# Generalized additive models (GAMLSS) work better than traditional method in dealing with the right-censored survival data

Yuling Xiong
yxiong@murraystate.edu

Christopher Mechlin
cmechlin@murraystate.edu

## Abstract

Two methods are generated to test whether Generalized Additive Models for Location, Scale, and Shape (GAMLSS) performed better than traditional methods (like Logrank test, Cox regression, and Weibull regression) in right censored survival analysis. Based on the simulation results, method I MVSS, found that GAMLSS 2 is more suitable to deal with non-normal and heteroscedastic right censored survival data. Since GAMLSS has higher power over other traditional methods in detecting the difference of unequal variances between two groups, and has similar power in detecting the difference of unequal means between two groups. Although there are many distribution families fit GAMLSS model, we find that Weibull model to weibull data distribution is the best one matching the GAMLSS model without incurring a high type I error here. In method II hazard ratio, GAMLSS 2 has superiority over other methods in large sample size with the lowest type I errors but similar power comparing with other traditional methods.

## Introduction

In health field, survival analysis is often used to test whether a new drug is better than an old traditional one in treating the disease. Almost all survival data is right censored for the reason that it occurs after the end of observation. For instance, an experiment lasted for 100 weeks. If one patient disappeared in the $12^{th}$ week due to the immigration, financial problem, car accident or any other non-disease related reasons, random right censoring occurred; if one patient still survived at the end of the 100 weeks, it was unknown how long he or she would survive beyond that point, and then fixed right censoring occurred. The survival time defined by Forthofer, Lee, Hernandez (2007) is the length of time from a meaningful starting point until the time at which either some well-defined event happens, such as death or relapse to certain condition, or the study ends.

Due to the censored characters of the survival data, the survival analysis methods are different from the traditional t-test which suits for detecting the difference between means. Secondly, most survival data disobeys the assumptions of normal distribution and variance homogeneity for a t-test or ANOVA. The classic methods for survival analysis fall into three categories: non-parametric which includes the Kaplan-Meier estimator and the Log-rank test that is based on Kaplan-Meier estimator; semi-parametric which includes cox regression; and non-parametric which includes Weibull regression. Non-parametric methods have the advantage of fewer misinterpretations and it is more robust. We can guess that type I error for non-parametric methods would be lower than other parametric methods. At the same time, it loses power comparing with other parametric methods. As for the semi-parametric method, no assumptions about the baseline hazard are made comparing with the parametric method.

The objective of this paper is to determine whether Generalized Additive Models for Location Scale and Shape (GAMLSS) outperforms the traditional survival methods listed above in

right-censored survival data analysis, which will be assessed by comparing the powers of each method. Simulations will be run to test the powers of these methods in a variety of scenarios. The hypothesis is that GAMLSS will have higher power than traditional survival analysis methods.

## Methodology

1. GAMLSS

Generalized additive models for Location, Scale and Shape (GAMLSS), (Stasinopoulos, Rigby and Akantziliotou, 2008) are based on generalized additive models (GAM), which expand the exponential family distribution assumption of response variable to a general family distribution including highly skew or kurtotic distributions. A GAMLSS model assumes independent observations $y_i \, for \, i = 1,2, \dots, n$ with probability (density) function $f(y_i|\theta)$ conditional on $\theta$ where $\theta = (\theta_{i1}, \ \theta_{i2}, \dots, \theta_{ip})$ is a vector of p parameters, each of which is related to the explanatory variables. In many practical situations at most p=4 distribution parameters are required. The R implementation denotes these parameters as $(\mu_i, \ \sigma_i, \ \gamma_i, \ \tau_i)$, μ for location, $\sigma$ for scale, and shape for $\gamma \ and \ \tau$.

In this project, only the shape and the scale parameters are considered. Firstly, the simplest parametric linear GAMLSS model is: $g_k(\theta_k) = n_k = X_k\beta_k$, which has no additive terms in any of the distribution parameters. Secondly, after adding non-linear semi-parametric additive terms, it can be extended to non-linear parametric sub-model of GAMLSS: $g_k(\theta_k) = n_k = X_k\beta_k + \sum_{j=1}^{J_k} h_{jk}(x_{jk})$, which involves the use of non-parametric smoothing functions. Thirdly, in order to allow random effects terms to be included in the model for , $\gamma$, $\tau$ , it has been extended below:

$g_k(\theta_k) = n_k = X_k\beta_k + \sum_{j=1}^{J_k} Z_{jk}\gamma_{jk}$,

which fits the assumptions for survival analysis in dealing with the right censored data.

Two GAMLSS regression models were taken into consideration. The first one used a Weibull distribution that only modeled the location parameter ($\mu$), and the second one modeled both the location ($\mu$) and the scale ($\sigma$). We can guess that GAMLSS 2 has higher power than GAMLSS 1.

2. Choosing data distribution to generate survival times

Choosing the optimizing data distribution for the models is the first most important step before simulation. Most survival data have two characters: positive and right-skewed, which makes many continuous data distribution model fitting it, like exponential, Weibull, Gamma, Log-normal, Log-logistic, Gompertz distributions et al. In order to compare the GAMLSS with other traditional survival analysis methods (Log-rank test, Cox regression, and Weibull Regression), two methods based on effect sizes and powers are introduced. One method is MVSS (GAMLSS is compared in four independent directions, different means, variances, sample sizes and statuses). The other method is based on Hazard Ratio, which is used to re-check the validity of our hypothesis that GAMLSS works as well as or better than traditional methods in survival analysis. The data distributions should match not only the characters of survival times but also these two methods. Ralf, Thomas and Maria (2005) found that the results of simulation studies may depend on the choice of the distribution of the generated survival times rather than to develop the best measurement error model in a complex situation.

In method II, the hazard ratio is: $\theta = e^{\beta x} = h_1(t)/h_0(t)$. The hazard at time t for the group is: $h_i(t) = \frac{f_i(t)}{S_i(t)}$, i=0,1, where $f_i(t)$ is the rate of

failure events that happens per time unit, $S(t)$ is the probability of survival to time t. Because Gamma, Log-normal, and Log-logistic distributions disobey the assumption of proportional hazards property, only exponential, Weibull, and Gompertz data distributions are left to be considered. However, in method I MVSS, two groups (the experimental group and the placebo group) are manipulated in five different effect sizes in variance-direction. Gompertz distribution is excluded because of the complicated variance function, which is worthless by adding too much calculation work. As for the exponential distribution, it is the special case of the Weibull distribution when the sigma in Weibull equals to 1. Besides, if the survival times are generated by means of the exponential distribution, either the number of deaths or the attained age were too high in the simulated data. Mara (2008) got the conclusion that the likelihood-ratio test for overall model fit is significant only for the Weibull and its p-value is the only one less than 0.05 even though the log-normal has the largest maximized likelihood when he compared the regression models between Weibull, log-logistic, and log-normal distributions.

In summary, weibull data distribution maybe not the most accurate regression model to generate realistic survival data, but it is the most optimizing distribution to fit our model here and simplify the calculation.

## 3. Generating censored survival times

In R, a random Weibull sample data of n are generated by $x = rWEI(n, \mu, \delta)$ . Let $x_i$ represents the survival time of the $i^{th}$ individual. Due to the right censored characters, another random weibull sample data of n are generated by $y = rWEI(n, a * \mu, \delta)$, where $a \in [0,1]$ . Let $y_i$ represents the survival time of the $i^{th}$ individual. If $x_i$ is lower than $y_i$, then the $i^{th}$ individual is censored, which is recorded as 1 in the status, otherwise it is uncensored and recorded as 0. The parameter $a$ controls the percentage of censored data in the group.

## 4. Method I MVSS

M represents mean. In traditional methods, the significant difference between groups is generally determined by mean value test under equal variance condition. Cohen (2001) introduced effect size, which is useful to generate two groups with same variance but different means. Conhen's d is defined as the difference between two means divided by a standard deviation for the data:

$$Conhen'd = effect\ size = \frac{mean1 - mean2}{S_{pooled}}$$

Where $S_{pooled} = \frac{(n_1-1){s_1}^2+(n_2-1){s_2}^2}{n_1+n_2-2}$

Two groups with same variances but different means are manipulated by means of

Group1: y + effect .size*Spooled

Group2: y

Where y is the random data taken from the Weibull distribution $rWEI(n, \mu = 5, \delta = 2)$.

When a constant (effect .size*Spooled) is added, mean for group1 is changed but not variances. However, the coefficients chosen for $\mu$ and $\delta$ should make the constant (effect.size*Spooled) small enough. Otherwise the group1 generated by us maybe breaks the Weibull distribution.

An effect size of zero corresponds to no difference between groups, and an effect size of one is interpreted as one standard deviance between means. Effect sizes that are used in detecting the mean difference are 0, 0.1, 0.3, 0.5, and 1. Under these different effect sizes, power was evaluated through simulation. For each effect size, 1000 simulations are performed. Type I error is examined at the effect size of zero. The type I error of GAMLSS is no more than the

alpha value we set, which is the prerequisite before comparing the power at other effect sizes, and to detect in which effect size range GAMLSS has higher power to identify the difference between the data sets. Alpha value is the maximum probability that reject the null hypothesis when the null hypothesis is true, which is the same as controlling the type I error to a specified level. Generally alpha value equaling to 0.05 is often used. However, it is quite flexible choice that is up to the researchers. Knowledge of situation at hand will contribute to the choices of an appropriate alpha value. And researchers in biostatistics typically have the choice made for them by government regulations (Mecklin 2003).

V represents for the variances. First, two groups generated from the Weibull distribution are manipulated with the same mean but the different variances. The effect size defined for the different variances is analogous with Cohen's method:

$$effect\ size = \frac{var_1 - var_2}{var_1}$$

where $var_1$ is the variance of experimental group, and $var_2$ is the variance of placebo group. The effect sizes for the variance-direction are 0, 0.1, 0.2, 0.3, and 0.5.

The mean and variance for Weibull distribution are:

$$Weibull\ mean = \mu * gamma(1 + \frac{1}{\sigma})$$

$$Weibull\ variance = \mu^2(gamma\left(1 + \frac{2}{\sigma}\right) - gamma^2(1 + \frac{1}{\sigma}))$$

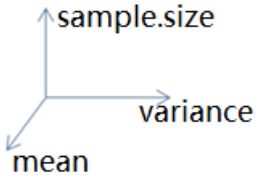
Since the property of $E(x + c) = E(x) + c, Var(x + c) = Var(x)$ , two groups that fit the assumptions in different variance effect size but with same mean are generated by manipulating the data below:

$$x1 = rWEI(n, \mu_1, \sigma) + mean1 - mean2$$
$$x2 = rWEI(n, \mu_2, \sigma)$$

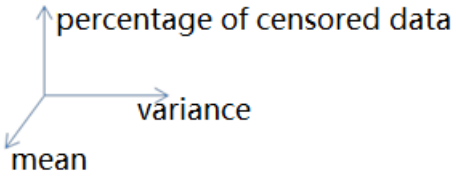
Where $\mu_2$ is $\sqrt{\frac{(var_1 - effect\ size * var_1)}{gamma\left(1+\frac{2}{\sigma}\right) - gamma^2(1+\frac{1}{\sigma})}}$.

In variance-direction, simulations are run to test the power with parameters $\mu_1 = 2\ and\ \sigma = \frac{1}{2}$. The constant (mean1-mean2) added should also be very small. Otherwise the random data x1 may disobey the Weibull distribution to some extent.

SS stand for sample size and percentage of censored status. The two comparing groups generated before are both in the same sample size and the same percentage of status. However, in realistic situation, maybe there are only 25 individuals in the experimental group, and 100 individuals in the placebo group. GAMLSS is tested again in different means and variances direction with the unequal sample sizes:




It works similar to the percentage of censored status direction, which is controlled by parametric $a$ of $a * \mu$ in the weibull function. Then GAMLSS is re-tested in different means and variances with the unequal percentage of censored data:




Any right censored survival data collected in real life could be written as the linear combination of MVSS: $survival\ data = p * mean + q * variance + r * sample\ size + s *$

$status$, where p, q, r, s are the weights in this four directions.

5. Method II hazard ratio

The foible of method I is that a small constant is added to the random data to generate two comparing groups, which may makes the new revised random data disobey the Weibull distribution to some extent and improves the measuring errors. So, hazard ratio is introduced to build a new effect size definition, which assists us to re-check the validity of GAMLSS.

The hazard ratio is:

$$\theta = effect\ size = e^{\beta x} = h_1(t)/h_2(t)$$

Where $h_1(t)$ is the hazard at time t for the experimental group, $h_2(t)$ is the hazard at time t for the placebo group. The hazard at time t for weibull distribution is $h(t) = \frac{\sigma * t^{\sigma-1}}{\mu^{\sigma}}$

And, $effect\ size = \frac{h_1(t)}{h_2(t)} = (\frac{\mu_2}{\mu_1})^{\sigma}$ , the effect sizes taken into consideration are 1, 1.5, 2, 2.5, and 3.

Then, the two groups generated by different $\mu$ below:

$$x1 = rWEI(n, \mu_1, \sigma)$$
$$x2 = rWEI(n, \mu_2, \sigma)$$

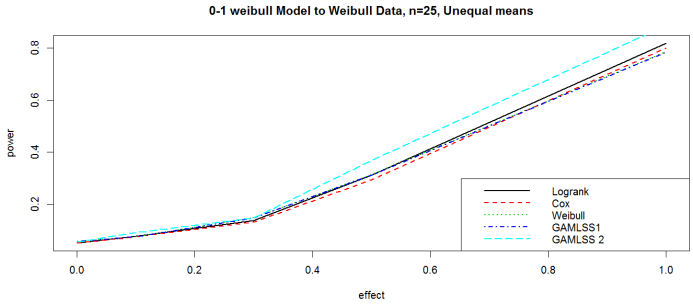
Both mean and variance for these two comparing groups are changed at the same time. The two moving directions couldn't be separated from each other in this method. But the random data are exactly taken from the Weibull distribution, which is a good way to recheck the hypothesis that GAMLSS works as well as or better than traditional methods in survival analysis.

6. Simulation

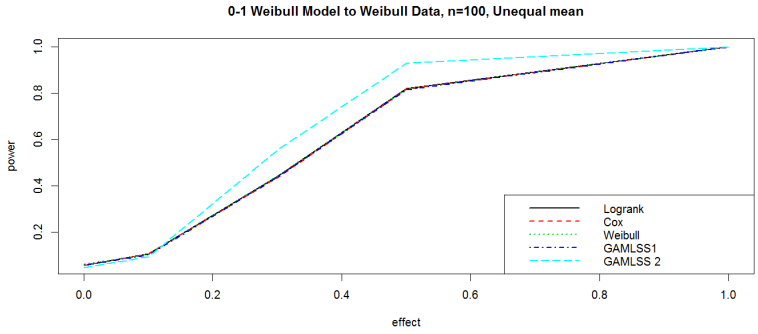
Only Weibull distribution is chosen to use in the simulation study, two sample sizes (25 and 100) were used to see if a large sample size improve the power. Alpha value is 0.05. For the convenience to recheck the correctness of the program, we set seed for each simulation.

## Results

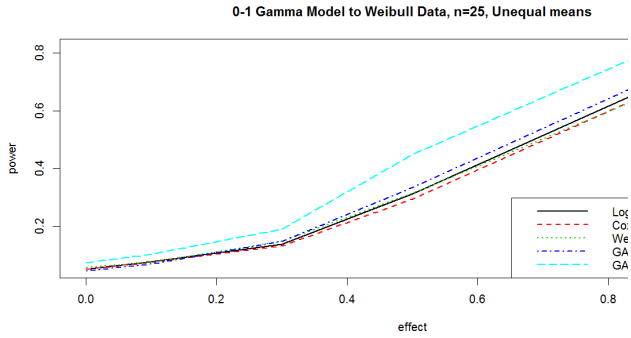
The simulation study proved quite effective. The elements analyzed were powers at different effect sizes and type I errors. In method I MVSS, simulation results for unequal means with the same sample size, the same percentage of censored data, and the same variances are showed below:

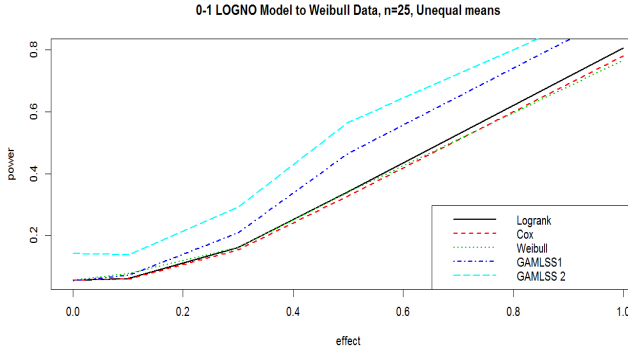



From the figure, it can be seem, GAMLSS 2 had higher power than traditional methods without incurring a high type I error in a small sample size n=25. As expected, higher power was achieved when fitting the GAMLSS 2 with both the location ($\mu$) and scale ($\sigma$) parameters comparing with GAMLSS 1 with just the location parameter.



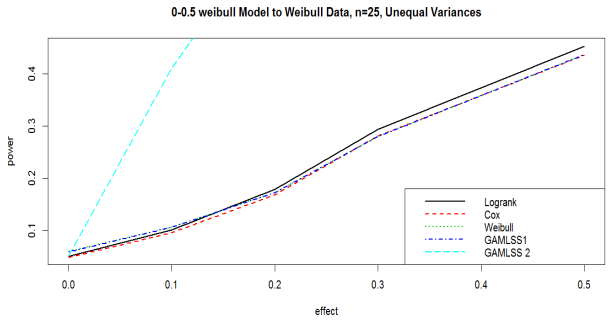

When sample size increased to 100, the type I error for the GAMLSS 2 was 0.046, which was the lowest one among five methods. Additionally, they all achieved higher power with larger sample size.

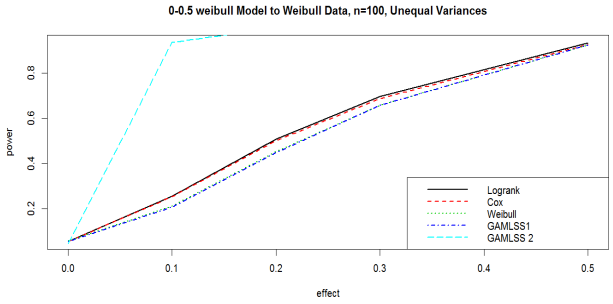


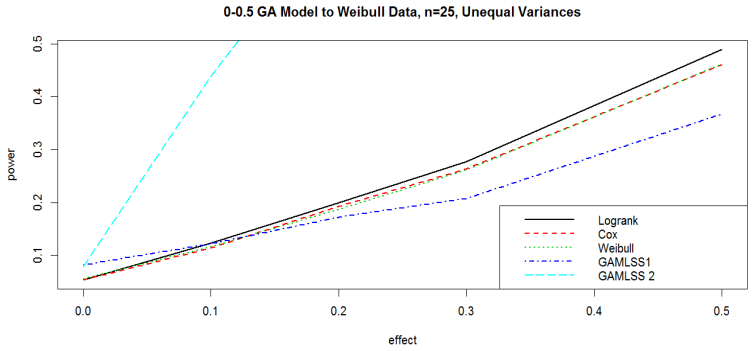

In order to check the assumption that Weibull family distribution is the best model fitting the GAMLSS, more simulations were done by adding Gamma and Log-normal models. From the figures above, GAMLSS 2 also had higher power. However, there was also a high type I error, which disobeys the prerequisite. Weibull family distribution to Weibull data is the best matching model.

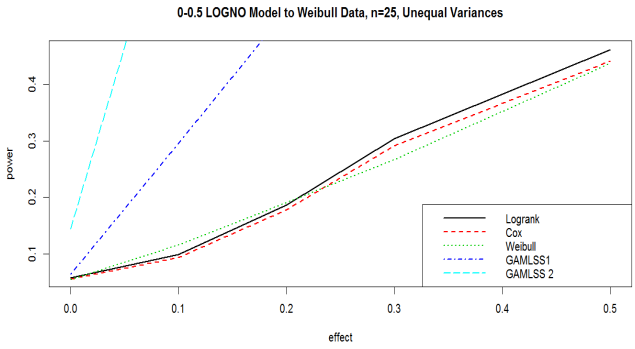
In the situation of unequal variances with the same sample size, the same percentage of censored data, and the same mean, simulation results are showed below:




GAMLSS 2 had advantages not only in low type I error but also in achieving higher power even in the small sample size comparing with other methods. And, it worked better in big sample size. Moreover, GAMLSS 2 is much more sensitive to test the difference in variance direction comparing with the mean direction.

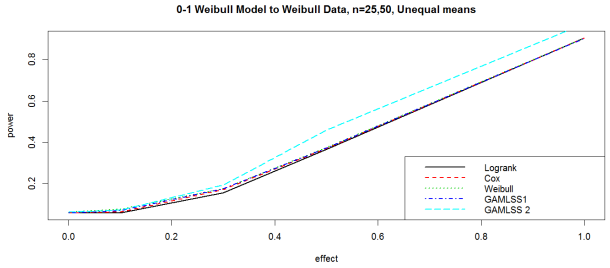



To check the Weibull family distribution model picked up for the GAMLSS, is the best one matching the model, Gamma and Log-normal models were added to run the simulation again. Quite similar to the result gained in mean-direction, both of them had high type I errors which wasn't at a manageable level. And the

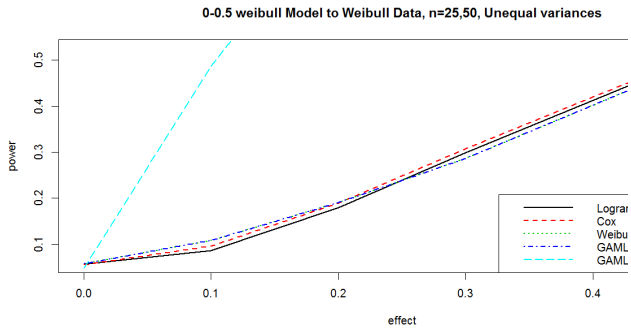
type I error in log-normal model is much higher than the Gamma one.

In the situation of unequal sample size (25 for group1 and 50 for group 2) of mean-direction, with the same percentage of censored data and the same variance



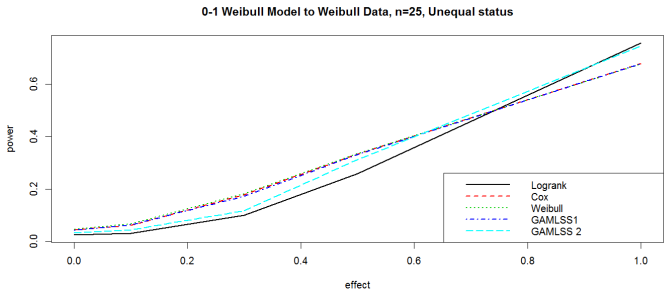

Comparing with the situation of equal sample size of mean-direction, no significant difference was found on the tendency of GAMLSS and the type I error in the situation of unequal sample size.

In the situation of unequal sample size (25 for group1 and 50 for group 2) of variance-direction, with the same percentage of censored data and the same mean



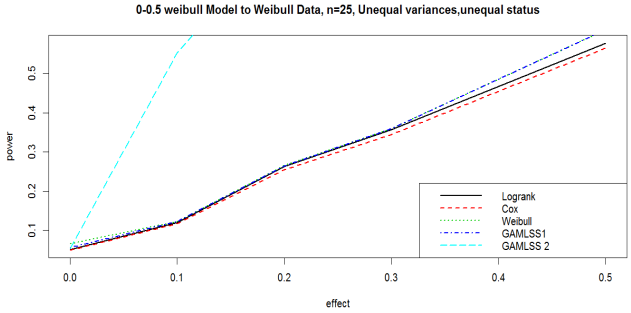

Similarly, no significant difference was found on the tendency of GAMLSS and the type I error in the situation of unequal sample sizes comparing with the situation of equal sample size of variance-direction. It seems they are all not sensitive to the change of sample sizes in both mean and variance directions. More simulations are needed by enlarging the difference of these two sample sizes to see whether the analysis methods are sensitive to it.

In the situation of unequal percentage of censored data of the mean-direction, with the same variance and sample size




GAMLSS performed almost the same as traditional methods. In group one 0.25 $\mu$ generated nearly 25% censored data. And in group two $1\mu$ generated nearly 50% censored data. Still more simulations are needed by changing the parameter $a$ of $a * \mu$ to see whether there is a turning point that GAMLSS works better.

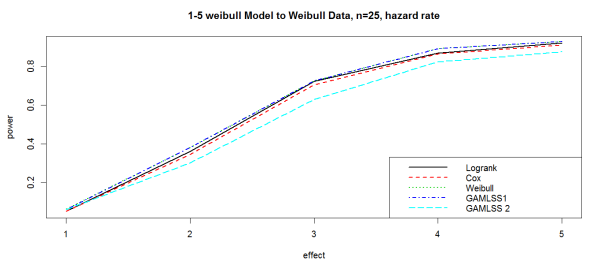
In the situation of unequal percentage of censored data of the variance-direction, with the same mean and sample size

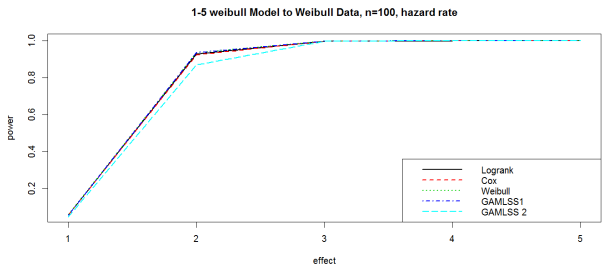



GAMLSS 2 continued to outperform other traditional methods in variance-direction even though in the situation of unequal status.

In method II hazard ratio

The two groups were generated exactly from the Weibull data distribution. Simulation results are showed below:




GAMLSS was not more powerful than traditional methods in small sample size, which indicated that it was worthless to loss that degree of freedom when GAMLSS was just as effective as traditional methods. However, GAMLSS performed almost the same as the traditional methods in large sample size with the lowest type I error again.

The complete tables and figures can be found in attachment I. The program for the simulation can be found in attachment II.

## Conclusion

In method I MVSS, Generalized Additive Models for Location, Scale, and Shape in survival analysis performed better than traditional methods, especially in variance-direction. GAMLSS is designed to deal with unequal variances with higher power. However, GAMLSS don't have significant advantage in mean-direction. Besides, GAMLSS is not significant to the changes occurred in sample size direction and the percentage of censored data direction. Log rank test has the smallest type I error in the small sample size. It is explainable for the reason that Log rank test belongs to non-parametric tests which are more robust and have fewer mistakes. However, when the sample size increases to 100, GAMLSS 2 takes the position as the lowest one in type I error. In order to manipulate data in these independent directions, measuring errors, resulted from adding constants to the groups for generating different effect sizes, couldn't be ignored.

Additionally, comparing with other distributions like gamma and log normal distributions, Weibull model to Weibull data distribution is the best one matching the GAMLSS model without incurring a high type I error.

In method II hazard ratio, when two groups are exactly generated from Weibull data distribution, the power for GAMLSS is almost the same with other methods, which is not effective to persuade others to use this method. Still, GAMLSS 2 has the lowest type I error in large sample size.

In conclusion, after considering the simulation results from method I and II, it seems that GAMLSS 2 has superiority over other methods in large sample size with the lowest type I errors. In some sophisticated experiments, type I error is required to much lower. Method I shows that GAMLSS 2 is more suitable to deal with non-normal and heteroscedastic data. As for the survival data taken from the Weibull distribution, method II shows that GAMLSS 2 performs the same with other traditional survival analysis methods.

It would be desirable to do more simulations by change the parametric in future research, which will assist us to understand the characters of GAMLSS better:

1. The sample sizes considered in this project are small (n=25) and large (n=100) ones. More different sample sizes should be tried between 25 and 100 to see whether there is a turning point which makes the GAMLSS 2 definitely has a lower type I error comparing with other methods.

2. Only two groups were compared here. It would be interesting to consider more than two groups at the same time to see what happens, which is a litter similar to ANOVA.

3. As for the sample size direction and the percentage of censored data direction, we should try more parametric to see whether GAMLSS has higher power when the difference between the sample sizes or the percentage of censored data is large enough.

GAMLSS might also has a high possibility to fit for the left and interval censored data analysis not only the right one. We will continue to strive for a greater understanding of the characters of GAMLSS before we can say the hypothesis that GAMLSS definitely has higher or almost the same power comparing with traditional survival analysis methods, or at least in some limited scenarios.

## References


Forthofer, R. N., Lee, E. S., & Hernandez, M. (2006).*Biostatistics: A guide to design, analysis, and discovery*. (2nd ed., p. 297). Elsevier.

Stasinopoulos, M., Rigby, B., & Akantziliotou, C. (2008).*Instructions on how to use the gamlss package in r*. (2nd ed., p.13-15) Unpublished manuscript, Retrieved from http://studweb.north.londonmet.ac.uk/~stasinom/papers/gamlss-manual.pdf

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate cox proportional hazards models. *Statistics in medicine*, *24*, 1718-1721. doi: 10.1002/sim.2059

Tableman, M. (2008). *Survival analysis using s/r*. (p.78) Unpublished manuscript, Department of Mathematics & Statistics, Portland State University, Portland, Retrieved from http://stat.ethz.ch/wbl/Skript_SurvivalAnalysis.pdf

Zhou, M. (n.d.). *Use software r to do survival analysis and simulation. a tutorial*. Unpublished manuscript, Department of statistics, University of Kentucky, Retrieved from http://www.ms.uky.edu/~mai/Rsurv.pdf

Suaray, K. (2012, 7). *Simulating censored data in r*. , Nashville, USA.

Cohen, J. (1992). A power primer. *Psychological Bulletin*,*112*(1), 155-159. doi: 10.1037/0033-2909.112.1.155

Mecklin, C. J. (2003). A comparison of equivalence testing in combination with hypothesis testing and effect size. *Journal of modern applied statistical methods*, *2*, 329-340.