

Survival Analysis

2024-10-12

1 Current Life Table

现时寿命表是加上同时出生的一代人，按照某年某地实际人口年龄别死亡率计算出的年龄别死亡率计算出的年龄别死亡率陆续死亡，直到全部人口死完为止，用寿命表的方法计算出这一代人在不同年龄组的“死亡人数”和刚满某年龄时的“尚存人数”以及“预期寿命”等寿命表指标。

每一组是左闭右开的区间 $[x, x + n)$ 。考虑从横断面数据获得以下指标：

1. x : exact age.
2. ${}_n P_x$: x 岁组的平均人口数, x 表示年龄组下限, n 表示组距。
3. ${}_n D_x$: x 岁组实际死亡人数。
4. ${}_n m_x$: $\frac{{}_n D_x}{{}_n P_x}$.

假定同时出生的一代人为 10 万人，那么其初期人数 $l_0 = 100000$ 。则有：

1. 0 岁组年龄别死亡概率为 $q_0 = \frac{{}_0 D_0}{l_0} = \frac{{}_0 D_0}{100000}$ 。中间组年龄别死亡概率 ${}_n q_x = \frac{{}_n D_x}{{}_n P_x}$ 。最后一组死亡概率为 1。
2. person-year of survival: $L_0 = l_1 + a_0 \cdot d_0$; 中间组 ${}_n L_x = \frac{l_x + l_{x+n}}{2} \cdot n$; 最后一组 $L_\omega = \frac{l_\omega}{m_\omega}$

由初期的 10 万人，以及每个组的年龄别死亡概率可计算得到每个组的理论死亡人数，进而得到每个组的初期人数 $l_{x+n} = l_x - d_x$ ；由此，可计算得到每个组的人年数，从最后一组开始累加即得到每个组生存人总年数 $T_x = \sum {}_n L_x$ ，进而求得预期寿命 $e_x = \frac{T_x}{l_x}$ 。

去死因寿命表与一般寿命表的差异在于其根据去死因生存概率计算每组的期初人数。

1. ${}_n r_x^{-i} = \frac{{}_n D_x - {}_n D_x^{-i}}{{}_n D_x}$.

2. ${}_n P_x^{-i} = ({}_n P_x)^{n r_x^{-i}}$.
3. $l_{x+n}^{-i} = l_x^{-i} \cdot {}_n p_x^{-i}$.
4. ${}_n d_x^{-i} = l_x^{-i} - l_{n+x}^{-i}$

2 Survival Analysis

直接观察满 n 年的人数中活满 n 年的比率。其计算公式为：

$$n = \frac{n}{n} \cdot 100\%$$

考虑 x, L_x, W_x, D_x 分别为随访年数, 期初观察人数, 期内失访人数, 期内死亡人数, 那么有：

1. $L_{x+1} = L_x - W_x - D_x$.
2. $N_x = L_x - \frac{W_x}{2}$. 校正人数, 该公式假设每个失访者在这个区间内被随访的时间是区间长度的一半, 那么 N_x 相当于是实际随访的人年数。
3. $q_x = \frac{D_x}{N_x}$. 死亡概率, 表示活满 x 年的人在第二年死亡的概率。
4. $p_x = 1 - q_x$. 生存概率, 表示活满 x 年的人再活一年的概率。

那么, 其 n 年生存率：

$${}_n P_0 = \prod_{x=0}^{n-1} p_x$$

$$S_{n P_0} = {}_n P_0 \cdot \sqrt{\sum_{i=0}^{n-1} \frac{q_i}{p_i \cdot N_i}}$$