
生存分析中生存率的计算

作者：陶立元，赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/1741.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

近日连续遇到好几

个同事在写临床研究论文时问及生存率的问题，有人还是对生存率

的计算有点儿理不清楚，本期就这一问题作一讨论。当然这一讨论仅限于简单的临床研究数据中的Kaplan-Meier法

。对于流行病学数据的生存率估计，有更为复杂的估算方法，也需要考虑更多的影响因素。

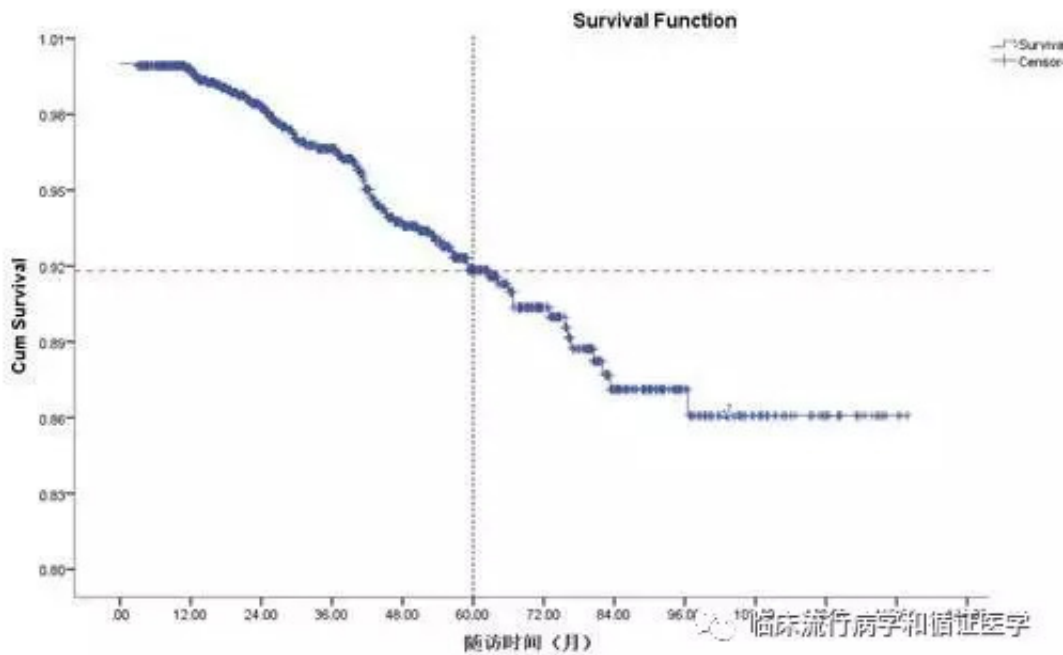
假设我们有某肿瘤患者的以下数据，包括患者的基本信息、诊断时间、治疗手段、随访时间和存活与否等指标。简单理解的生存率，就是患者的存活率，是指在这些患者中经过一段时间的随访后，尚存活的患者的比例。它的计算方法=随访满一定时间的尚存活的病例数/开始随访时的总病例数×100%。如2年生存率的计算，是指随访满2年的尚存活的病例/总病例数。上面的计算的确有着严谨的逻辑，但这种生存率的计算在临床研究中往往是不被采用的。因为这种情况下生存率的计算，会漏掉不少随访时间不够2年的病例。随访时间不够2年的原因可能是：1)、这些病例从入院到现在还没有2年的时间，2)、这些病例随访不足2年时就失访了。

就像上面的这个数据，我截屏了前20个病人，其中13人都是不能纳入上面说的那个2年生存率的计算的，因为他们的随访时候都不够24个月。

但是我们在做生存分析时往往不是这样计算的生存率的，我们是按照KM的生存概率表读取的某一个时间的生存率，比如下表中5年的累积生存率为91.8%。理论上应该读取第838行的累积生存率(因为它正好对应着60个月)，但是这儿没有数，所以读取他的上一个数(第826行)来代替。

Survival Table						
	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	2.633	死亡	.999	.001	1	1206
2	3.000	删失	-	-	1	1205
3	3.000	删失	-	-	1	1204
4	3.267	删失	-	-	1	1203
5	3.367	删失	-	-	1	1202
6	3.533	删失	-	-	1	1201
7	3.567	删失	-	-	1	1200
8	3.567	删失	-	-	1	1199
...
823	59.000	删失	-	-	57	384
824	59.033	死亡	.921	.010	58	383
825	59.100	删失	-	-	58	382
826	59.133	死亡	.918	.011	59	381
827	59.133	删失	-	-	59	380
828	59.233	删失	-	-	59	379
829	59.267	删失	-	-	59	378
830	59.400	删失	-	-	59	377
831	59.500	删失	-	-	59	376
832	59.500	删失	-	-	59	375
833	59.500	删失	-	-	59	374
834	59.700	删失	-	-	59	373
835	59.733	删失	-	-	59	372
836	59.733	删失	-	-	59	371
837	59.767	删失	-	-	59	370
838	60.067	删失	-	-	59	369
839	60.067	删失	-	-	59	368
840	60.100	删失	-	-	59	367
841	60.167	删失	-	-	59	366
842	60.200	删失	-	-	59	365
843	60.267	删失	-	-	59	364
844	60.333	删失	-	-	59	363
845	60.333	删失	-	-	59	362
846	60.333	删失	-	-	59	361
847	60.333	删失	-	-	59	360
848	60.333	删失	-	-	59	359
849	60.333	删失	-	-	59	358
850	60.333	删失	-	-	59	357
851	60.333	删失	-	-	59	356
852	60.333	删失	-	-	59	355
853	60.333	删失	-	-	59	354
854	60.333	删失	-	-	59	353
855	60.333	删失	-	-	59	352
856	60.333	删失	-	-	59	351
857	60.333	删失	-	-	59	350
858	60.333	删失	-	-	59	349
859	60.333	删失	-	-	59	348
860	60.333	删失	-	-	59	347
861	60.333	删失	-	-	59	346
862	60.333	删失	-	-	59	345
863	60.333	删失	-	-	59	344
864	60.333	删失	-	-	59	343
865	60.333	删失	-	-	59	342
866	60.333	删失	-	-	59	341
867	60.333	删失	-	-	59	340
868	60.333	删失	-	-	59	339
869	60.333	删失	-	-	59	338
870	60.333	死亡	.916	.011	60	337
871	60.333	删失	-	-	60	336
872	60.333	删失	-	-	60	335
873	60.333	删失	-	-	60	334
874	60.333	删失	-	-	60	333

这个5年的累积生存率也跟生存曲线对应着，它的生存曲线如下。图中红色的线对应的就是60个月的累积生存率。



这儿我们一直把它称为累积生存率，因为它有别于前面所说的生存率，累积生存率的计算是将各个时点的生存概率进行乘积计算而来的，考虑了删失数据的影响。但是在生存分析的论文写作中大家早已习惯了将累积生存率简称为生存率。如果不存在删失数据，那么累积生存率就完全等于咱们前面所说的那个简单生存率的计算，即等于t时刻仍存活的例数/观察总例数。

由于上述的表格中不存在时间正好等于60.000个月时的生存率，所以我们选用它上一个最接近的时间点59.133个月的数据(0.918)来代替。除此之外，我们还可以采用比例的方法来估算60.000个月的生存率，因为我们知道59.133个月是0.918，62.533个月是0.916，假设这个生存率在这一较小的时间段内是等比例下降的。那么可以算出这一时间段内每1个月的平均下降速度为 $(0.918-0.916)/(62.533-59.133)=0.000588$ ，60个月距离59.133个月为0.867个月，乘以 $0.000588=0.00051$ ，再用59.133个月的 $0.918-0.00051=0.91749$ ，即为正好60个月时的估算生存率。

呃，这么一通计算折腾下来，跟选取最接近的数据来代替，两个生存率也只差了0.51%。本例题的样本量比较大，事件与事件之间时间点相距较近。但有时候研究对象较少时(如只有20例)，出现事件的时间点可能会相距较远，此时差别会大一些。当然读者应该酌情考虑是否采用这一计算方法。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发