

# score-statistics-formula-v01

*Gail Gong*

*2013 August 30*

## Contents

<b>1. The score statistic</b>	<b>1</b>
<b>2. Regression examples</b>	<b>3</b>
2A Example 1: $z_{nj} = 1$ . . . . .	3
2B Example 2: $z_{nj} = \frac{\exp( r_n - \bar{r} )}{\exp(\max(r_n) - \bar{r})}$ . . . . .	4
2C Example 3: $z_{nj} = (z_{nj1}, \dots, z_{njK})^T$ . . . . .	4

trunk/projects/calibration/07-rmap-v-0.01-06/07-corn-www-rmap/doc/score-statistics-formula-v01.Rmd

Copied from trunk/projects/calibration/04-the-next-rmap-package/03-rmap-v0.01-05-first-try/website-support-material/score-statistics-formulas/score-statistics-formulas-v01.Rmd

## 1. The score statistic

A person's model-assigned risk has the form  $r = r(t^*)$  where

**Equation (1)**

$$r(t^*) = \int_0^{t^*} \lambda_1(u) \exp\left\{-\left(\Lambda_1(u) + \Lambda_2(u)\right)\right\} du \quad (1)$$

and where  $\lambda_1$  and  $\lambda_2$  denote the model-assigned hazard rates for disease and death respectively.

A person's observed data has the form  $(t, \varepsilon_1, \varepsilon_2)$ ;  $t = \min(t_1, t_2, c)$ , where  $t_1$  and  $t_2$  are her latent times to disease and death respectively, and  $c \leq t^*$  is her time of last observation;  $\varepsilon_1 = 1(t = t_1)$  and  $\varepsilon_2 = 1(t = t_2)$ .

Our goal is to test the null hypothesis that the risk model conforms to the survival data of the people in the cohort study. If this hypothesis is rejected, we want to investigate subsets of people for whom the model fits poorly.

We embed the hypothesized risk model in a larger model involving an unknown parameter vector  $\beta$  with the property that  $\beta = 0$  when the null hypothesis holds. We use the likelihood-based efficient score for  $\beta$  to form a test statistic.

Let  $z$  denote a vector of known predictors for the person, for example multinomial subgroups of assigned risks, or simply  $z = 1$ . The expanded hazards are

$$\lambda_j(t; \beta, z = (z_1, z_2)) = \lambda_j(t) e^{\beta_j^T z_j}, \quad j = 1, 2 \quad (2)$$

where  $\beta_j$  and  $z_j$  are of dimension  $K$ .

If times to disease, death, and censoring are independent, the probability of the  $n$ -th person's observed data is proportional to

$$L_n = \prod_{j=1}^2 \left\{ \left( \lambda_{nj}(t_n) e^{\beta_j^T z_{nj}} \right)^{\varepsilon_{nj}} \exp \left( \Lambda_{nj}(t_n) - e^{\beta_j^T z_{nj}} \right) \right\} \quad (3)$$

$$(4)$$

where

**Equation (2)**

$$\Lambda_{nj}(t_n) = \int_0^t \lambda_{nj}(u) du, \quad n = 1, \dots, N, \quad j = 1, 2 \quad (5)$$

I emphasize here that  $\Lambda_{nj}(t_n)$  is evaluated at the **event time**  $t_n$  from the person's observed time  $t_n = \min(t_{n1}, t_{n2}, c_n)$ .

The person's contribution to the efficient score (for  $j = 1, 2$ ) is

$$U_{nj} = \frac{\partial}{\partial \beta_j} \log L_n(\beta) \Big|_{\beta_j=0} = \left\{ \varepsilon_{nj} - e^{\beta_j^T z_{nj}} \Lambda_{nj}(t_n) \right\} z_{nj} \Big|_{\beta_j=0} = \left\{ \varepsilon_{nj} - \Lambda_{nj}(t_n) \right\} z_{nj} \quad (6)$$

And the person's contribution to the variance is

$$V_{nj} = -\frac{\partial}{\partial \beta_j} \left( \frac{\partial}{\partial \beta_j} \right)^T \log L_n(\beta) \Big|_{\beta_j=0} = e^{\beta_j^T z_{nj}} \Lambda_{nj}(t_n) Z_{nj} Z_{nj}^T \Big|_{\beta_j=0} = \Lambda_{nj}(t_n) Z_{nj} Z_{nj}^T \quad (7)$$

The score test statistic  $T_j$  is

$$T_j = U_j^T V_j^{-1} U_j \quad (8)$$

$$U_j = \sum_{n=1}^N U_{nj} \quad (9)$$

$$V_j = \sum_{n=1}^N V_{nj} \quad (10)$$

$T_1 = T_{\text{outcome}}$  and  $T_2 = T_{\text{mortality}}$  tests the model inadequacy of the outcome hazard and mortality hazard respectively. We can also obtain  $T_+ = T_{\text{combined}}$  that tests the model inadequacy of either hazard by setting  $z = z_1 = z_2$  and  $\beta = \beta_1 = \beta_2$  above.

The person's contribution to the efficient score is

$$U_{n+} = \left. \frac{\partial}{\partial \beta} \log L_n(\beta) \right|_{\beta=0} = \left\{ \varepsilon_{n+} - e^{\beta^T z_n} \Lambda_{n+}(t_n) \right\} z_n \Big|_{\beta=0} = \left\{ \varepsilon_{n+} - \Lambda_{n+}(t_n) \right\} z_n, \quad (11)$$

the person's contribution to the variance is

$$V_{n+} = - \left. \frac{\partial}{\partial \beta} \left( \frac{\partial}{\partial \beta} \right)^T \log L_n(\beta) \right|_{\beta=0} = e^{\beta^T z_n} \Lambda_{n+}(t_n) Z_n Z_n^T \Big|_{\beta=0} = \Lambda_{n+}(t_n) Z_n Z_n^T \quad (12)$$

and the test statistic is

$$T_+ = U_+^T V_+^{-1} U_+ \quad (13)$$

$$U_+ = \sum_{n=1}^N U_{n+} \quad (14)$$

$$V_+ = \sum_{n=1}^N V_{n+} \quad (15)$$

## 2. Regression examples

### 2A Example 1: $z_{nj} = 1$

We have  $K = 1$  and

$$U_j = \sum_{n=1}^N \left\{ \varepsilon_{nj} - \Lambda_{nj}(t_n) \right\} \quad (16)$$

$$V_j = \sum_{n=1}^N \Lambda_{nj}(t_n) \quad (17)$$

$$T_j = U_j^2 / V_j \quad (18)$$

For the covariate  $z = z_1 = z_2$ , we obtain the above equations but with  $+$  replacing  $j$ .

**2B Example 2:**  $z_{nj} = \frac{\exp(|r_n - \bar{r}|)}{\exp(\max(r_n) - \bar{r})}$

Again  $K = 1$  and

$$\bar{r} = \frac{1}{N} \sum_{n=1}^N r_n \quad (19)$$

$$U_j = \sum_{n=1}^N \left\{ \varepsilon_{nj} - \Lambda_{nj}(t_n) \right\} z_n \quad (20)$$

$$V_j = \sum_{n=1}^N \Lambda_{nj}(t_n) z_{nj}^2 \quad (21)$$

$$T_j = U_j^2 / V_j \quad (22)$$

I emphasize here that  $r_n = \int_0^{t^*} \lambda_{n1}(u) S_{n1}(u) S_{n2}(u) du$ , that is the right hand limit of integration is  $t^*$ . For the covariate  $z = z_1 = z_2$ , we obtain the above equations but with  $+$  replacing  $j$ .

**2C Example 3:**  $z_{nj} = (z_{nj1}, \dots, z_{njK})^T$

$z_{nj}$  is a multinomial vector of indicators for membership in one of the  $K$  subgroups in a partition of the sample.

For simplicity, suppose I have two subgroups (or two quantiles).

$$z_{nj} = \begin{cases} \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \text{if } n \text{ is in the first subgroup} \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \text{if } n \text{ is in the second subgroup} \end{cases} \quad (23)$$

$$z_{nj} z_{nj}^T = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \text{if } n \text{ is in the first subgroup} \\ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} & \text{if } n \text{ is in the second subgroup} \end{cases} \quad (24)$$

And then I can write

$$U_j = \sum_{n=1}^N \left\{ \varepsilon_{nj} - \Lambda_{nj}(t_n) \right\} z_{nj} = \begin{pmatrix} \sum_{n \in \text{Subgroup 1}} \left\{ \varepsilon_{nj} - \Lambda_{nj}(t_n) \right\} \\ \sum_{n \in \text{Subgroup 2}} \left\{ \varepsilon_{nj} - \Lambda_{nj}(t_n) \right\} \end{pmatrix} \quad (25)$$

$$V_j = \sum_{n=1} \Lambda_{nj}(t_n) z_{nj} z_{nj}^T = \begin{pmatrix} \sum_{n \in \text{Subgroup } 1} \Lambda_{nj}(t_n) & 0 \\ 0 & \sum_{n \in \text{Subgroup } 2} \Lambda_{nj}(t_n) \end{pmatrix} \quad (26)$$

And now I can write the score statistic

$$T_j = U_j^T V_j^{-1} U_j = \sum_k \frac{\left( \sum_{n \in \text{Subgroup } k} \left\{ \varepsilon_{nj} - \Lambda_{nj}(t_n) \right\} \right)^2}{\sum_{n \in \text{Subgroup } k} \Lambda_{nj}(t_n)} \quad (27)$$

The  $k$ -th subgroup-specific residual is

$$\delta_{jk} = \frac{\sum_{n \in \text{Subgroup } k} \left\{ \varepsilon_{nj} - \Lambda_{nj}(t_n) \right\}}{\sqrt{\sum_{n \in \text{Subgroup } k} \Lambda_{nj}(t_n)}} \quad (28)$$

For the covariate  $z = z_1 = z_2$ , we obtain the above equations but with  $+$  replacing  $j$ .