非寿险精算大作业二答案

1.

$$N_i \sim \text{Poi}(\lambda_i v_i) \quad \text{with} \quad \lambda_i = \lambda(\boldsymbol{x}_i) = \exp\langle \beta, \boldsymbol{x}_i \rangle \tag{1}$$

其中,

$$\begin{aligned}\ln \lambda_i =& \beta_0 + \beta_1 \mathbb{1}_{\text{AGE}_i=\text{B}} + \beta_2 \mathbb{1}_{\text{AGE}_i=\text{C}} + \beta_3 \mathbb{1}_{\text{AGE}_i=\text{D}} + \beta_4 \mathbb{1}_{\text{AGE}_i=\text{E}} + \\ & \beta_5 \mathbb{1}_{\text{AGE}_i=\text{F}} + \beta_6 \mathbb{1}_{\text{USE}_i=\text{B}} + \beta_7 \mathbb{1}_{\text{USE}_i=\text{C}} + \beta_8 \mathbb{1}_{\text{USE}_i=\text{D}}\end{aligned} \tag{2}$$

基本假设:

(a) $N_i, i = 1, \ldots, n$ 互相独立.

(b) N_i 的方差等于期望.

(c) N_i 服从泊松分布.

(d) 连接函数为对数.

2. 参数的极大似然估计为:

$$\begin{aligned}\ln \hat{\lambda}_i =& -0.46 - 0.24 \mathbb{1}_{\text{AGE}_i=\text{B}} - 0.43 \mathbb{1}_{\text{AGE}_i=\text{C}} - 0.38 \mathbb{1}_{\text{AGE}_i=\text{D}} - 0.47 \mathbb{1}_{\text{AGE}_i=\text{E}} + \\ & 0.37 \mathbb{1}_{\text{AGE}_i=\text{F}} - 0.17 \mathbb{1}_{\text{USE}_i=\text{B}} - 0.13 \mathbb{1}_{\text{USE}_i=\text{C}} - 0.40 \mathbb{1}_{\text{USE}_i=\text{D}}\end{aligned} \tag{3}$$

离散系数估计为 $\hat{\phi}_P = 1.09$, 满足泊松分布假设, 方差等于期望.

3. 在 H_0 下, 把年龄组 E 和年龄组 F 合并. 求得检验统计量为

$$D^*(\hat{\beta}, \hat{\beta}_{H_0}) = D^*(\hat{\beta}_{\text{full}}, \hat{\beta}_{H_0}) - D^*(\hat{\beta}_{\text{full}}, \hat{\beta}) = 5.07 > \chi_1^2(0.95) = 3.84$$

拒绝假设, 年龄组 E 和年龄组 F 的索赔频率有差异.

4. 该假设等价于 $H_0 : \beta_6 = 0$. 可通过正态检验得到 p 值为 0.0035, 所以拒绝假设, 车龄组 A 和车龄组 B 的索赔频率有差异. 也可以求得卡方检验统计量为 8.39, 同样拒绝假设.

5. (a) 预测平均索赔次数为 0.50, 预测均方误差为 0.0056. (b) 预测索赔次数为 0.50, 预测均方误差为 0.4968.

6. 使用个体保单数据的对数似然函数为

$$\sum_{i=1}^{n} N_i \ln \lambda_i + N_i \ln v_i - v_i\lambda_i - \ln N_i! \tag{4}$$

对数似然函数中只有下式对参数估计有影响

$$\sum_{i=1}^{n} N_i \ln \lambda_i - v_i\lambda_i \tag{5}$$

通过表 1 易知, 所有的保单被分在 24 组, 在同一组保单的索赔频率相同, 记为 $\lambda_j^+, j = 1, \ldots, 24$. 上式可以变为

$$\sum_{j=1}^{24} N_j^+ \ln \lambda_j^+ - v_j^+\lambda_j^+ \tag{6}$$

其中 $N_j^+ = \sum_{i=1}^n N_i \mathbb{1}_{第i个保单在第j组}$ 为在第 j 组保单的总索赔次数, $v_j^+ = \sum_{i=1}^n v_i \mathbb{1}_{第i个保单在第j组}$ 为在第 j 组保单的总已赚风险单位数. 而式(6)也为累积数据的对数似然函数中影响参数估计的部分. 故建立在个体保单数据上的索赔次数模型等价于建立在累积数据上的索赔次数模型.

7. 参数的极大似然估计和第 2 问完全一样. 离散系数估计为 $\hat{\phi}_p = 0.8773$.

8. 由于在年龄组 A 车龄组 D 的风险单位数为 0, 故有 $4\times6-1 = 23$ 个参数. 参数的极大似然估计为该组合的经验平均索赔次数 $\hat{\lambda}_j^+ = N_j^+/v_j^+$. 该模型为饱和模型.