ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 175, 2004

*Agnieszka Rossa**

ANALYSIS OF CENSORED LIFE-TABLES WITH COVARIATES BY MEANS OF LOG-LINEAR MODELS

Abstract. In survival analysis the subject of observation is duration of time until some event called failure event. Often in such studies only partial information on the length of failure time is available what yields the so-called right-censored observations. The main interest in survival analysis is either to estimate the distribution of the true failure time or to identify the relationship between the true failure time and a set of some covariates. Additional troublesome point of theory and application of survival techniques is treatment of grouped observations (life-tables) along with incorporating covariates.

In the paper a new approach is considered which allows to treat the censored life-table with qualitative covariates as a standard contingency table. Such a table can be further analysed by means of log-linear models or other standard multivariate inference techniques.

Key words: survival analysis, censored data, life-tables, log-linear models.

1. INTRODUCTION

The usual representation of the right-censored random sample with covariates takes the form

$$(T_i, \delta_i, \mathbf{X}_i), \quad \mathrm{i} = 1, 2, \ldots, \mathrm{n} \tag{1}$$

where $\delta_i = 1$ if an *i*-th individual actually failed at time T_i, $\delta_i = 0$ if an individual was right-censored at time T_i and X_i is a *p*-dimensional vector of known covariates, for example, sex, age and other characteristics of an individual.

Nearly all the statistical methods for censored survival data are based on the assumption that censoring mechanism is not related to mechanism causing failures. Thus the usual model for censored survival analysis assumes independent random censoring. In this model variables T_i and δ_i can be defined as follows

* PhD, Department of Statikstical Methods, University of Łódź.

$$T_i = \min(Y_i, Z_i), \quad \delta_i = \begin{cases} 1, & \text{if} \quad T_i = Y_i \\ 0, & \text{if} \quad T_i = Z_i \end{cases} \tag{2}$$

where Y_i are independent copies of a positive random variable Y representing true failure time with a cumulative distribution function (cdf) F. Similarly, Z_i are independent copies of a positive random variable Z with a cdf G. It is assumed that variables Y and Z are independent, conditionally on $\mathbf{X}$. Thus the observed variables T_i represent here independent copies of a variable $\min(Y, Z)$ with a cdf H satisfying the equality $H = 1 - (1 - F)(1 - G)$.

A special case of independent censoring occurs in studies where failure time is measured from entry into the study and one observes the true failure times of those individuals who fail by the time of analysis and censored times for those individuals who do not. In such a case all censoring times Z_i are known and the sequence

$$(T_i, Z_i, \delta_i, \mathbf{X}_i), \quad i = 1, 2, ..., n \tag{3}$$

instead of sequence (1) is observed. It is worth noting that in the repeesentation (3) variables δ_i are redundant and therefore can be omitted.

The main interest in survival analysis is either to estimate the distribution of the true failure time Y represented by F or the so-called survival function $\overline{F} = 1 - F$ or to identify the relationship between the true failure time Y and a set of covariates $\mathbf{X}$. Additional troublesome point of theory and application of survival techniques is treatment of grouped observations (life-tables) along with incorporating covariates.

2. LIFE-TABLE ANALYSIS

Standard life-tables techniques are the oldest techniques most extensively used by actuaries, medical statisticians and demographers, starting from the work of J. Graunt in 1662 (cf. D. V. Glass (1950), B. Benjamin (1978)).

The life-table data arise from a partition of the range $[0, T^*]$ of observations into some time intervals $\Omega_k = [t_k, t_{k+1})$, $k = 0, 1, ..., K - 1$ where the endpoints $0 = t_0 < t_1 < ... < t_K < T^*$ are pre-specified. The life-table data can be characterized by defining numbers of individuals alive at the beginning of each time interval and by defining numbers of failures and censored observations in these intervals.

The main purpose is to estimate conditional probabilities of failure in the intervals Ω_k given survival to t_k or to estimate probabilities of survival past t_{k+1} for $k = 0, 1, ..., K-1$ (see E. L. Kaplan and P. Meier (1958), C. L. Chiang (1968)).

D. R. Cox (1972) gave a first systematic study of use of covariates in the analysis of failure time. He proposed a regression model for a hazard function and introduced a vector of unknown regression parameters specifying the effect of covariates on survival. If the covariates are not time-varying then Cox's model can be termed "proportional hazards" because the ratio of hazard functions for any two individuals is independent of time. Sub-sequent papers by J. D. Kalbfleisch and R. L. Prentice (1973), N. Breslow and J. Crowley (1973), N. Breslow (1974, 1975), O. O. Aalen (1978), P. K. Andersen and R. D. Gill (1982) are the substantial contributions to this subject.

T. R. Holford (1976) introduced the proportional hazards model for life-table data. In his model the baseline hazard function was assumed to be constant within each time interval Ω_k, what implies piecewise exponential distributions for failure times.

This approach was further developed by T. R. Holford (1980) and N. Laird and D. Olivier (1981), who discussed application of log-linear analysis techniques to life-tables with categorical covariates. Their key result refers to two important observations. First, log-linear model for cell means of Poisson contingency table data is equivalent to log-linear model for a hazard function in piecewise exponential survival model. Second, the likelihoods for both models are equivalent. Thus, the statistical inference methods based on maximum likelihood for these models are also equivalent.

The broad survey of the development of the survival analysis throughout the twentieth century can be found in T. R. Fleming and D. Y. Lin (2000) or D. Oakes (2001).

3. LOG-LINEAR MODELS FOR LIFE-TABLES WITH CATEGORICAL COVARIATES

Log-linear models provide a flexible and popular tool of treating the multivariate categorical data arranged in a multidimensional contingency table. Some of the more attractive features of this approach are the easy of model specification, flexlibility in treating both dependent and independent variables and the fact that the equivalent maximum likelihood estimates of model parameters may be obtained from different sampling distributions, such as Poisson, multinomial and product multinomial distributions.

As it was pointed out by N. Laird and D. Olivier (1981), log-linear techniques can be easily applied in life- tables analysis to identify the relationship between the survival time and a set of categorical covariates.

T. R. Holford (1976) considered the following representation for the hazard function $h(y;\mathbf{X})$ of failure time Y

$$h(y;\mathbf{X}) = h_k \cdot \exp\{\mathbf{X}^T\beta\} \quad \text{for} \quad y \in \Omega_k, \ k = 0, 1, ..., K-1 \tag{4}$$

where h_k denotes a constant baseline hazard in the time interval Ω_k, $\mathbf{X}$ is a fixed covariate vector and β is a vector of unknown parameters. Non-proportional hazards model can be reformulated from (4) by allowing the baseline hazard h_k, $k = 0, 1, ..., K-1$ to depend also on $\mathbf{X}$.

The representation (4) implies that, conditional on $\mathbf{X}$, hazard function $h(y;\mathbf{X})$ is a stepwise function of time and failure times have piecewise exponential distributions. Log-linear hazard model proposed by Laird and Olivier flows directly from Holford's model and takes the form

$$h(y;\mathbf{X}) = \ln h_k + \mathbf{X}^T\beta \quad \text{for} \quad y \in \Omega_k, \ k = 0, 1, ..., K-1 \tag{5}$$

Let us assume that the vector $\mathbf{X}$ specifies the levels of p categorical covariates and each covariate X_s of $\mathbf{X}$ has I_s levels indexed by i_s, $s = 1, 2, ..., p$. Denote for simplicity by i_0 the index of time intervals, $i_0 = 0, 1, 2, ..., K-1$. Thus, for the given time interval Ω_{i_0} and for the fixed set of covariates at levels $(i_1, i_2, ..., i_p)$ the hazard function $h(y;\mathbf{X})$ given in (4) takes a constant value, which can be denoted by $\theta_{i_0 i_1 ... i_p}$. Then employing the usual log-linear "u-terms" notation, introduced by M. W. Birch (1963), the model (5) can be rewritten in the following form

$$\ln\theta_{i_0 i_1 ... i_p} = u + u_{i_0}^{\{0\}} + u_{i_1}^{\{1\}} + ... + u_{i_1 i_2}^{\{12\}} + ... + u_{i_1 i_2 ... i_p}^{\{12...p\}} \tag{6}$$

where parameters {effects} on the right-hand side of (6) satisfy the following linear constrains

$$u_+^{\{0\}} = u_+^{\{1\}} = ... = u_{+i_2}^{\{12\}} = u_{i_1 +}^{\{12\}} = ... = u_{+i_2 ... i_p}^{\{12...p\}} = ... = u_{i_1 i_2 ... +}^{\{12...p\}} = 0.$$

The non-proportional hazards model can be introduced here by a simple generalization of (6)

$$\ln\theta_{i_0 i_1 ... i_p} = u + u_{i_0}^{\{0\}} + u_{i_1}^{\{1\}} + ... + u_{i_0 i_1}^{\{12\}} + ... + u_{i_1 i_2 ... i_p}^{\{12...p\}} \tag{7}$$

where

$$u_+^{\{0\}} = u_+^{\{1\}} = ... = u_{+i_1}^{\{01\}} = u_{i_0 +}^{\{01\}} = ... = u_{+i_1 ... i_p}^{\{01...p\}} = ... = u_{i_0 i_1 ... +}^{\{01...p\}} = 0.$$

Thus, the problem of estimating survival distributions under the model (6) or (7) reduces to estimating the u-parameters, what can be done by means of slightly modified Iterative Proportional Fitting Routines (see N. Laird and D. Olivier (1981) for details).

The formula of estimating the log-survival function $\ln S(t)$ derived from the piecewise exponential distribution is expressed as follows

$$\ln\hat{S}(t) = -\exp\{\hat{u}^*\}\left[\sum_{j-1}^{k-1}(t_{j+1}-t_j)\exp\{\hat{u}_j^{(0)}\} + (t-t_k)\exp(\hat{u}_k^{(0)}\}\right], \quad t\in\Omega_{k+1} \tag{8}$$

where $\hat{u}^*$ represents here the estimated total covariate effect.

The modified log-linear model for censored life-table data proposed here allows to handle many IPF routines for log-linear models without any modification. The proposed model is closely related to the one given in (7), however we will assume that $\theta_{i_0 i_1 \ldots i_p}$ represent probabilities of failure in time intervals Ω_{i_0} for fixed sets of covariates at levels $(i_1, i_2, \ldots, i_p)$. This approach is based on the extended life-table data and is based on a method called here "the completion method". The approach allows to apply standard inference.

4. EXTENDED LIFE-TABLES WITH RIGHT-CONSORED DATA

For simplicity, let us assume that the covariate vector **X** is not observed. Let $T^*>0$ be a fixed real number such that $H(T^*)<1$. Let $0=t_0<t_1<\ldots<t_K=T^*<\infty$ constitute a partition of $[0,T^*]$ into K sub--intervals of the form $\Omega_k=[t_k, t_{l+1}]$ for $k=0, 1, \ldots, K-1$. Let us also assume that $\Omega_K=[t_K, \infty)$.

Let us assume that individuals enter the follow-up study at random time points. For an i-th individual we observe a pair of random variables (T_i, Z_i), where T_i and Z_i are defined in Section 1. The observation of individuals terminates when for s items ($s\geqslant 2$ is a fixed integer) we obtain $T_{i_j}\geqslant T^*$, $j=1, 2, \ldots, s$. Let N_s denote the total number of individuals observed in the experiment. Thus, N_s is a random variable distributed according to the negative binomial distribution with parameters s and $p=1-H(T^*)$.

We will consider an extended life-table data characterized by the following statistics

$$
\begin{aligned}
&D_k = \sum_{i=1}^{N_s} 1(T_i \in \Omega_k, Z_i \geqslant t_{k+1}), \quad k = 0, 1, ..., K-1 \\
&D_K = 0, \\
&O_k = \sum_{i=1}^{N_s} 0(T_i \in \Omega_k, Z_i \in \Omega_k), \quad k = 0, 1, ..., K \\
&M_k = \sum_{i=1}^{N_s} 0(T_i \geqslant t_k, Z_i \geqslant t_{k+1}), \quad k = 0, 1, ..., K-1 \\
&M_K = 0, \\
&W_k = O_k + M_k, \quad k = 0, 1, ..., K
\end{aligned}
\tag{9}
$$

where $\mathbf{1}(A)$ denotes a characteristic function of a set A. Notice that there is $W_0 O_k + M_0 = N_s$ and $W_K = O_K$.

Statistics defined in (9) constitute *an extended censored life-table* and will be employed in the procedure called "a completion method".

5. COMPLETION METHOD FOR EXTENDED LIFE-TABLES

Let us consider a probability $q_{k|k}$ defined as follows

$$q_{k|k} \equiv P(Y \in \Omega_k | Y \geqslant t_k), \quad k = 0, 1, ..., K-1 \tag{10}$$

This is the probability of failure in Ω_k, conditional on survival past t_k. This probability will be estimated by means of the following statistics

$$\hat{Q}_{k|k} = \frac{D_k}{M_k - 1}, \quad k = 0, 1, ..., K-1 \tag{11}$$

The similar estimator of $q_{k|k}$ was firstly considered by E. L. Kaplan and P. Meier (1958) for the sample with a fixed size. It is usually called the Reduced-Sample Estimator (RSE). Let us define a probability $q_{k|l}$ as follows

$$q_{k|l} \equiv P(Y \in \Omega_k | Y \geqslant t_l), \quad k = 1, 2, ..., K-1, \quad l = 0, 1, ..., k-1 \tag{12}$$

This is the probability of failure in the interval Ω_k conditional on survival past t_l for $0 < l < k$ and can be estimated from the following recurrent formula

$$\hat{Q}_{k|l} = \frac{W_{l+1} - 1}{M_l - 1} \hat{Q}_{k|l+1}, \quad k = 1, 2, ..., K-1, \quad l = 0, 1, ..., k-1 \tag{13}$$

and the estimated number of failures in the interval Ω_l for O_l individuals, who survived past t_l, can be calculated as

$$\tilde{D}_{k,l} = O_l \cdot \hat{Q}_{k|l'}; \quad l = k, k-1, ..., 1, 0, \quad k = 0, 1, ..., K-1 \tag{14}$$

Let

$$\hat{D} = D_k + \sum_{l=0}^{k} \tilde{D}_{k,l}, \quad k = 0, 1, ..., K-1$$

$$\hat{D}_K = N_s - \sum_{k=0}^{K-1} \hat{D}_k \tag{15}$$

The sum $\Sigma_{l=0}^{k} \tilde{D}_{k,l}$ on the right-hand site of (15) can be treated as an estimated number of failures in the interval Ω_k for those items for which $Z_i < t_{k+1}$. Thus, $\hat{D}_k$ is an estimated total numbers of failures in the intervals Ω_k for $k = 0, 1, ..., K$.

The set of estimates $\hat{D}_k$ determines *a completed version* of an extended censored life-table defined by the statistics (9). This completion procedure is esplained in details in an example in Section 6.

Generally, we can consider extended life-tables constructed for a categorical covariate vector **X** fixed at levels $(i_1, i_2, ..., i_p)$ and calculate the estimated numbers of failures similarly as in Tab. 2. Proceeding in such a way for each combination of levels $(i_1, i_2, ..., i_p)$ of **X** we obtain as a result a $p+1$-dimensional contingency table with estimated numbers of failures for each combination $(i_1, i_2, ..., i_p)$ and for each time interval Ω_{i_0} in its body. Such a table can be next analysed by means of standard log-linear techniques mentioned in Section 3.

6. A NUMERICAL EXAMPLE

We will consider a sample of patients who have had received a valve implantation (bioprothesis or mechanical valve) and had to be reopered because of some valve complications. Patients enter the study at random time points. The subject of observation was the length of their life after reoperation (in years). The study was terminated when $s = 8$ patients survived past $T^* = 7$ years. Thus, the length of life after reoperation is a random right-censored variable, for some of the patients were alive by the end of the study. The total number of patients N_s observed in such an experiment is a random variable. Its realization observed here was equal to 50.

Let $t_0 = 0$, $t_1 = 1$, $t_2 = 7$ and $\Omega_0 = [0, t_1)$, $\Omega_1 = [t_1, t_2)$, $\Omega_2 = [t_2, \infty)$. We will consider an extended life-table determined by the statistics D_k, M_k, O_k for $k = 0, 1, 2$ (see Tab. 1).

Table 1

The Extended Life-Table

Time	D_k	O_k	M_k
$\Omega_0 = [0, 1)$	8	9	41
$\Omega_1 = [1, 7)$	2	23	10
$\Omega_2 = [7, \infty)$	0	8	0

From Tab. 1 we can now estimate total numbers of deaths $\hat{D}_k$ in each interval Ω_k by means of formulae given in (15). These estimates constitute "a completed version" of Tab. 1 (see Tab. 2).

Table 2

Completed Version of Tab. 1

Time intervals	$\hat{D}_k$
$\Omega_0 = [0, 1)$	9.8
$\Omega_1 = [1, 7)$	8.0
$\Omega_2 = [7, \infty]$	32.2

Note, that first two estimates in the second column of Tab. 2 represent estimated values of $\hat{D}_k$ for $k = 0, 1$, and the last value is calculated as $N_s - \Sigma_{k=0}^{1} \hat{D}_k$.

7. SOME THEORETICAL RESULTS

Theorem. The estimators $\hat{Q}_{k|k}$ and $\hat{Q}_{k|l}$ defined in (11) and (13) are unbiased estimators of respective conditional probabilities $q_{k|k}$ and $q_{k|l}$.

Proof.

Let us assume the following notation

$$p_k = P(T \geqslant x_k) \quad \text{for} \quad k = 0, 1, ..., K,$$

$$q_k = P(T \geqslant x_k, Z \geqslant x_{k+1}) \quad \text{for} \quad k = 0, 1, ..., K-1.$$

For $0 \leqslant l < k$ and $1 \leqslant k \leqslant K-1$ the estimator $\hat{Q}_{k|l}$ according to (13) equals to

$$\hat{Q}_{k|l} = \frac{W_{l+1}-1}{M_l-1} \cdot \frac{W_{l+2}-1}{M_{l+1}-1} \cdot ... \cdot \frac{W_{k-1}-1}{M_{k-2}-1} \cdot \frac{W_k-1}{M_{k-1}-1} \cdot \frac{D_k}{M_k-1}.$$

For $l = k$ and $0 \leqslant k \leqslant K-1$ we have from (11)

$$\hat{Q}_{k|k} = \frac{D_k}{M_k-1}.$$

The expression $D_k/(M-1)$ can be also written equivalently as $1-(W_{k+1}-1)/(M_k-1)$, thus for $0 \leqslant l < k$ and $1 \leqslant k \leqslant K-1$

$$\hat{Q}_{k|l} = \frac{W_{l+1}-1}{M_l-1} \cdot \frac{W_{l+2}-1}{M_{l+1}-1} \cdot ... \cdot \frac{W_{k-1}-1}{M_{k-2}-1} \cdot \frac{W_k-1}{M_{k-1}-1} \cdot \left(- \frac{W_{k+1}-1}{M_k-1} \right),$$

and for $l = k$ and $0 \leqslant k \leqslant K-1$

$$\hat{Q}_{k|k} = 1 - \frac{W_{k+1}-1}{M_l-1}.$$

Let us denote by $A_{r|l}$ the following expression of the form

$$A_{r|l} = \frac{W_{l+1}-1}{M_l-1} \cdot \frac{W_{l+2}-1}{M_{l+1}-1} \cdot ... \cdot \frac{W_{r-1}-1}{M_{r-2}-1} \cdot \frac{W_r-1}{M_{r-1}-1} \cdot \frac{W_{r+1}-1}{M_r-1} \tag{16}$$

where $r \geqslant l$. Now the estimator $\hat{Q}_{k|l}$ expresses as follows

$$\hat{Q}_{k|l} = \begin{cases} A_{k-1|l} - A_{k|l} & \text{for} \quad 0 \leqslant l < k, \; 1 \leqslant k \leqslant K-1 \\ 1 - A_{k|k} & \text{for} \quad l = k, \; 0 \leqslant k \leqslant K-1 \end{cases} \tag{17}$$

Let us find the expectation of $A_{r|l}$ defined in (16) using the joint distribution of the variables N_s, M_l, W_{l+1}, ..., M_{r-2}, W_{r-1}, M_{r-1}, W_r, M_r, W_{r+1}.

Notice, that the sample size N_s is a random variable with a negative binomial distribution and that given $N_s = n$ the variables M_l, W_{l+1}, ..., M_{r-2}, W_{r-1}, M_{r-1}, W_r, M_r, W_{r+1} have a multinomial distribution. Thus the joint probability distribution function is of the form

$$P(N_s = n, M_l = m_l, W_{l+1} = w_{l+1}, \ldots, M_{r-1} = m_{r-1}, W_r = w_r, M_r = m_r, W_{r+1} = w_{r+1}) =$$

$$= \binom{n-1}{s-1}\binom{n-s}{m_l - s}\binom{m_l - s}{w_{l+1} - s} \cdot \ldots \cdot \binom{m_{r-1} - s}{w_r - s}\binom{w_r - s}{m_r - s}\binom{m_r - s}{w_{r+1} - s}$$

$$(1 - q_l)^{n - m_l}(q_l - p_{l+1})^{m_l - w_{l+1}} \cdot$$

$$\cdot (p_{l+1} - q_{l+1})^{w_{l+1} - m_{l+1}} \cdot \ldots (q_{r-1} - p_r)^{m_{r-1} - w_r}(p_r - q_r)^{w_r - m_r}(q_r - p_{r+1})^{m_r - w_{r+1}}$$

$$(p_{r+1} - p_K)^{w_{r+1} - s} p_K^S,$$

where

$$n = s, \ s+1, \ \ldots,$$

$$m_l = s, \ s+1, \ \ldots, \ n,$$

$$w_{i+1} = s, \ s+1, \ \ldots, \ m_i, \quad i = l, \ l+1, \ \ldots, \ r,$$

$$m_i = s, \ s+1, \ \ldots, \ w_i \quad i = l+1, \ \ldots, \ r.$$

The distribution function of N_s, M_l, W_{l+1}, ..., M_{r-2}, W_{r-1}, M_{r-1}, W_r, M_r, W_{r+1} can be expressed also equivalently as

$$P(N_s = n, M_l = m_l, W_{l+1} = w_{l+1}, \ldots, M_{r-1} = m_{r-1}, W_r = w_r, M_r = m_r, W_{r+1} =$$
$$= w_{r+1}) =$$

$$= \binom{n-1}{m_l - 1}\binom{m_l - 1}{w_{l+1} - 1} \cdot \ldots \cdot \binom{m_{r-1} - s}{w_r - 1}\binom{w_r - 1}{m_r - 1}\binom{m_r - 1}{w_{r+1} - 1}\binom{w_{r+1} - 1}{s - 1}$$

$$(1 - q_l)^{n - m_l} \, q_l^{m_l} \cdot$$

$$\cdot \left(\frac{q_l - p_{l+1}}{q_l}\right)^{m_l - w_{l+1}} \left(\frac{p_{l+1}}{q_l}\right)^{w_{l+1}} \left(\frac{p_{l+1} - q_{l+1}}{p_{l+1}}\right)^{w_{l+1} - m_{l+1}} \left(\frac{q_{l+1}}{p_{l+1}}\right)^{m_{l+1}} \cdot \ldots \cdot$$

$$\cdot\left(\frac{q_{r-1}-p_r}{q_{r-1}}\right)^{m_{r-1}-w_r}\left(\frac{p_r}{q_{r-1}}\right)^{w_r}\cdot$$

$$\cdot\left(\frac{p_r-q_r}{p_r}\right)^{w_r-m_r}\left(\frac{q_r}{p_r}\right)^{m_r}\left(\frac{q_r-p_{r+1}}{q_r}\right)^{m_r-w_{r+1}}\left(\frac{p_{r+1}}{q_r}\right)^{w_{r+1}}\left(\frac{p_{r+1}-p_K}{p_{r+1}}\right)^{w_{r+1}-s}\left(\frac{p_K}{p_{r+1}}\right)^{s},$$

where

$$w_{r+1}=s,\ s+1,\ \ldots,$$

$$m_i=w_{i+1},\ w_{i+1}+1,\ \ldots,\quad i=r,\ r-1,\ \ldots,\ l+1,\ l,$$

$$w_i=m_i,\ m_i+1,\ \ldots,\quad i=r,\ r-1,\ \ldots,\ l+1,\ l,$$

$$n=m_l,\ m_l+1,\ \ldots$$

Thus, the expectation of $A_{r|l}$ is equal to

$$E(A_{r|l})=\sum_{w_{r+1}=s}^{\infty}\sum_{m_r=w_{r+1}}^{\infty}\cdots\sum_{w_{l+1}=m_{l+1}}^{\infty}\sum_{m_l=w_{l+1}}^{\infty}\sum_{n=m_l}^{\infty}\frac{w_{l+1}-1}{m_l-1}\cdot\ldots\cdot\frac{w_r-1}{m_{r-1}-1}\cdot\frac{w_{r+1}-1}{m_r-1}\cdot$$

$$\cdot P(N_s=n,\ldots,W_{r+1}=w_{r+1})=$$

$$\sum_{w_{r+1}=s}^{\infty}\binom{w_{r+1}-1}{s-1}\left(\frac{p_{r+1}-p_K}{p_{r+1}}\right)^{w_{r+1}-s}\left(\frac{p_K}{p_{r+1}}\right)^{s}\sum_{m_r=w_{r+1}}^{\infty}\frac{w_{r+1}-1}{m_r-1}\binom{m_r-1}{w_{r+1}-1}\cdot$$

$$\cdot\left(\frac{q_r-p_{r+1}}{q_r}\right)^{m_r-w_{r+1}}\left(\frac{p_{r+1}}{q_r}\right)^{w_{r+1}}\cdot\ldots\cdot$$

$$\cdot\sum_{w_{l+1}=m_{l+1}}^{\infty}\binom{w_{l+1}-1}{m_{l+1}-1}\left(\frac{p_{l+1}-q_{l+1}}{p_{l+1}}\right)^{w_{l+1}-m_{l+1}}\left(\frac{q_{l+1}}{p_{l+1}}\right)^{m_{l+1}}\sum_{m_l=w_{l+1}}^{\infty}\frac{w_{l+1}-1}{m_l-1}\binom{m_l-1}{w_{l+1}-1}\cdot$$

$$\cdot\left(\frac{q_l-p_{l+1}}{q_l}\right)^{m_l-w_{l+1}}\left(\frac{p_{l+1}}{q_l}\right)^{w_{l+1}}\cdot$$

$$\cdot\sum_{n=m_l}^{\infty}\binom{n-1}{m_l-1}(1-q_l)^{n-m_l}q_l^{m_l}=\frac{p_{r+1}}{q_r}\cdot\frac{p_r}{q_{r-1}}\cdot\ldots\cdot\frac{p_{l+2}}{q_{l-1}}\cdot\frac{p_{l+1}}{q_l}=$$

$$=\frac{P(T\geqslant t_{r+1})}{P(T\geqslant t_r,Z\geqslant t_{r+1})}\cdot\frac{P(T\geqslant t_r)}{P(T\geqslant t_{r-1},Z\geqslant t_r)}\cdot\ldots\cdot\frac{P(T\geqslant t_{l+2})}{P(T\geqslant t_{l+1},Z\geqslant t_{l+2})}\cdot$$

$$\cdot\frac{P(T\geqslant t_{l+1})}{P(T\geqslant t_l,Z\geqslant t_{l+1})}=$$

$$= \frac{P(Y \geqslant t_{r+1}, Z \geqslant t_{r+1})}{P(Y \geqslant t_r, Z \geqslant t_{r+1})} \cdot \frac{P(Y \geqslant t_r, Z \geqslant t_r)}{P(Y \geqslant t_{r-1}, Z \geqslant t_r)} \cdot \ldots \cdot \frac{P(Y \geqslant t_{l+2}, Z \geqslant t_{l+2})}{P(Y \geqslant t_{l+1}, Z \geqslant t_{l+2})} \cdot$$

$$\cdot \frac{P(Y \geqslant t_{l+1}, Z \geqslant t_{l+1})}{P(Y \geqslant t_l, Z \geqslant t_{l+1})} =$$

$$= \frac{P(Y \geqslant t_{r+1})}{P(Y \geqslant t_r)} \cdot \frac{P(Y \geqslant t_r)}{P(Y \geqslant t_{r-1})} \cdot \ldots \cdot \frac{P(Y \geqslant t_{l+2})}{P(Y \geqslant t_{l+1})} \cdot \frac{P(Y \geqslant t_{l+1})}{P(Y \geqslant t_l)} = \frac{P(Y \geqslant t_{r+1})}{P(Y \geqslant t_l)}.$$

From the result just obtained and using the definition (17) we have for $0 \leqslant l < k$, $1 \leqslant k \leqslant K-1$

$$E(\hat{Q}_{k|l}) = E(A_{k-1|l}) - E(A_{k|l}) = \frac{P(Y \geqslant t_k)}{P(Y \geqslant t_l)} - \frac{P(Y \geqslant t_{k+1})}{P(Y \geqslant t_l)} = \frac{P(Y \in \Omega_k)}{P(Y \geqslant t_l)} = q_{k|l},$$

and for $l = k$, $0 \leqslant k \leqslant K-1$

$$E(\hat{Q}_{k|k}) = 1 - E(A_{k|k}) = 1 - \frac{P(Y \geqslant t_{k+1})}{P(Y \geqslant t_k)} = \frac{P(Y \in \Omega_k)}{P(Y \geqslant t_k)} = q_{k|k},$$

what completes the proof.

REFERENCES

Aalen O. O. (1978), *Nonparametric Inference For a Family of Counting Processes*, "The Annals of Statistics", **6**, 701–726.

Andersen P. K., Gill R. D. (1982), *Cox's Regression Model For Counting Processes: A Large Sample Study*, "The Annals of Statistics", **10**, 1100–1120.

Benjamin B. (1978), *John Graunt*, [in:] (eds.) W. H. Kruskal, J. M. Tanur, *International Encyclopedia of Statistics*, Vol. 1, Free Press, Macmillan Publishing Co., New York, 435–437.

Birch M. W. (1963), *Maximum Likelihood in Three-Way Contingency Tables*, J. Royal Stat. Soc., Ser. B, **25**, 220–233.

Breslow N. (1974), *Covariance Analysis of Censored Survival Data*, "Biometrics", **30**, 89–99.

Breslow N. (1975), *Analysis of Survival Data Under the Proportional Hazards Model*, Int. Stat. Rev., **43**, 45–58.

Breslow N., Crowley J. (1973), *Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship the Annals of Statistics*, **2**, 437–453.

Chiang C. L. (1968), *Introdnction to Stochastic Processes in Biostatistics*, John Wiley, New York.

Cox D. R. (1972), *Regression Model And Life-Tables*, J. Royal Stat. Soc., Ser. B, **34**, 187–220, (with discussion).

Fleming T. R., Lin D. Y. (2000), *Survival Analysis in Clinical Trials: Past Developments and Future Directions*, "Biometrics", **56**, 971–983.

Glass D. V. (1950), *Graunt's Life Table*, J. Inst. Actuaries, **76**, 60–64.

Holford T. R. (1976), *Life-Tables with Concomitant Information*, "Biometrics", **32**, 587–597.
Holford T. R. (1980), *The Analysis of Rates and Survivorship Using Log-linear Models*, "Biometrics", **36**, 299–306.
Kalbfleisch J. D., Prentice R. L. (1973), *Marginal Likelihoods Based on Cox's Regression and Life Model*, "Biometrika", **60**, 267–278.
Kaplan E. L., Meier P. (1958), *Nonparametric Estimation From Incomplete Observations*, J. Amer. Stat. Assoc., **53**, 457–481.
Laird, N., Olivier D. (1981), *Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques*, J. Amer. Stat. Assoc., **76**, 231–240.
Oakes D. (2001), *Biometrika Centenary: Survival Analysis*, "Biometrika", **88**, 99–142.

Agnieszka Rossa

ANALIZA TABLIC TRWANIA ŻYCIA DLA DANYCH CENZUROWANYCH Z WYKORZYSTANIEM MODELI LOGARYTMICZNO-LINIOWYCH

W pracy przedstawiono propozycję analizy tablicy trwania życia dla danych prawostronnie cenzurowanych. Przedstawiona metoda pozwala na sprowadzenie takiej tablicy do wielowymiarowej tablicy kontyngencyjnej, którą można analizować standardowymi technikami wielowymiarowego wnioskowania statystycznego, np. za pomocą modeli logarytmiczno-liniowych.