Journal of Quantitative Economics Studies | Volume07, Number01 (2021), PP.427-435 | EISSN: 2602-5183 * PISSN: 1033-2437

A New Method for the Comparison of Survival functions using the Ranking of Median

HAMIMES Ahmed[*,1], BENAMIROUCHE Rachid[2], CHELLAI Fatih[3]

[1]École nationale de la statistique et d'économie appliquée, université de Constantine 3, Algérie,(ahmedhamimes@yahoo.com).

[2] École nationale de la statistique et d'économie appliquée, Algérie, (rbena2002@hotmail.com)

[3]Faculté d'économie, de commerce et de gestion, Université de Ferhat Abbas, Algérie, (fatih.chellai@univ-setif.dz)

Received: 24/08/2020; **Revised:** 17/11/2020; **Accepted:** 10/07/2021

Summary

The value of the median in the duration models is very important either in the interpretation of the results obtained or in the comparison between survival functions in the situation where the risk functions $h_1(t)$ and $h_2(t)$converge in the time t. These test methods are considered more effective than log-rank and Wilcoxon tests when the survival curves are not proportional. In this article we use a classification function in the objective is to probability the median parameter, this method makes it possible to determine an interval with different probabilistic weights and simplifies the comparison between the survival functions.

Keywords: median, survival functions, classification function.

Jel Classification Codes:C41,C12, C25.

Résumé

La valeur de la médiane dans les modèles de durées est très importante soit dans l'interprétation des résultats obtenus ou dans les comparaisons entres des fonctions de survies dans la situation où les fonctions de risque $h_1(t)$et $h_2(t)$ convergent dans le temps$t$. Ces méthodes de tests sont considérées comme plus efficaces que les tests log-rank et Wilcoxon lorsque les courbes de survie ne sont pas proportionnelles. Dans cette article on utilise une fonction de classification dans l'objectif est de probabilisé le paramètre de la médiane, cette méthode permet de déterminer un intervalle avec des poids probabiliste différents et de simplifie la comparaison entre les fonctions de survies.

Mots-clés : médiane, fonctions de survies, une fonction de classification.

Codes de classification Jel: C41, C12, C25.

* Corresponding author.

I. Introduction

Survival analysis techniques are generally used in the fields of medicine, biology, public health, clinical trials, and epidemiology. The unique feature of survival data is that there is censorship. In other words, the case might be of interest to some topics, others not. Lack of experience may result from the end of the study period, loss of follow-up or withdrawal from the study by the subject. This is also called right-censored information. An event of interest can be described as death, onset of illness, recurrence of illness, or recovery, one of which is called "failure". Traditional statistical models, such as multiple linear regression, cannot handle direct censorship, which is why survival models are designed specifically for managing data using censored observations. If censorship is ignored in a study. the results seem to underestimate the probability of survival. The basic goals of survival analysis include estimating and interpreting survival probabilities, comparing survival probabilities between classes, and assessing possible risk factors related to the probability of survival.

In this article, we focus on methods of comparing survival curves between two different groups. We use the median which represents an alternative to the classical method of comparison because of their specific interpretation and the ease of calculation.The median survival time is thought to be a significant measure of the survival distribution due to the skewness of survival results. If two crossing survival curves (Chen and al, 2015) are different but their median survival times are identical, then it is more fitting to address some research questions by comparing the survival media or quantiles rather than the curves. comparing median survival times(Chen(2014), Brookmeyer R, Crowley (1982)). Brookmeyer and Crowley proposed a test (hereafter called the BC test) based on the weighted survival curve and the weighted median(Brookmeyer R, Crowley,1982). In their study, it is appropriate to measure
the individual survival curve for each group and their weighted median survival time using the Kaplan-Meier estimator; then the probability of survival of all groups are compared at the time point of the calculated weighted survival median The authors have shown that the test statistics have an asymptotic chi-square distribution with degrees of freedom equal to the number of groups minus one, assuming that the groups have the same distribution of survival. To obtain this test statistics, we need to estimate a matrix of covariances and determine its generalized inverse, which may be a tedious task(Tang and Jeong (2012)). Another drawback of the BC test is its exaggerated error rate of type I, which can cause serious problems if the sample size is small (Rahbar and al (2012)).
In our article to simplify the test we use the Bayesian approach. Thus it is possible to calculate the Bayesian p-value or place the frequentist P-value and introduce a priori information if it exists. This method plays a dual role; it allows the comparison of survival functions, also the determination of the median interval.

II. Kaplan-Meier estimator and the Bayesian approach

The Kaplan-Meier (KM) estimation method is also called "Product Limit Estimations (PLE)" by Anglo-Saxon statisticians. This estimator, which generalizes the notion of empirical distribution function, is based on the following idea: living after a while means being alive just before and not dying over time, that is,

$$P(X > t) = P(X > t_{i-1}, X \geq t_i)$$
$$= P(X > t_i / X \geq t_{i-1}) P(X > t_{i-1})$$
$$= P(X > t_i / X \geq t_{i-1}) P(X > t_{i-1} / X \geq t_{i-2}) P(X > t_{i-2})$$

Wecanwrite:

$$S(t_i) = P(X > t_i / X \geq t_i) * S(t_{i-1})$$

the proportion q_i of individuals who experienced the event at time t_i corresponds to

$$q_i = \frac{d_i}{n_i}$$

this quantity estimates the value of the risk function $\lambda(t)$ for $t = t_i$, where : t_i represents the follow-up time since inclusion in the study for each patient i;
$d_i$est le nombre de décès au temps t_i;
$n_i$is the number of subjects at risk of presenting the event studied at the instant t_i , i.e. the number of patients who have not yet undergone the event nor the censorship just before t_i.

$(1-q_i)$represents the proportion of people who did not experience the event.
The probability of survival in t_ide then becomes:

$$S(t_i) = S(t_{i-1})(1-h_i)$$
$$= S(t_{i-1})\frac{n_i - d_i}{n_i}$$

Selon cette équation, la probabilité de survie t_i sachant qu'on était en vie en t_{i-1} est estimée de la manière suivante :

$$\hat{S}(t_i/t_{i-1}) = \frac{n_i - d_i}{n_i}$$

By extension, if we consider $t_1 < t_2 < \cdots < t_n$ the distinct survival times of n individuals, $\hat{S}(t)$corresponds to the product of all the probabilities of not having known the event since the start of the observation, but this estimator can be written in two different ways depending on whether there is no joint or the presence of a joint as follows:

- Case of absence of a joint (ex aequo):

L'estimateur de Kaplan-Meier est donné par

$$\hat{S}(t) = \begin{cases} \prod_{t_i \le t}\left(1 - \frac{1}{n-i-1}\right)^{\delta_i} & \text{if } t \ge t_1 \\ 1 & \text{if } t < t_1 \end{cases}$$

avec

$$\delta_i = \begin{cases} 1 & \text{realized event,} \\ 0 & \text{censored subject.} \end{cases}$$

- Ex aequo presence:

In the case of application, we are confronted with the presence of events of different natures, we consider that the uncensored observations take place before the censored ones, we have

$$\hat{S}(t) = \begin{cases} \prod_{t_i \le t}\left(1 - \frac{d_i}{n_i}\right) = \prod_{t_i \le t}(1-h_i) & \text{if } t \ge t_1 \\ 1 & \text{if } t < t_1 \end{cases}$$

In the frequentist approach the number of deaths in the interval of time is an realization of a Binomial law written by:

$d_i \sim \beta in(n_i, q_i)$

where

$$q_i = 1 - \frac{d_i}{n_i} \qquad (1)$$

From a Bayesian perspective we assume an a priori for$q_{i,}$, and when the distribution used in the case of proportions is that of Beta, we set:

$$q_{i\sim} beta(\alpha, \beta)$$

For the hyper parameters (α, β), we find several propositions, we use a vague prior law, it is a proper law with a very large variance, according to this distribution, the prior law is considered to be weak informative, and this law is used for regularization and stabilization, it provides solutions in the use of algorithms. We ask:

$q_i \sim \beta(0{,}01, 0{,}01)$

1. Calculate the interval of the median

To determine the interval of the median we follow three steps as follows:

1ère step

In this step we find the value for the classification as follows:

$$SR_i = |S(t_i) - 0.5|$$

2ème step

In this step we find the classification value as follows:

$$rank = \sum_j I(SR_j \le SR_i)$$

such as

$rank \equiv rank_i$

3ème step

We seek in this step to determine the minimum order probability of the function in OpenBUGS equals (rank (sr [], i), 1) = 1 if the ith element of SR_i has the lowest value, and 0 if not; the mean is the probability that the ith element has the lowest value.

So the code that contains the three steps is as follows:

```
for (i in 1:m) {
sr[i]<- abs(s[i]-0.5)
numbers1toN1[i] <- i
rank[i] <- rank(sr[], i) # rank of hospital i
}
for (i in 1:m1) {
prob.lowest1 [i] <- equals (rank[i], 1)
}
```

2. La comparaison entres les fonctions de survies

For the comparison between the survival functions, we compare the classification probabilities of each group, then:

$$\alpha_i = equals(rank(sr2\ [],i),1) - equals\ (rank\ (sr1\ [],i)\ ,1)$$

and the test statistic is the sum of the deviations, so:

$$\bar{\alpha} = \sum_{i=1}^{min\ (m_1,m_2)} \alpha_i = \sum_{i=1}^{min\ (m_1,m_2)} equals(rank(sr2\ \ [],i),1) - equals\ (rank\ (sr1\ [],i)\ ,1)$$

So the code that contains this step in OpenBUGS is as follows:

```
for (i in 1:m1) {
sr1[i]<- abs(s1[i]-0.5)
numbers1toN1[i] <- i
rank1[i] <- rank(sr1[], i) # rank of hospital i
}
for (i in 1:m1) {
prob.lowest1[i] <- equals(rank1[i] ,1 )
}
for (j in 1:m2) {
sr2[j]<-abs(s2[j]-0.5)
numbers1toN2[j] <- j
rank2[j] <- rank(sr2[], j) # rank of hospital i
}
for (i in 1:m2) {
prob.lowest2[i] <- equals(rank2[i] ,1 )
}
for (i in 1:m1) {
alpha[i]<-abs(prob.lowest2[i]-prob.lowest1[i])
}
alphap<- sum(alpha[])
p.value<- step(alphap - 0)
}
```

As a decision scale we use the following table:

Tab1:the decision scale table.

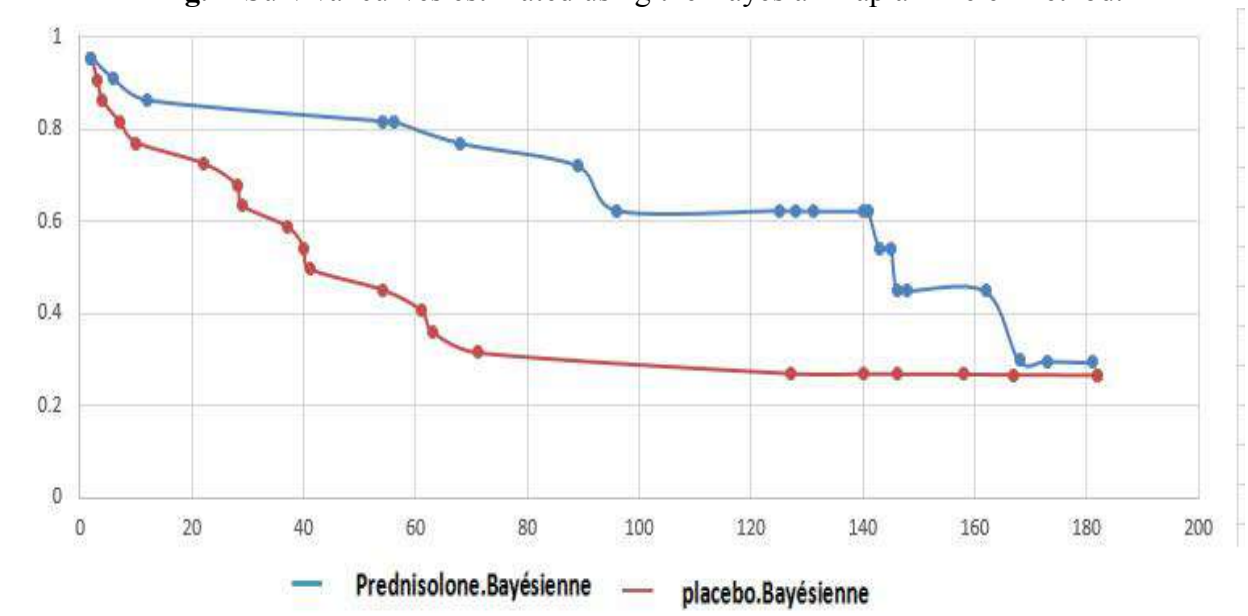
$p-valeur$	*Interpretastion*
0	Equality of survivalfunctions
$0<p-valeur \leq 0.5$	Lowsimilarity
$0.5<p-valeur \leq 1$	A strongsimilarity
1	Deferenceisdecisive

Source: Developed by us.

III. Application

In this section, survival function is estimated in a clinical study for two pharmaceuticals (placebo and prednisolone), this example uses survival times for 42 patients with chronic active hepatitis. These patients were randomized into two equal groups, one was treated with prednisolone, the other received a placebo (see Held, 2010).

Fig.1- Survival curves estimated using the Bayesian Kaplan-Meier method.



Source: Developed by us.

The survival function of the two pharmaceutical products demonstrates that patients in the placebo group have a lower probability of survival than those in the prednisolone group, and according to this result prednisolone has an efficacy compared to the comparator (placebo).

We now want to go beyond the graphical comparison of the two treatment groups to perform an appropriate statistical test. The Student t test does not lend itself to this, because it asks to ignore that in a part of the patients, the event has not yet taken place. For this, the tests in Table 1 are used.

Tab.2- Tests for the equality of survival functions.

Statistic	Observed value	Critical value	p-value	Alpha
Log-Rank	4,028	3,841	0,045	0,05
Wilcoxon	5,686	3,841	0,017	0,05
Tarone -Ware	5,255	3,841	0,022	0,05

Source: Developed by us.

From the results summarized in Tab (2), the p-value of the log-rank, Wilcoxon and Taron-Ware tests (see Table (2)) is below the threshold (5%), which means that we proved that there is a statistically significant difference between the survival probabilities of the two groups.

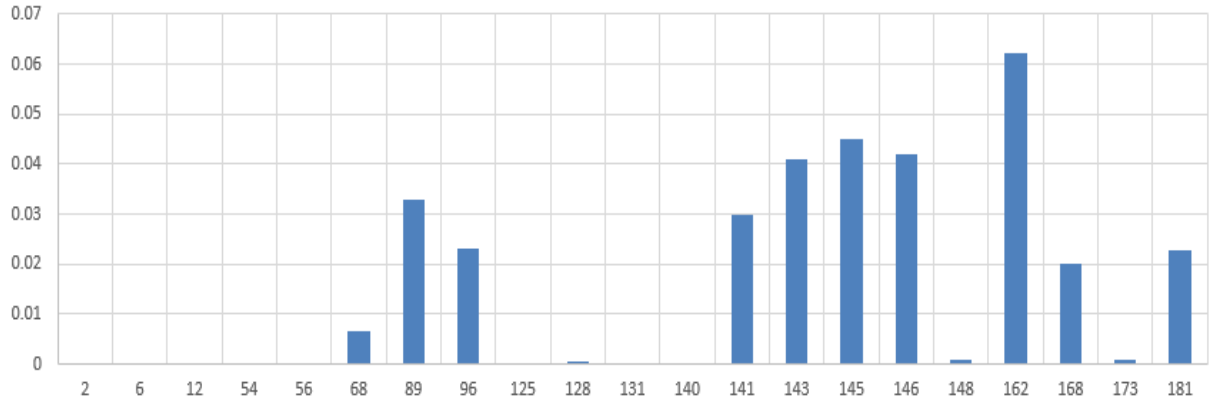
From Table (3) the hypothesis of the difference between the survival probabilities of the two groups is statistically significant when the value of P-value is equal to 1.

Tab.3- The Bayesian test for equality of survival functions.

	mean	sd	MC_error	val2.5pc	median	val97.5pc
p. Value	1.0	0.0	7.071E-13	1.0	1.0	1.0

To confirm the results, a graphical comparison between the medians of the two functions is used.

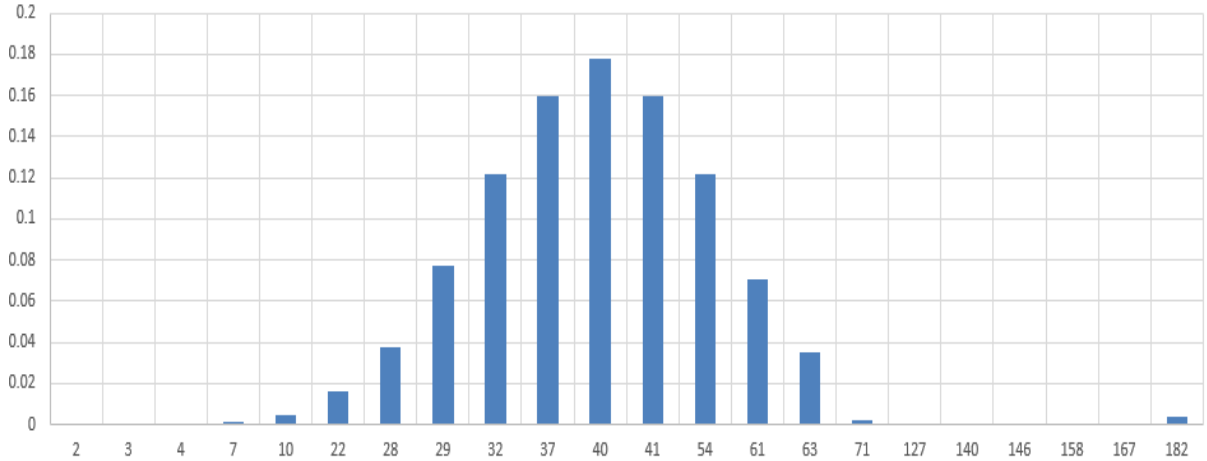
Fig 2 : the probable times and the median value in the prednisolone group.



Source: Developed by us.

In figure (2), we can see graphically that the majority of the mass is located after the diurea 141 months. This indicates that the median survival is between 141 months and 146 months.

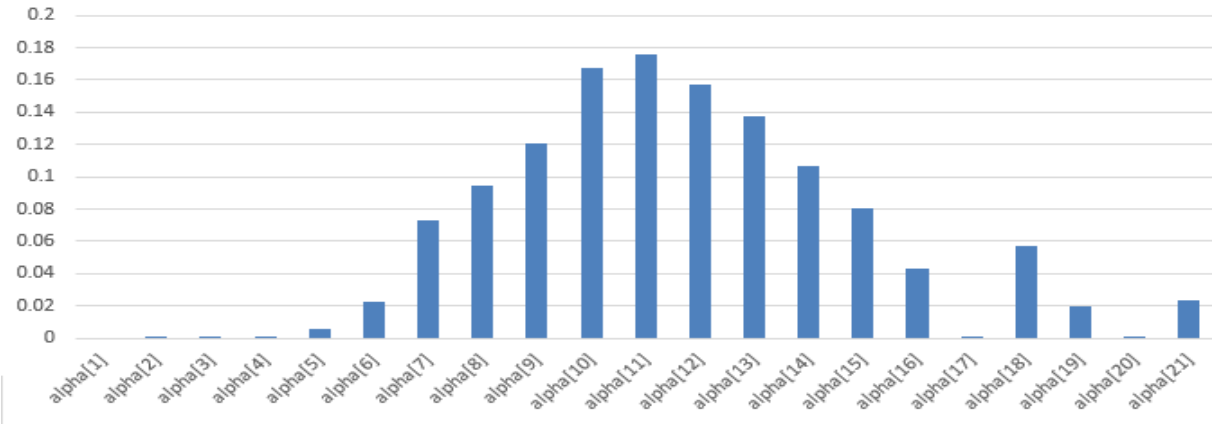
Fig 3 :the probable times and the median value in the placebo group.



Source: Developed by us.

In Figure (3), we find a distribution with a single mass so the median interval is easy to read and is between 22 months and 63 months for the placebo group.

Fig 4 : the difference between the probable times and the value of the median in the groups (Prednisolone VS Placébo).



Source: Developed by us.

The difference between the probable times and the median value in the placebo and prednisolone groups helps to understand the magnitude of the difference in this study. in figure (4) we notice that this difference is large in we accept the hypothesis of difference between the survival functions.

IV. Conclusion

In this article the Bayesian approach and the MCMC simulation method is used to calculate the P-value which allow a more in-depth comparative analysis in the comparison of survival functions, the result found through the example shows a simple, straightforward method and efficient.

References

- Brookmeyer R, Crowley J. A (1989),**k-sample median test for censored data**. J Am Stat Assoc,77(378):433–40.
- Chen Z, Huang H, QiuP(2015),**Comparison of multiple hazard rate functions**. Biometrics, (in press), doi:10.1111/biom.12412.
- Chen Z (2014), **A nonparametric approach to detecting the difference of survival medians**. Commun Stat Simulat.(in press), doi:10.1080/03610918.2014. 964804.
- Held, U (2010),**Représentation graphique et comparaison de courbes de survie**. Forum Med Suisse, 10(33), 548-550.
- Rahbar MH, Chen Z, Jeon S, Gardiner JC, Ning J(2012),**A nonparametric test for equality of survival medians**. Stat Med.31(9):844–54
- Tang S, Jeong JH (2012),**Median tests for censored survival data; a contingency table approach**. Biometrics.68:983–9.

Appendices

```
model
{
for (i in 1:m1) {
d1[i]~dbin(q1[i],n1[i])
q1[i]~dbeta(0.01,0.01)
}
for (i in 1:m2) {
d2[i]~dbin(q2[i],n2[i])
```

```
q2[i]~dbeta(0.01,0.01)
}
for (i in 1:m1){
ce1[i]~dbin(0.01,0.01)
}
for (i in 1:m2){
ce2[i]~dbin(0.01,0.01)
}
for (i in 1:m1){
p1[i]<-1-q1[i]
}
for (i in 1:m2){
p2[i]<-1-q2[i]
}
n1[1]<- 22
n2[1]<- 22
for(i in 2:m1){
n1[i]<-n1[i-1]-d1[i-1]-ce1[i-1]
}
for(i in 2:m2){
n2[i]<-n2[i-1]-d2[i-1]-ce2[i-1]
}
for (i in 2:m1){
s1[i]<-s1[i-1]*p1[i]
}
s1[1]<-p1[1]
for (i in 2:m2){
s2[i]<-s2[i-1]*p2[i]
}
s2[1]<-p2[1]
for (i in 1:m1) {
sr1[i]<- abs(s1[i]-0.5)
numbers1toN1[i] <- i
rank1[i] <- rank(sr1[], i) # rank of hospital i
}
for (i in 1:m1) {
prob.lowest1[i] <- equals(rank1[i] ,1 )
}
for (j in 1:m2) {
sr2[j]<-abs(s2[j]-0.5)
numbers1toN2[j] <- j
rank2[j] <- rank(sr2[], j) # rank of hospital i
}
for (i in 1:m2) {
prob.lowest2[i] <- equals(rank2[i] ,1 )
}
for (i in 1:m1) {
alpha[i]<-abs(prob.lowest2[i]-prob.lowest1[i])
}
alphap<- sum(alpha[])
p.value<- step(alphap - 0)
}
list(m1=21,d1=c(1,1,1,1,0,1,1,2,0,0,0,0,
0,1,0,1,0,0,1,0,0),
ce1=c(0,0,0,0,1,0,0,0,1,1,1,
1,1,0,1,0,1,1,0,1,1),m2=22, d2=c(1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,0,0,
```

0,0,0,0),ce2=c(0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,
1,1,1,1,1,1

How to cite this article by the APA method:

HAMIMES Ahmed, BENAMIROUCHE Rachid, CHELLAI Fatih (2021), A New Method for the Comparison of Survival functions using the Ranking of Median, Journal of quantitative economics studies, Volume 07 (Number 01), Algeria: Kasdi Marbah University Ouargla, PP. 427-435.