Technical Note—A Note on the Nonparametric Analysis of the Stollmack-Harris Recidivism Data

Bruce W. Turnbull,

OPERATIONS RESEARCH, Vol. 25, No. 4, July-August 1977

Technical Notes

A Note on the Nonparametric Analysis of the Stollmack-Harris Recidivism Data

BRUCE W. TURNBULL

Cornell University, Ithaca, New York

(Received original June 1976; final, November 1976)

We show how quantitative methodology from the field of medical trials can be applied to the analysis of recidivism data from correctional release programs. The methods used are nonparametric and so do not depend on any assumptions concerning an underlying model. The data of Stollmack and Harris are used for illustration.

IN AN INTERESTING and stimulating paper, Stollmack and Harris [9] use failure-rate analysis to compare the results of two types of release programs in the District of Columbia. One of their stated aims is to show how methods in reliability theory can be applied to recidivism data. Using some results of Barlow and Proschan [1] and others, they develop an F-statistic for testing the null hypothesis H_0 of no difference between the experiences of the two programs. The statistical test is based on the assumption that time to rearrest (failure) is exponentially distributed (constant hazard rate). The authors recommend that a preliminary goodness-of-fit test for exponentiality should be performed, and they describe such a test, which also leads to an F-statistic. For uncensored data, this goodness-of-fit test has been developed further by Harris [3], and, in fact, has wider applicability. For example, Thiagarajan and Harris [10] have shown how to adapt the test for Weibulls. However, as Stollmack and Harris (p. 1197) point out, their goodness-of-fit test can have poor power properties against certain alternatives. Also, the F-test in general is not very robust (see [8], Section 10.2). Stollmack and Harris (p. 1204) leave open the question of how to handle the problem when the constant hazard rate assumption does not obtain. Another problem in applying their goodness-of-fit test is that a substantial number of success times are greater than the largest failure time. An alternative goodness-of-fit test based on a likelihood ratio statistic with an asymptotic χ^2 distribution is described in Turnbull and Weiss [11].

There are two methods that have been used in medical contexts that seem to be applicable to recidivism data. The methods are nonparametric and so, in particular, do not depend on the assumption of a constant failure rate. The first approach is due to Gehan [2] (see also Mantel [5]). Let $m_A=185$ be the number of cases in the first group and $m_B=211$ be the number in the second group. Consider the pooled sample of $m_A+m_B=396$ cases. For the ith case define the score U_i to be the number of remaining 395 observations that are definitely greater than it minus the number that are definitely less than it. Gehan's statistic is $W=\sum_A U_i$, where the summation is taken over just the scores of the 185 cases in group A. Under the null hypothesis H_0 of homogeneity, W has mean zero and variance $V=m_Am_B\sum_{i=1}^{m_A+m_B} U_i^2/(m_A+m_B)(m_A+m_B-1)$. An exact test can be performed, but asymptotically $W/\sqrt{V}$ has a standard normal distribution under H_0. For the data in Tables I and II of [9], the observed value of $W/\sqrt{V}$ is 1.624, which yields a one-tailed P-value (significance level) of 0.052.

The second approach is due to Mantel [4] and Haenszel. We now order the pooled $n_A=61$ observed failure times in the first sample and the $n_B=49$ failure times in the second and denote them $t_1, \cdots, t_{110}$. Denote the total number of observations definitely greater than or equal to t_r by δ_r and the number of failures at t_r by d_r (generally one, except for ties). Let δ_r' be the number of cases in group A with failure or success times greater than or equal to t_r, and d_r' be the number of failures in group A at t_r. Then, conditional on the statistics t_r, δ_r, d_r and δ_r', the distribution of d_r' under H_0 is hypergeometric with $E\,(d_r')=\delta_r'\,(d_r/\delta_r)$ and

$$\text{Var}\,(d_r') = \delta_r'(\delta_r-\delta_r')\,d_r(\delta_r-d_r)/\delta_r^2(\delta_r-1).$$

A significance test for H_0 can be based on $\sum_{r=1}^{110} d_r'$, which is asymptotically normally distributed. Thus we can compute the standardized normal deviate $z=\sum\,[d_r'-E\,(d_r')]/(\sum \text{Var}\,(d_r'))^{1/2}$ where the summations are over $1\leqq r\leqq 110$. Carrying out the arithmetic we find the expected number of failures in program A, $\sum E\,(d_r')\simeq 50.9$, under H_0, which is to be compared with the observed number of $\sum d_r'=61$. The standard deviation of $\sum d_r'$ was calculated to be 5.22 and hence $z=1.91$ for a P value of 0.028. It is pleasing that this significance level agrees closely with the value $P\simeq 0.024$ that Stollmack and Harris find using the F-test under the exponential assumption.

Of course, there are some disadvantages in taking a nonparametric rather than a parametric approach. First, a parametric model may allow one to gain more insight into the structure of the underlying phenomena.

Second, it is easier to incorporate a parametric model into some larger model of the general correctional system. Last, if the failures are indeed exponentially distributed, there is loss of power for small sample sizes. However, for large sample sizes one is not losing much power by using the Mantel-Haenszel test since Peto and Peto [7] have shown it to be asymptotically efficient for the general class of Lehmann alternatives.

Finally, as an alternative to the graphical analysis in the appendix of [9], the cumulative hazard plots of Nelson [6] may be easier to construct and to interpret. They are also more generally applicable to other parametric assumptions, e.g., lognormal and Weibull, as well as exponential. These plots could give some more insight into the important question of whether the tendency to return to criminal behavior decreases with time after release from prison.

REFERENCES

1. R. E. Barlow and F. Proschan, "A Note on Tests for Monotone Failure Rate Based on Incomplete Data," *Ann. Math. Statist.* **40,** 595–600 (1969).
2. E. A. Gehan, "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples," *Biometrika* **52,** 203–223 (1965).
3. C. M. Harris, "A Note on Testing for Exponentiality," *Nav. Res. Logist. Quart.* **23,** 169–175 (1976).
4. N. Mantel, "Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration," *Cancer Chemother. Rep.* **50,** 163–170 (1966).
5. N. Mantel, "Ranking Procedures for Arbitrarily Restricted Observation," *Biometrics* **23,** 65–78 (1967).
6. W. Nelson, "Theory and Applications of Hazard Plotting for Censored Failure Data," *Technometrics* **14,** 945–966 (1972).
7. R. Peto and J. Peto, "Asymptotically Efficient Rank Invariant Test Procedures," *J. Roy. Statist. Soc.*, Ser. A, **135,** 185–206 (1972).
8. H. Scheffé, *The Analysis of Variance*, John Wiley & Sons, New York, 1959.
9. S. Stollmack and C. Harris, "Failure Rate Analysis Applied to Recidivism Data," *Opns. Res.* **22,** 1192–1205 (1974).
10. T. R. Thiagarajan and C. M. Harris, "A Ratio-type Goodness-of-Fit Test for 2-Parameter Weibull Distributions," *IEEE Transactions. Reliability,* **R-25,** 340–343 (1976).
11. B. W. Turnbull and L. Weiss, "A Likelihood Ratio Statistic for Testing Goodness-of-Fit with Randomly Censored Data," Technical Report No. 307, Department of Operations Research, Cornell University, Ithaca, N. Y., 1976.

Published in Operations Research on August 01, 1977 as DOI: 10.1287/opre.25.4.706.
This article has not been copyedited or formatted. The final version may differ from this version.