ISSN 2278 – 0211 (Online)

Survival Modeling of Unemployment Duration Experience of Staff of National Bureau of Statistics, Ibadan, Nigeria

Ronke, F. Odeleye

Student, Department of Statistics, Federal Polytechnic, Bida, Nigeria

Bushirat T. Bolarinwa

Chief Instructor, Department of Statistics, Federal Polytechnic, Bida, Nigeria

Ismaila A. Bolarinwa

Chief Lecturer, Department of Statistics, Federal Polytechnic, Bida, Nigeria

Abstract:

This article sought to model unemployment duration of staff of National Bureau of Statistics (NBS), Ibadan Office. Structured questionnaires were administered on 150 staff out of which 119 were duly completed and returned. Survival model employed was Kaplan Meier model. Survival curves of males and females were compared using the Log-rank test. It was found that 91.2% of males and 91.6% of females were unemployed for 12 months, 20.9% of males were unemployed for 144 months while 35.5% of females were unemployed for 216 months; no male was unemployed after 264 months while no female was unemployed after 240 months. Over 50% of males and females were unemployed for 60 months (5 years). The survival probabilities are higher for females than for males for survival times ($t = 12, 24, 36, 48$ and 60) shared by both males and females. However, log-rank test produced a p -value of 0.259 which led to non-rejection of the null hypothesis of no difference in population curves of males and females. It was concluded that there was no significant difference in the unemployment duration of male and female staff. The need to encourage entrepreneurial education, revisit power sector privatization and sustain efforts at ending insurgency, banditry, kidnapping and the likes was recommended.

Keywords: Survival, unemployment, Kaplan Meier, log-rank

1. Introduction

Survival analysis is a class of statistical method aimed at studying the occurrence and timing of events of interest. The analysis is propelled by event while time is the central operator. In medicine, event may be death, relapse, in which wise, time may be time to death and time to relapse respectively. In economics, the event may be employment while time is time to employment, that is, unemployment duration. In Civil Engineering, the event may be completion of a project while time is project duration; the event may also be appearance of a crack on a building while the time is time that lapses between project completion and appearance of the crack. It is thus, a concept applicable to virtually all aspects of human endeavour where event and time to event are clearly defined.

Unemployment duration study is very important as it explains changes in the labour market situation. Such a study is capable of informing policy makers about effectiveness of policies formulated to tackle unemployment in an economy. An unemployment duration study via survival analytical approach that signifies that unemployment duration has reduced significantly will naturally be pleasing to the government. A contrary result will suggest that government approach to tackling unemployment needs to be revisited. The need for such study from time to time can therefore, not be overemphasized.

A large amount of efforts has gone into survival analysis research. Although most of the efforts seem to be in the medical sciences, some applications have been made in other areas also. Leo and Go (1977) reviewed the common statistical technique employed to analyze survival data in public research; Tatsiramos (2006) studied the effect of unemployment insurance on unemployment duration and the subsequent employment stability using mixed proportional hazard model; Blanchard and Diamond (2008) examined unemployment duration dependence and suggested that this affects both the matching and the wage function. Daniela-Emanuela and Cirnu (2014) studied unemployment duration in Romania using survival methods; Novella and Duvivier (2015) examined the relationship between unemployment duration and education in Belgium; Oujezsky, Horvath, and Skorpil (2016) applied survival methods to analyze botnet command and control traffic using Kaplan-Meier estimator. Echeburua, Gomez and Freixa (2017) examined schizophrenia patients with gambling disorder using Cox survival model.

These research objectives are to apply Kaplan-Meier survival model to unemployment duration data and to also compare the unemployment duration experience of males and females.

The remaining part of the article is organized as follows: Section 2 presents Methodology; Section 3 presents Results and Discussion while Section 4 concludes the article and recommends.

2. Methodology

This section presents data collection and discusses survival analytical tools employed.

3. Data

A total of 150 questionnaires were administered on staff of National Bureau of Statistics (NBS), Ibadan, Nigeria out of which 119 duly were filled and returned. The critical event of interest was first employment and time to first employment was time before securing the first employment after graduation from school, that is, unemployment duration. Unemployment duration was measured in months.

3.1. Kaplan-Meier Survival Model

There exist three approaches to estimating the survival function. These approaches fall under three main headings namely, nonparametric, parametric and semi-parametric. The Kaplan-Meier model due to Kaplan and Meier (1958) belongs to the first class and is probably the most popularly used of all methods. This estimator is a product limit estimator, likelihood-based estimator (Cox & Oakes, 1984) and redistribute to the right estimator. The development of the model can be motivated from each of the three stated perspectives. In explaining this model, let it be known that survival within the context of this study implies unemployment. Survival time hence, implies unemployment duration.

Let us think of the time t as beginning of a short time interval terminating at time $(t+1)$ and n_t as the number of subjects available at the beginning of the interval and hence, at the risk of experiencing the event of interest (being employment in this case) during the short interval afterwards. Let us denote as d_t the number of subjects that experience the event in the short interval after t , number of subjects surviving the interval is $(n_t - d_t)$. Consequently, this number (being n_{t+1}) is the number of subjects starting interval $(t+1)$.

If we define by p_j the conditional probability of surviving the j th month after having survived month $(j-1)$, then, the probability, $S(t)$ of surviving months after the initial event, called the survival function is defined

$$S(t) = p_1 \cdot p_2 \cdots p_t$$

But

$$p_t = 1 - \frac{d_t}{n_t}$$

It follows that

$$s(t) = \left(1 - \frac{d_1}{n_1}\right) \left(1 - \frac{d_2}{n_2}\right) \cdots \left(1 - \frac{d_t}{n_t}\right)$$

The survival function $S(t)$ is therefore, defined

$$s(t) = \prod_t \left(1 - \frac{d_t}{n_t}\right)$$

Successive survival probability can be obtained by utilizing

$$s(t) = s(t-1) p_t$$

It is worthy of mention that at $t = 0$, $S(0) = 1$, meaning that all subjects are available at time zero.

3.2. Comparison of Survival Curves

In many survival studies, interests include a comparison of two or more survival curves. In medicine, it may be of interest to compare the survival experience of patients on two or more different treatments for a particular disease; in economics, interest may focus on comparing unemployment duration of graduates of different disciplines. Parametric and non-parametric methods exist in the literature for comparing survival curves. Non-parametric methods in the literature include: Mantel-Haenszel logrank test, Peto and Peto's version of logrank test, Generalized Wilcoxon test, and non-parametric version of Cox's F-test. The non-parametric tests are fairly robust, simple and intuitive and efficient relative to their parametric counterparts.

The Mantel-Haenszel logrank test, sometimes called Mantel-Cox test is the most widely used method for comparing survival curves (Machin, Cheung & Parmar, 2006). The procedure can be motivated as follows:

If we present the K distinct and ordered event times by $t_1, \dots, t_K$ and for a particular event time, say j -th, we denote by d_{Aj} and d_{Bj} , the number of subjects that experience the event in Groups A and B respectively; let r_{Aj} and r_{Bj} represent the number at risk at j -th event time in Groups A and B respectively. The Mantel-Haenszel logrank test is based on the statistic:

$$\chi^2_{Logrank} = \frac{\left\{ \sum_{j=1}^K (d_{Aj} - r_{Aj} \frac{d_j}{r_j}) \right\}^2}{\sum_{j=1}^K \frac{r_{Bj} r_{Aj} (r_j - d_j)}{r_j^2 (r_j - 1)}} \sim \chi_1^2 \quad (2.1)$$

If the contingency tables are independent, (2.1) is chi-square distributed with 1 degree of freedom. Equation 2.1 is equivalent to

$$\chi^2_{Logrank} = \left\{ \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B} \right\} \sim \chi^2_1 \quad (2.2)$$

where O_A and O_B are the total numbers of observed event in Groups A and B respectively and E_A and E_B are the total numbers of expected event.

Under the null hypothesis that there is no difference between two population survival curves, Equations 2.1 and 2.2 are chi-square distributed with 1 degree of freedom. The log-rank test can also be used to compare three or more survivor curves, as usual, the null hypothesis remains that all survival curves are the same (Kleinbaum & Klein, 2005).

4. Results and Discussion

The results of analysis are condensed in Tables 1 to 3. Survival experience of males is presented in Table 1 while that of females is in Table 2. Table 3 presents the results of log-rank test.

Time	n.risk	n.event	Survival	Std. Error	Lower 95% CI	Upper 95% CI
12	68	6	0.912	0.0344	0.6488	0.982
24	53	5	0.826	0.0481	0.7367	0.926
36	39	5	0.720	0.0609	0.6099	0.850
48	29	5	0.596	0.0714	0.4711	0.753
60	20	3	0.506	0.0771	0.3758	0.682
72	15	1	0.473	0.0790	0.3406	0.656
84	13	1	0.436	0.0808	0.3034	0.627
96	10	1	0.393	0.0837	0.2585	0.596
132	5	1	0.314	0.0970	0.1714	0.576
144	3	1	0.209	0.1072	0.0768	0.571
264	1	1	0.000	NaN	NA	NA

Table 1: Kaplan Meier Results for Males

Table 1 presents the survival probabilities for eleven distinct and ordered survival times and the associated 95% confidence intervals. The probability that a male remained in unemployment for up to 12 months is 0.912, implying that 91.2% of males remained in unemployment for up to 12 months. As high as 82.6 % remained unemployed for 24 months while 20.9 % remained unemployed for 144 months. No male remained in unemployment for more than 264 months.

Time	n.risk	n.event	Survival	Std. Error	Lower 95% CI	Upper 95% CI
12	51	2	0.961	0.0272	0.909	1.000
24	37	4	0.857	0.0547	0.756	0.971
36	24	1	0.821	0.0630	0.707	0.954
48	19	3	0.692	0.0868	0.541	0.884
60	13	1	0.638	0.0950	0.477	0.855
96	6	1	0.532	0.1253	0.335	0.844
216	3	1	0.355	0.1672	0.141	0.893
240	1	1	0.000	NaN	NA	NA

Table 2: Kaplan Meier Results for Females

Table 2 suggests that the probability that a female remained in unemployment for 12 months is 0.961. A staggering 85.7% of females remained in unemployment for 24 months while 35.5% remained for 216 months. Females that remained unemployed for 60 months constituted 63.8 % of female job seekers. No female remained unemployed beyond 240 months. Tables 1 and 2 share five survival times ($t=12, 24, 36, 48$ and 60) in common.

A comparison of the survivorship probabilities for these survival times shows that for each case of t , female has higher probability of remaining unemployed than male. This is rather a surprise as it contradicts the yet to be substantiated general belief that females tend to secure paid jobs faster than their male counterparts due to "connection". Other survival times are not directly comparable owing to differences in survival times, even if they were, statistical procedure will usually be required to establish a significant difference.

A situation where Over 50% of males and females were unemployed for 60 months (5 years) can only be suggestive of an unacceptably high unemployment rate. It paints a precarious situation for Nigeria's economy. When one considers this against the backdrop that exists no social welfare scheme for the citizenry, one should wonder about the impact of such scenario on the living standards of the citizenry.

The results portray unemployment situation years ago. This situation has consistently worsened owing to poor governance and the present situation should be better imagined than described. Efforts at revamping the power sector have failed woefully due to corruption. Unfortunately, no meaningful level of employment can be generated where power supply is grossly inadequate or unaffordable. The unemployment situation has been further aggravated by insecurity in form of insurgency, kidnapping, banditry and the likes. Security of life and property is a necessary ingredient for appreciable economic development.

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Gender=0	68	30	26.5	0.455	1.28
Gender=1	51	14	17.5	0.691	1.28

Table 3: Results of Log-rank Test
Chisq=1.3 on 1 degrees of freedom, p=0.259

A statistical comparison of the survivor curves for males and females produced a p-value of 0.259. The null hypothesis of no difference in population survival curves could not be rejected at 5% level. Hence, there is no significant difference in unemployment duration of males and females.

5. Conclusion and Recommendation

The article has modeled the unemployment duration experience of staff of NBS, Ibadan; Over 50% of males, and females were still unemployed five years after leaving school. No significant difference was found between unemployment duration of males and females.

Towards reducing unemployment duration and hence, unemployment rate, the following recommendations are hereby, made:

- Entrepreneurial education and improved access to capital should be encouraged for the youths to become job creators rather than job seekers.
- Power sector privatization should be revisited for the desired much better performance.
- Coupled with right policies aimed at discouraging unsolicited importation, the citizenry should be educated on the need to patronize home made goods.
- Efforts at ending insurgency, banditry, kidnapping and other social vices that discourage both local and foreign investors should be sustained.
- A study of relationship between unemployment duration and educational level should be worth pursuing by other researcher(s).

6. References

- Blanchard, O.J. & Diamond, P. (2008). The flow approach to labour market. *The American Economic Review*, 8(2), 345-359.
- Cox, D.R. & Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman & Hall.
- Daniela-Emanuela, D. & Cirnu, D. (2014). Unemployment duration in Romania. *Procedia Economics and Finance*, 15, 1173-1184.
- Echeburua, E., Gomez, M. & Freixa, M. (2017). Prediction of relapse after chronic behavioral treatment of gambling disorder in individuals with chronic schizophrenia: A survival analysis. *Behavior Therapy*, 48(1), 69-75.
- Kaplan, E.L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(1,2), 457-481.
- Kleinbaum, D.G. & Klein, M. (2012). *Survival Analysis* (3rd ed.). New York: Springer.
- Lee, E.T. & Go, O.T. (1997). Survival analysis in public health research. *Annual Review of Public Health*, 18, 105-134.
- Machin, D., Cheung, Y.B. & Parmar, M.K.B. (2006). *Survival Analysis-A Practical Approach* (2nd ed.). West Sussex: Wiley.
- Novella, M.L. & Duvivier, M. (2015). The relationship between unemployment duration and education- The case of school leavers in Belgium. (Federal Planning Bureau Working Paper 10-15). Brussels: Federal Planning Agency.
- Oujesky, V., Horvath, T. & Skorpi, V. (2016). Modeling Bonet C & C Traffic Lifespans from Netflow using survival analysis (50-55), *Proceedings of 39th International Conference on Telecommunications and Signal Processing (TSP 2016)*. New York: Curran Associates.
- Tatsiramos, K. (2006). The effect of unemployment insurance on unemployment duration and the subsequent employment stability. (The Institute for the Study of Labor (IZA) Discussion Paper 2280). Bonn: IZA.