., (X_n, C_n). Under random censoring, n sets of observations can be expressed as (T_i,δ_i), i=1,2,…,n, where

$$T_i = \min.(X_i, C_i) = \begin{cases} X_i, & if \quad X_i \le C_i \\ C_i, & if \quad X_i > C_i \end{cases} \tag{1}$$

$$\delta_i = I[X_i \le C_i] = \begin{cases} 1, & if \quad X_i \le C_i \\ 0, & if \quad X_i > C_i \end{cases} \tag{2}$$

min.(.) represents the minimum value; I[.] is the indicator function, and δ_i is used to indicate the state of data T_i. When δ_i =1, this indicates that T_i is a survival time; when δ_i =0, T_i is a censored time. Note that (T_i,δ_i), i=1,2,…,n, were n sets of independently and identically distributed (iid) observations; the likelihood function, L, of the n sets of observation data can be expressed as [9]

$$\begin{aligned} L &= \prod_{i=1}^{n} [f(T_i)\bar{G}(T_i)]^{\delta_i} [g(T_i)\bar{F}(T_i)]^{1-\delta_i} \\ &= \prod_{i=1}^{n} [f(T_i)^{\delta_i} \bar{F}(T_i)^{1-\delta_i}][\bar{G}(T_i)^{\delta_i} g(T_i)^{1-\delta_i}] \end{aligned} \tag{3}$$

where

$$\bar{F}(T_i) = 1 - F(T_i) \tag{4}$$

$$\bar{G}(T_i) = 1 - G(T_i) \tag{5}$$

C. Hazard Function And Survival Function

As defined by survival analysis [6], the probability that survival time variable X will exceed a specific value x is termed the survival rate at time x. The survival function S is defined as follows:

$$S(x) = \Pr(X > x) = \int_x^{\infty} f(t)dt \tag{6}$$

in which f is the probability density function of survival time X. The survival function S is a function that monotonically decreases with time to zero; Thus, $S(0)$=1 and $S(\infty)$=0. The relationships among the survival function, distribution function F and probability density function f are described below:

$$F(x) = \Pr(X \le x) = 1 - S(x) \tag{7}$$

$$f(x) = \frac{dF(x)}{dx} = -\frac{dS(x)}{dx} \tag{8}$$

In survival analysis, the hazard function h represents the instantaneous failure rate at time x. According to stochastic process theory, the hazard function is an intensity function indicating the occurrence frequency of equipment failure within a unit of time. $h(x)\ge 0$. The hazard function is defined as follows:

$$h(x) = \lim_{\Delta x \to 0} \frac{\Pr(x \le X < x + \Delta x \,|\, X \ge x)}{\Delta x} \tag{9}$$

The relationship among the hazard function, probability density function, and survival function is expressed below:

$$h(x) = \frac{f(x)}{S(x)} = -\frac{d}{dx}\log[S(x)] \tag{10}$$

Another important function is the cumulative hazard function, H, defined as

$$H(x) = \int_0^x h(t)dt = -\log[S(x)] \tag{11}$$

IJARE

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

The survival function can be expressed as

$$S(x) = \exp[-H(x)] = \exp[-\int_0^x h(t)dt] \quad (12)$$

In this study, the hazard function and cumulative hazard function include important information related to the occurrence of equipment failure in water chillers. These functions are particularly useful when applying survival analysis to identify the probability model for the interarrival time of equipment failure in water chillers. Commonly used types of hazard function in survival analysis include constant function, incremental function, decreasing function, bathtub-shaped function, and hump function.

D. Probability model of the interarrival time of equipment failure

This study hypothesized that the probability distribution of the interarrival time of two consecutive failures in a water chiller was a single-parameter exponential distribution. The Kaplan-Meier method was used to construct the survival function for water chillers, which was then used to analyze and compare the maintenance quality of different water chillers. The probability density function of exponential distribution is as follows:

$$f(x) = \lambda \exp(-\lambda x) \quad (13)$$

where λ is the parameter of exponential probability distribution. The corresponding hazard function, cumulative hazard function, and survival function are respectively shown below:

$$h(x) = \lambda \quad (14)$$

$$H(x) = \int_0^x h(t)dt = \lambda x \quad (15)$$

$$S(x) = \exp(-\lambda x) \quad (16)$$

By using the maximum likelihood estimation (MLE) method to estimate, we can obtain the maximum likelihood estimator of λ. The cumulative hazard function of the exponentially distributed survival time variable has a linear relationship with survival time x, indicating that the occurrence rate of equipment failure does not change over time. When models are being tested, this property can be used to validate the suitability of the hypothesized exponential distribution.

E. Maximum likelihood estimation of model parameters

We hypothesized that the joint probability density function of T and δ included p parameters (θ_1, θ_2,…, θ_p), and that the parameter vector was $\Theta=(\theta_1, \theta_2,\ldots, \theta_p)$. n groups of random samples were considered ((T_1,δ_1), (T_2,δ_2),…, (T_n,δ_n)), and the likelihood function L can be expressed as follows:

$$L(\Theta) = \prod_{i=1}^{n} f(\Theta; T_i, \delta_i) \quad (17)$$

The maximum likelihood estimator of the parameter can be obtained through the maximum likelihood function [10]. In other words, the maximum likelihood estimator of parameter Θ satisfied the partial differential Eq. 18, and Hessian was determined to be a negative-definite parameter vector.

$$\frac{\partial}{\partial \theta_j} L(\Theta) = 0 \qquad j = 1, 2, \ldots, p \quad (18)$$

A closed-form solution can be obtained for the maximum likelihood estimators of the parameters (θ_1, θ_2,…, θ_p) described above. The maximum likelihood estimator is the function of random sample groups (T_1,δ_1), (T_2,δ_2),…, (T_n,δ_n).

If we assume that the probability distribution of survival time variable X and censored time variable C is exponential, then

$$f(x;\lambda) = \lambda e^{-\lambda x} \quad (19)$$

$$f(c;\mu) = \mu e^{-c\mu} \quad (20)$$

where x>0 and c>0.

The variables X and C are mutually independent, as are the variables T and δ. Variable conversion was used to obtain the joint probability density function of T and δ:

$$f(T,\delta) = \lambda^{\delta} \mu^{1-\delta} e^{-(\lambda+\mu)T} \quad (21)$$

where T>0, δ=0 or 1, μ>0, and λ>0. Equation 21 and n groups of random sample data ((T_1,δ_1), (T_2,δ_2),..., (T_n,δ_n)) were integrated into Eq. 17; the likelihood function can be expressed as

$$L(\lambda,\mu) = (\lambda)^{\sum_{i=1}^{n}\delta_i} (\mu)^{n-\sum_{i=1}^{n}\delta_i} \exp\left\{-(\lambda+\mu)\sum_{i=1}^{n} T_i\right\} \quad (22)$$

Substituting Eq. 22 into Eq. 18, the maximum likelihood estimators for λ and μ can be respectively expressed below:

$$\hat{\lambda}_{MLE} = \frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} T_i} \quad (23)$$

$$\hat{\mu}_{MLE} = \frac{n-\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} T_i} \quad (24)$$

III. EVALUATION OF MAINTENANCE QUALITY

In research on facilities management, the influence of maintenance methods or environmental conditions on the quality of maintenance is a topic of considerable interest. Regardless of the degree of similarity in the procedures used, maintenance quality often differs according to environmental conditions. For property management companies, higher

IJARE

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

maintenance quality can reduce costs and enhance market competitiveness.

This study used the log-rank method [11-13] to test for a significant difference in the survival curves of two groups of water chillers, and thereby infer the difference in maintenance quality. This widely applied nonparametric test method was proposed by Nathan Mantel [11] and termed the log-rank test by Richard and Julian Peto [12]. The log-rank test is used to test the null hypothesis that there is no difference between the populations in the probability of an event at any time point. The test is most likely to detect a difference between groups when the risk of an event is consistently greater for one group than another. However, the log-rank test is suitable when a proportional hazard rate exists between two groups of survival time data.

Consider two groups of observational data including observations from water chillers of Group 1: (T_{1i},δ_{1i}), i=1, 2,…,n_1 and Group 2: (T_{2i},δ_{2i}), i=1,2,…,n_2. d was set as the sum of the survival time data of these two groups (i.e. sum of the observations of equipment failure); n_1+n_2-d indicates the sum of censored time data. In d number of observations, there are m disparate survival times. These m dissimilar survival times were arranged in ascending order of value and can be expressed as $t_1<t_2<\ldots<t_m$; thus, $m\leq d\leq n_1+n_2$.

In statistical hypothesis testing, the null hypothesis H_0 and the alternative hypothesis H_1 were as follows:

H_0:$h_1(t_i)= h_2(t_i)$

H_1:$h_1(t_i)< h_2(t_i)$ or H_1:$h_1(t_i)>h_2(t_i)$

The null hypothesis established at each dissimilar time point t_i (i=1,2,…,m) and the alternative hypothesis established at least one time point t_i (i=1,2,…,m). We compiled a 2x2 contingency table at each dissimilar time point t_i (i=1,2,…,m), as shown in Table 1. In the table, n_{ji} (j=1,2) represents the observed number of water chillers in Group j at the moment prior to survival time t_i. d_{ji} (j=1,2) indicates the observed number of malfunctioning water chillers in Group j at survival time t_i. The random variable D_{1i} represents the number of malfunctioning water chillers in Group 1 at survival time t_i. The variables D_{11},D_{12},D_{13},...,D_{1m} are mutually statistically independent.

Under the conditions of the null hypothesis H_0:$h_1(t_i)= h_2(t_i)$, the expected value $E(D_{1i})$ and variance $Var(D_{1i})$ of D_{1i} are as follows:

$$E(D_{1i})=\frac{n_{1i}d_{.i}}{n_{.i}} \tag{25}$$

$$Var(D_{1i})=\frac{n_{1i}n_{2i}(n_{.i}-d_{.i})d_{.i}}{n_{.i}^2(n_{.i}-1)} \tag{26}$$

Table I. 2x2 contingency table at event time t_i.

Groups	Status		Sum total
	Number of malfunctioning water chillers	Number of properly functioning water chillers (survival number)	
Group 1	d_{1i}	$n_{1i}-d_{1i}$	n_{1i}
Group 2	d_{2i}	$n_{2i}-d_{2i}$	n_{2i}
Sum total	$d_{.i}$ **	$n_{.i}-d_{.i}$	$n_{.i}$ *

* $n_{.i}=n_{1i}+n_{2i}$; ** $d_{.i}=d_{1i}+d_{2i}$

Then the statistic of the log-rank test can be given as

$$\Lambda=\sum_{i=1}^{m}[d_{1i}-E(D_{1i})] \tag{27}$$

Under the conditions of the null hypothesis H_0:$h_1(t_i)= h_2(t_i)$, the test statistic Λ complies with a normal distribution. The expected value $E(\Lambda)$ and variance $Var(\Lambda)$ of Λ are as follows:

$$E(\Lambda)=0 \tag{28}$$

$$Var(\Lambda)=\sum_{i=1}^{m}Var[D_{1i}] \tag{29}$$

Let $Z=\Lambda/\sqrt{Var(\Lambda)}$, then the statistic Z complies with the standard normal distribution.

IV. EMPIRICAL ANALYSIS

This study used the water chillers in the stations of the Red Line and the Green Line, Taipei Metro, Taiwan as case studies, and assessed the quality of maintenance. The water chillers in the stations of the Red Line, Taipei Metro, Taiwan are designated as the water chillers of Group 1 and those in the Green Line are designated as the water chillers of Group 2. Two sets of data on the maintenance of water chillers in the Red Line and the Green Line were respectively collected. When conducting survival analysis, sample observations were obtained from the repair and maintenance logs of two groups of water chillers.

Figure 2 illustrates the survival curves (estimated using the Kaplan-Meier method) for water chillers of Groups 1 and 2. As shown in Fig. 2, compared to Group 1, water chillers of Group 2 have a higher survival rate. This indicates that within a specific period of time, water chillers of Group 1 have a higher probability of equipment failure. As the figure shows, the survival curves differ, but is this sufficient to conclude that in the population water chillers of Group 1 have worse survival than those of Group 2? This study used the log-rank method to test the statistical difference in survival rates between two groups of water chillers. The designated level of significance for the test was 0.05. Test results showed that p-value is equal to 0.03, which was below the 0.05 level of significance. The conclusion of the test was that at a 0.05 level of significance, the maintenance quality of water chillers of Group 2 was significantly higher than that of water chillers of Group 1.

IJARE

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

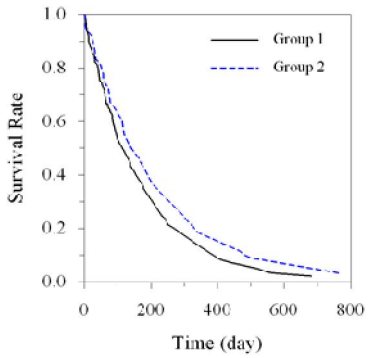


Fig 2. Survival curves for water chillers for Groups 1 and 2.

The solid lines and dotted lines in Fig. 3 represent the estimated values and theoretical values of the cumulative hazard functions of water chillers of Groups 1 and 2, respectively. The figure shows that the estimated cumulative hazard functions of both water chillers have a near linear relationship with the interarrival time of equipment failure. This indicates that the occurrence rate (i.e. hazard rate) of equipment failure does not change with time, and that the estimated values and theoretical values are relatively close. This also demonstrates that the probability distribution of the interarrival time of equipment failure in water chillers is approximately the exponential distribution. The theoretical values of cumulative hazard function in Fig. 3 were obtained from Eq. 15, and the maximum likelihood estimator of parameter was estimated using Eq. 23. By incorporating the parameter estimators of water chillers of Groups 1 and 2 into Eq. 16, we obtain the survival function equation for both water chillers.

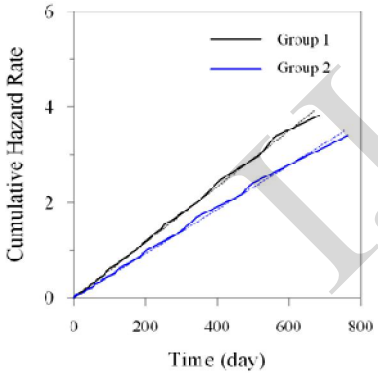


Fig 3. Cumulative hazard functions of water chillers for Groups 1 and 2.

V. CONCLUSIONS

This study conducted survival analysis of empirical data to estimate the survival curves of water chillers and provide a probability model of the interarrival time of equipment failure. Maximum likelihood estimates were used to estimate the parametric values of the probability model. We also applied the log-rank test to infer the differences in maintenance quality between two groups of water chillers. Research results show that the probability model of the interarrival time of equipment failure is exponential distribution. The survival time data of water chillers satisfy the proportional hazard model. The log-rank test can be used to verify the difference in maintenance quality between two groups of water chillers.

REFERENCES

[1] P. O. Fanger, Thermal Comfort: Analysis and Application in Environmental Engineering, McGraw-Hill, New York, 1970.
[2] J. C. Parker, "Air-conditioning related to passenger environment," Proceedings of the Railway Division Conference on Passenger Environment, IMechE C32/72, 1972, pp.58–63.
[3] S. C. Hu and J. H. Lee, "Influence of platform screen doors on energy consumption of the environment control system of a mass rapid transit system: case study of the Taipei MRT system," Energy Conversion and Management, vol.45, 2004, pp.639–650.
[4] F. Ampofo, G. Maidment, and J. Missenden, "Underground railway environment in the UK part 1: review of thermal comfort," Applied Thermal Engineering, vol.24, 2004, pp.611-631.
[5] M. Abbaspour, M. J. Jafari, N. Mansouri, F. Moattar, N. Nouri, and M. Allahyari, "Thermal comfort evaluation in Tehran metro using relative warmth index," Int. J. Environmental Science Technology, vol.5, 2008, pp.297-304.
[6] D. W. Hosmer, Applied Survival Analysis, Wiley Series in Probability and Statistics, New York, 1999.
[7] J. F. Lawless, Statistical Models and Methods for Lifetime Data, Wiley Series in Probability and Statistics, New York, 1981.
[8] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," Journal of the American Statistical Association, vol.53, 1958, pp.457–481.
[9] D. M. Dabrowska and K. A. Doksum, "Partial likelihood in transformation models with censored data," Scand. J. Statist., vol.15, 1988, pp.1-24.
[10] O. Borgan, "Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data," Scand. J. Statist., vol.11, 1984, pp.1-16.
[11] N. Mantel, "Evaluation of survival data and two new rank order statistics arising in its consideration," Cancer Chemotherapy Reports, vol.50, 1966, pp.163–70.
[12] R. Peto and J. Peto, "Asymptotically efficient rank invariant test procedures," Journal of the Royal Statistical Society, Series A (General), vol.135, 1972, pp.185–207.
[13] J. M. Bland and D. G. Altman, "The logrank test," British Medical Journal, vol.328, 2004, pp.1073.

IJARE

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)