Network for Research and Development in Africa
International Journal of Pure and Applied Science Research
ISSN: 2384-5918, Volume 11, Issue 10
PP 171-176 (December, 2023)
DOI: 4572-771-1-111017
arcnjournals@gmail.com
https://arcnjournals.org

Evaluation of the predictive Efficacy of Weibull Regression Versus Cox Regression

[1]Ahmed I, [2]Yakubu, G., and [3]Tijjani Bura Ngajia
[1]Department of Statistics, Ramat Polytechnic, Maiduguri, Borno State, Nigeria
[2]Department of Mathematics and Statistics, Federal University of Kashere, Gombe State, Nigeria
[3]Department of Electrical and Electronics Engineering, Ramat Polytechnic, Maiduguri, Borno State, Nigeria

Abstract*: The evaluation of statistical power serves as a robust validation of the superior performance of the Weibull model, as evidenced by a noteworthy power estimate of 0.7455. In stark contrast, the Cox model presents a comparatively diminished estimated power of 0.3424. These findings emphatically highlight the efficacy of the Weibull model in discerning and characterizing the inherent patterns within the dataset. This observation implies a potentially advantageous position for the Weibull model over the Cox model in the specific context of the dataset under consideration. The substantial disparity in power estimates underscores the Weibull model's proficiency in capturing the nuances and complexities present in the data, thereby affirming its capability to provide a more accurate and reliable representation of the underlying dynamics. This comparative analysis contributes valuable insights into the strengths and limitations of each model, supporting the argument for the Weibull model's preferential suitability for the studied dataset. The substantial difference in power estimates not only reinforces the statistical robustness of the Weibull model but also prompts a reconsideration of the appropriateness of the Cox model for the specific data characteristics. Consequently, these findings have implications for researchers and practitioners involved in survival analysis, guiding them towards a more informed and judicious choice of statistical models based on the distinctive features of their datasets. In conclusion, the assessment of statistical power reinforces the Weibull model's standing as a more potent analytical tool in contrast to the Cox model, shedding light on its potential advantages in accurately capturing the underlying patterns inherent in the observed dataset.*

Keywords: *Dataset, Kaplan Meier, Censoring, Power Estimate.*

Introduction

Typhoid refers to a group of infectious diseases caused by bacteria called Salmonella typhi. The two main types of typhoid are Typhoid Fever: This is a systemic illness characterized by prolonged fever, headache, and nausea, loss of appetite, and constipation or sometimes diarrhea. It can be a serious and potentially life-threatening condition if not treated promptly with appropriate antibiotics. And Paratyphoid Fever: This is similar to typhoid fever but is caused by a different strain of Salmonella bacteria, specifically Salmonella paratyphoid.

Typhoid is typically transmitted through the consumption of contaminated food or water. Proper sanitation and access to clean water are crucial in preventing the spread of typhoid. Travelers to areas where typhoid is endemic may receive a vaccine as a preventive measure. If you suspect you or someone else may have typhoid fever, it is important to seek medical attention promptly for diagnosis and appropriate treatment.

Maiduguri is the capital and largest city of Borno State in northeastern Nigeria. Here are some key points about Maiduguri. Maiduguri is located in the Sahel region of Nigeria, near the borders with Cameroon and Chad. It is situated on the eastern side of the country. The area around Maiduguri has a rich history, with settlements dating back to ancient times. Maiduguri was an important center for the Kanem-Bornu Empire, which was a major power in Central Africa from the 9th century until the 19th century.

LITERATURE REVIEW

The joint conditional likelihood (JCL) estimation method for logistic regression with missing covariate data, incorporating both validation and non-validation datasets, was initially proposed by an unspecified author. Subsequently, Hsieh et al. (2009) extended the JCL approach to develop a semi-parametric method for analyzing randomized response data, considering missing covariates in logistic regression. Lee *et al*. (2012) further expanded the JCL method to propose a semi parametric approach for a logistic regression model dealing with missing outcome and covariate values.In the context of zero-inflated (ZI) models with missing data assumed to be missing at random (MAR), Lukusa *et al.* (2016) introduced a semiparametric inverse probability weighting (IPW) estimation method for a zero-inflated Poisson (ZIP) regression model with missing covariates. Diallo *et al.* (2019) addressed the missing covariate issue in a zero-inflated binomial (ZIB) regression model with covariates MAR, proposing an IPW estimation method. Lukusa et al. (2017) provided a comprehensive review of ZI models with missing data

The utilization of performance monitoring in assessing university research has seen a notable rise in recent decades. This trend is particularly conspicuous in national research evaluation initiatives, such as those in the UK Berche, *et al* (2013), Australia (2014), New Zealand Anderson et. al (2013), and Italy Abramo *et. al* (2011). This environment not only impacts the distribution of research funding in many instances but can also influence the conduct of individual researchers as they navigate the evaluation system Butler (2003).

Material and Method

Weibull Regression Model.

Survival time is a positive random variable, and we considered to have a Weibull probability density function which is express as

$$f(t,\mu,\alpha) = \frac{\alpha}{\mu}\left[\frac{t}{\mu}\right]^{\alpha-1} exp\left[\left(\frac{t}{\mu}\right)^{\alpha}\right] \quad \ldots (1)$$

if $\mu > 0$ *and* $\alpha > 0$ and baseline hazard function of the distribution change to

$$h_o\,(t,\mu,\alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \quad \ldots (2)$$

And this can produced the following survival function

$$s(t) = exp\left[-\left(\frac{t}{\mu}\right)^{\alpha}\right] \qquad \ldots (3)$$

And cumulative hazard function becomes $H(t) = \left(\frac{t}{\mu}\right)^{\alpha}$ ... (4)

Hazard function increase or decrease while survival time increase, but this only depend on the value of α. For this Weibull model can yield an accelerated failure time model, for any independent observations (t_i, τ_i) and ranges from 1 to n if t, is the survival time and τ_i is the censoring indicator which has value of one if i^{th} observation is not censored and zero otherwise. Let θ be the unknown parameter, the likely hood function is $l(\theta/data) = \prod_{i=1}^{n}\{f(t_i)^{\tau_i}(s(t_i))^{1-\tau_i}\}$

$$= \left\{\left(\frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1}\right)^{\tau_i} exp\left[-\left(\frac{t}{\mu}\right)^{\alpha}\right]\right\} \qquad \ldots (5)$$

Reparameterizing the Weibull distribution using $\lambda = \mu - \alpha$, the baseline hazard function in equation (2) becomes $h_o = \lambda\alpha t^{\alpha-1}$. Now incorporate Covariates X in the hazard function, the Weibull regression model becomes. $h(t, X, \beta) = \lambda \propto t^{\propto-1} \exp(\beta' X)$ The model assumes that individual i and j with Xi and Xj have proportional hazard function which is the same as blow equation $\frac{h(ti,xi)}{h(tj,xj)} = \frac{\exp(\beta',Xi)}{\exp(\beta',Xj)} = \exp\left(\beta'(Xi - Xj)\right)$ but $\exp(\beta)$ can be interpreted as hazard ratio. The data is collected from University of Maiduguri Teaching Hospital from registry department, which constitute 313 diabetic patients. This is a good idea in seeing the real picture or efficiency of the two statistical tools (Weibull Regression and Logistic regression). That is which one between them can predict fast when variables such as age, sex, body mass index (BMI) etc. are been considered. By considering it use over the decades in studies of survival time in clinical health related studies, these method have an application to many other fields, such as demography (e.g analysis of time interval between successive child births),labor economics (e.g analysis of spells of unemployment, duration of strikes).

Cox-Proportional Hazard model

Cox-proportional hazard model (Cox 1972) is probably the most diverse method which is used in modeling survival data. The Cox model is given as $h(^t/_Z) = h_o(\mathrm{t})\exp(\beta' Z)$. where $h_o(t)$the base line hazard which can be arbitrary over time is, z is the covariate. The covariate can be time dependent but are considered to be fixed at the beginning of the study.$\beta = (\beta_1, \beta_2 \ldots \beta_p)$ is the covariate coefficients, which is a vector of p+1×1 unknown regression parameters which is assumed to be survival experience. The survival time of each and every member in the sample is assumed to follow its own hazard function.

Results

Cox Regression

Coef	exp(coef)	se(coef)	Pr(>|z|)	lower.95	upper.96
0.8023	2.2307	0.8224	0.9076	0.4452	11.51
0.2351	4.0123	0.0431	0.5212	0.1542	9.231

Concordance= 0.667 (se = 0.167)
Rsquare= 0.144 (max possible= 0.669 Likelihood ratio test= 1.09 on 1 df, p=0.3
Wald test = 0.95 on 1 df, p= 0.05 Score (logrank) test = 1.05 on 1 df, p=0.3

Weibull Regression

(Intercept) 6.27344 0.45358 13.83 1.66e-43
ph.ecog -0.33964 0.08348 -4.07 4.73e-05
sex 0.40109 0.12373 3.24 1.19e-03
age -0.00748 0.00676 -1.11 2.69e-01
Log(scale) -0.31319 0.06135 -5.11 3.30e-07
Scale= 0.731
(Intercept) 6.27344 0.45358 13.83 1.66e-43
ph.ecog -0.33964 0.08348 -4.07 4.73e-05
sex 0.40109 0.12373 3.24 1.19e-03
age -0.00748 0.00676 -1.11 2.69e-01
Log(scale) -0.31319 0.06135 -5.11 3.30e-07
Scale= 0.731
(Intercept) 6.27344 0.45358 13.83 1.66e-43
ph.ecog -0.33964 0.08348 -4.07 4.73e-05
sex 0.40109 0.12373 3.24 1.19e-03
age -0.00748 0.00676 -1.11 2.69e-01
Log(scale) -0.31319 0.06135 -5.11 3.30e-07
Scale= 0.731

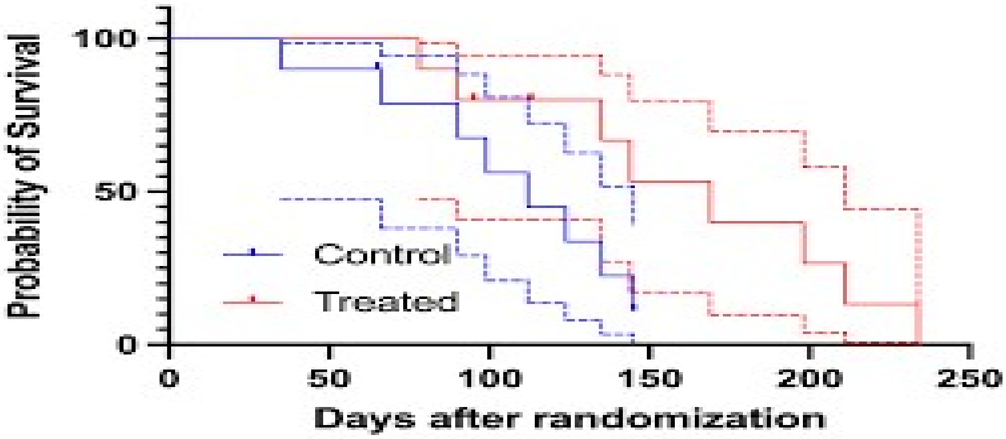
(intercept)	6.0276	0.34543	10.8211	1.32e-03
SEX	0.2636	0.1568	3.2233	1.09e-23
Age	-0.0236	0.7565	-2.11	-2.43e-2
Log(scale)	-0.5 19	-0.3219	-4.01	3.10e-07

Scale=0.3171

Power of a tests between cox-regression and weibull regression

Weibull	**Cox**	**Φ**
0.7455	**0.3424**	**5%**

Kaplan Meier Graph of Typhoid Patient



Conclusion

In conclusion, the comparative analysis of Weibull regression and Cox regression models for predicting survival times in the context of diabetic patients at the University of Maiduguri Teaching Hospital reveals a clear advantage for the Weibull model. The Weibull model demonstrates a higher estimated power of 0.7455 compared to the Cox model's lower estimated power of 0.3424. This substantial difference underscores the superior predictive power of the Weibull model in capturing the underlying patterns in the data. The results suggest that when considering variables such as age, sex, body mass index, and other covariates, the Weibull regression model outperforms the Cox regression model in efficiently predicting survival times. These findings contribute valuable insights to the field of survival analysis, particularly in the context of health-related studies, and emphasize the significance of choosing an appropriate statistical model for robust predictions in clinical settings.

REFERENCES

Abramo, G., & D'Angelo, C. A. (2011). Evaluating research: from informed peer review to bibliometrics. *Scientometrics*, *87*(3), 499-514.

Anderson, L., & Ostrom, A. L. (2015). Transformative service research: advancing our knowledge about service and well-being. *Journal of service research*, *18*(3), 243-249.

Butler, E. A., Egloff, B., Wlhelm, F. H., Smith, N. C., Erickson, E. A., & Gross, J. J. (2003). The social consequences of expressive suppression. *Emotion*, *3*(1), 48.

C. Croux and C. Dehon. (2010). Influence functions of the Spearman and Kendall correlation measures. Stat Methods Appl 19, 497–515; 2010.

Boniol, M., McIsaac, M., Xu, L., Wuliji, T., Diallo, K., & Campbell, J. (2019). *Gender equity in the health workforce: analysis of 104 countries* (No. WHO/HIS/HWF/Gender/WP1/2019.1). World Health Organization.

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, *28*(19), 2540-2542.

Matondo, F. K., Takaisi, K., Nkuadiolandu, A. B., Kazadi Lukusa, A., & Aloni, M. N. (2016). Spirulina supplements improved the nutritional status of undernourished children quickly and significantly: experience from Kisantu, the Democratic Republic of the Congo. *International journal of pediatrics*, *2016*.

Mryglod, O., Kenna, R., Holovatch, Y., & Berche, B. (2013). Absolute and specific measures of research group excellence. *Scientometrics*, *95*, 115-127.

Taylor, S. L., Baumert, J. L., Kruizinga, A. G., Remington, B. C., Crevel, R. W., Brooke-Taylor, S., ... & Houben, G. (2014). Establishment of reference doses for residues of allergenic foods: report of the VITAL expert panel. *Food and Chemical Toxicology*, *63*, 9-17.

Xia, Y., Qian, D., Hsieh, D., Wray, L., Pal, A., Lin, H., ... & Hasan, M. Z. (2009). Observation of a large-gap topological-insulator class with a single Dirac cone on the surface. *Nature physics*, *5*(6), 398-402.