

의학통계학 과제 6

이다연

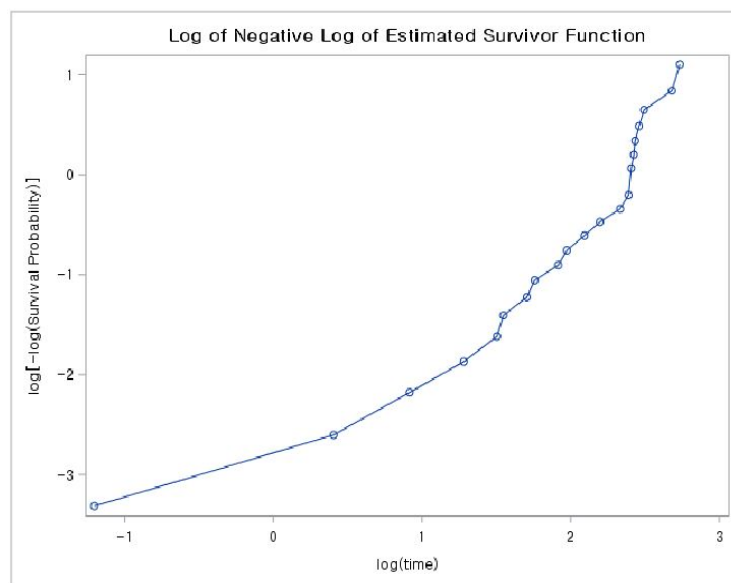
2018. 12. 17

연습문제 5번

- (1) 생존시간이 와이블분포를 따르는지에 대한 적합성 검토를 그래프를 이용하여 살펴보아라.
- (2) (1)에서 구해진 분포로 생존시간을 분석하고 각 회귀계수의 유의성을 검정하여라.

5.1

와이블분포의 경우, 시간 t 에 \log 를 씌운 $\log t$ 에 대한 $\log(-\log \hat{S}(t))$ 의 그래프가 직선이 되는가를 검토하면 된다. SAS의 PROC LIFETEST에서 option으로 PLOT=(LLS)를 쓰면 이 그래프를 얻을 수 있다.



위 그래프는 당뇨병환자의 자료에서 $\log t$ 에 대한 $\log(-\log \hat{S}(t))$ 를 그린 것이다. 직선 형태를 보이므로 와이블모형을 가정한 것이 타당하다는 것을 알 수 있다.

5.2

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
age	1	8.5013	0.0035
bmi	1	0.1018	0.7497
diag_age	1	4.6127	0.0317
smoke	1	2.9031	0.0884

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	3.2683	0.6987	1.8989	4.6377	21.88	<.0001
age		1	-0.0410	0.0141	-0.0686	-0.0135	8.50	0.0035
bmi		1	-0.0044	0.0138	-0.0315	0.0227	0.10	0.7497
diag_age		1	0.0282	0.0131	0.0025	0.0539	4.61	0.0317
smoke	0	1	0.3697	0.2170	-0.0556	0.7949	2.90	0.0884
smoke	1	0	0.0000
Scale		1	0.4185	0.0729	0.2975	0.5887		
Weibull Shape		1	2.3893	0.4159	1.6986	3.3609		

위 결과를 참고하여 생존시간의 모형식을 표현하면,

$$\log T = 3.2683 - 0.0410 * x_{age} - 0.0044 * x_{bmi} + 0.0282 * x_{diag_age} + 0.3697 * x_{smoke(0)} + 0.4185e.$$

와 같다. 여기서 x_{smoke} 는 흡연상태를 나타내며 흡연경력이 없을 때 0, 있을 때 1로 코딩되었다. 환자의 나이(x_{age})에 대한 회귀계수가 -0.0410으로 나이가 많을수록 환자의 생존시간이 줄어들게 되며 이는 5% 유의수준 하에서 유의하다(p-값=0.0035). 반면 진단시 환자의 나이(x_{diag_age})에 대한 회귀계수는 0.0282로 진단시 환자의 나이가 많을수록 생존시간이 길게 되며 이또한 5% 유의수준 하에서 유의하다(p-값=0.0317). 그러나 환자의 BMI지수(x_{bmi})나 흡연상태(x_{smoke})는 유의한 영향을 미치지 않는다(p-값은 각각 0.7497과 0.0884이다).

연습문제 6번

비례위험 모형을 이용하여 자료를 분석하고 각 설명변수들의 유의성을 검정해보아라.

6.

PROC PHREG에서 비례위험모형에 대한 분석을 할 수 있다. 다음은 급성골수성백혈병(Acute Myeloid Leukemia) 환자들 자료의 결과이다.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5.4260	2	0.0663
Score	5.4212	2	0.0665
Wald	5.0979	2	0.0782

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
age	1	1.01317	0.45740	4.9065	0.0268	2.754
myeloid	1	0.35025	0.43917	0.6360	0.4252	1.419

첫번째 표는 모든 회귀계수가 0이라는 가설에 대한 검정통계량이다. 세 통계량 모두 5% 유의수준 하에서는 가설을 기각하지 못하지만 p-값이 각각 0.0663, 0.0665, 0.0782로 무시할 수 없는 수준이다.

두번째 표, 회귀계수에 대한 출력결과를 보면 환자의 나이에 대한 회귀계수의 추정치가 1.0132로 유의하다(p-값=0.0268). 그러나 골수 응고여부의 p-값은 0.4252로 유의한 영향을 미치지 않는다.

비례위험모형을 사용하였으므로 나이가 1살 증가할 때마다 위험율이 $\exp(1.0132) = 2.75$ 배 증가하는 것을 알 수 있다.