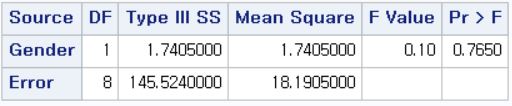
생명과학데이터분석 HW3

2014150137, 통계학과

박 정진

6-3

1. 반복 요인이 아니므로 구형성 검정의 필요성 X

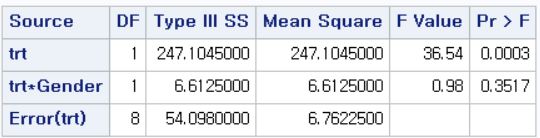
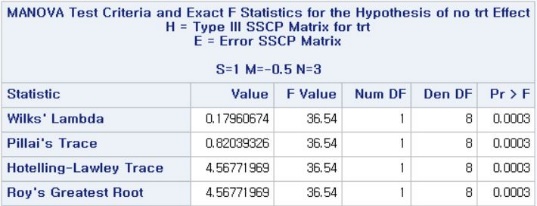


Not

P-value가 0.7650으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 성별과 폐활량 지수의 증가 간에는 유의한 상관관계가 없다.

1. 운동 방법은 수준이 2개뿐으로, 별도의 구형성 자동 만족(일변량 분석)

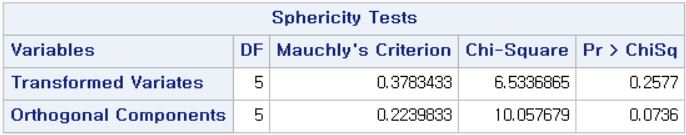


Not

P-value가 0.0003으로 유의수준 =0.05하에서 귀무가설을 기각한다.

즉 운동 방법에 따라 폐활량 지수의 증가에 유의한 차이가 있다.

1. -1 구형성 검정

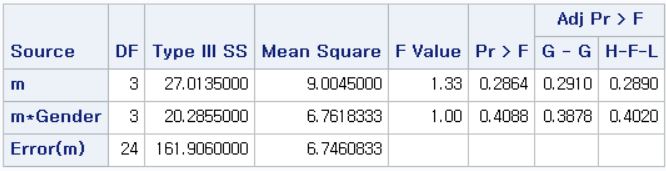


Not

P-value가 0.2577, 0.0736으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 시간은 구형성 가정을 만족한다.

-2 일변량 분석

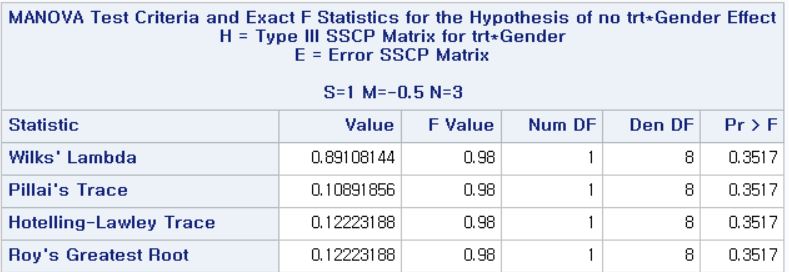


Not

P-value가 0.2864로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 시간에 따라 폐활량 지수가 유의하게 차이가 나지는 않는다.

1. –다변량

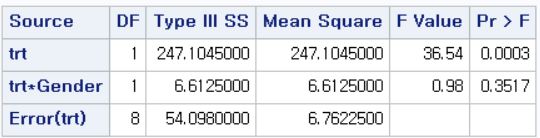


Not

P-value가 0.3517로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 두 요인 간에는 유의한 교호작용이 존재하지 않는다.

-일변량



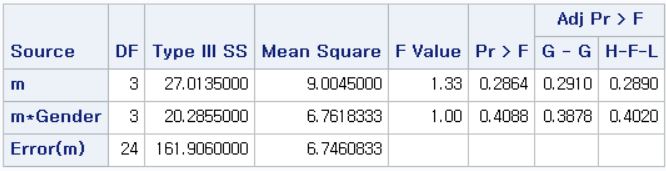
Not

P-value가 0.3517로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 두 요인 간에는 유의한 교호작용이 존재하지 않는다.

두 방법 모두 동일하게 유의한 교호작용이 존재하지 않는 것으로 나타났다.

1. 시간이 구형성 가정을 만족하므로 일변량 분석을 이용하여 교호작용을 본다.

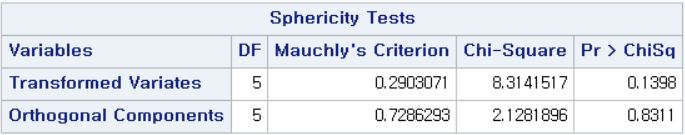


Not

P-value가 0.4088로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 성별과 시간 간에는 유의한 교호작용이 존재하지 않는다.

1. – 구형성 검정

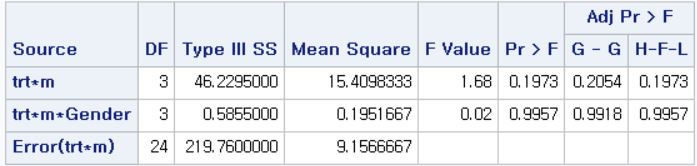


Not

P-value가 0.1398, 0.8311으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 시간과 운동방법의 교호작용은 구형성 가정을 만족한다.

* 일변량 검정



Not

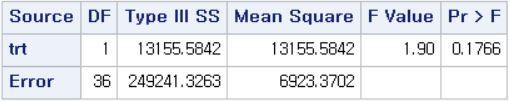
P-value가 0.1973으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 운동 방법과 시간 간에는 유의한 교호작용이 존재하지 않는다.

6-6

1. 약의 종류는 반복 요인이 아니므로 구형성 검정 X

* 일변량 분석

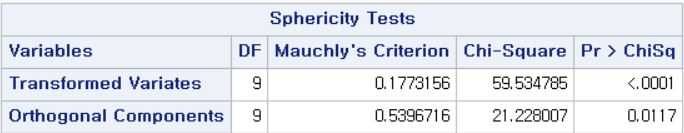


Not

P-value가 0.1766으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉, 약의 종류에 따라 운동거리에 유의한 차이는 없다.

1. – 구형성 검정

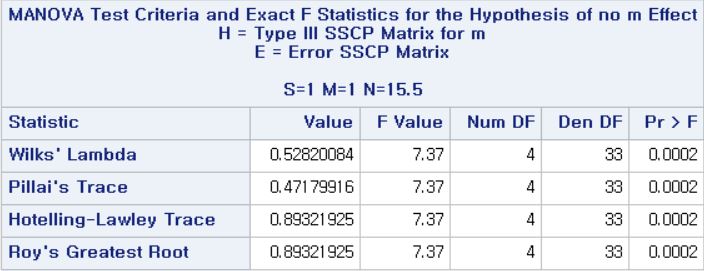


Not

P-value가 0.0001, 0.0117으로 유의수준 =0.05하에서 귀무가설을 기각한다.

즉 시간은 구형성 가정을 만족하지 못한다.

- 다변량 분석

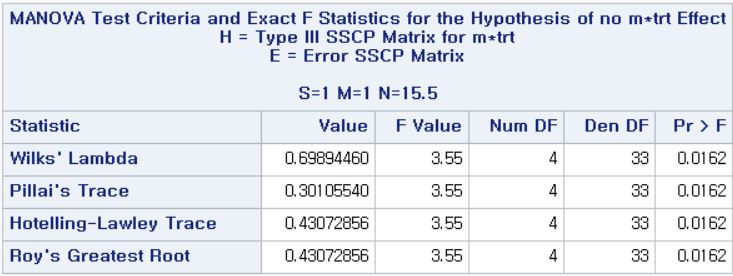


Not

P-value가 0.0002으로 유의수준 =0.05하에서 귀무가설을 기각한다.

즉, 시간에 따라 운동거리에 유의한 차이가 있다.

1. – 다변량

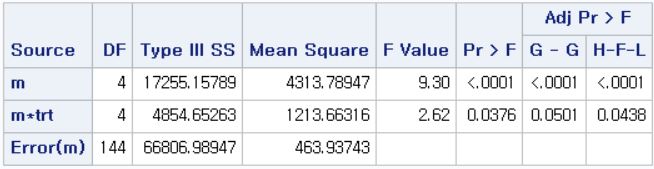


Not

P-value가 0.0162으로 유의수준 =0.05하에서 귀무가설을 기각한다.

즉, 시간에 따라 약의 종류 사이에 유의한 교호작용이 있다.

-일변량



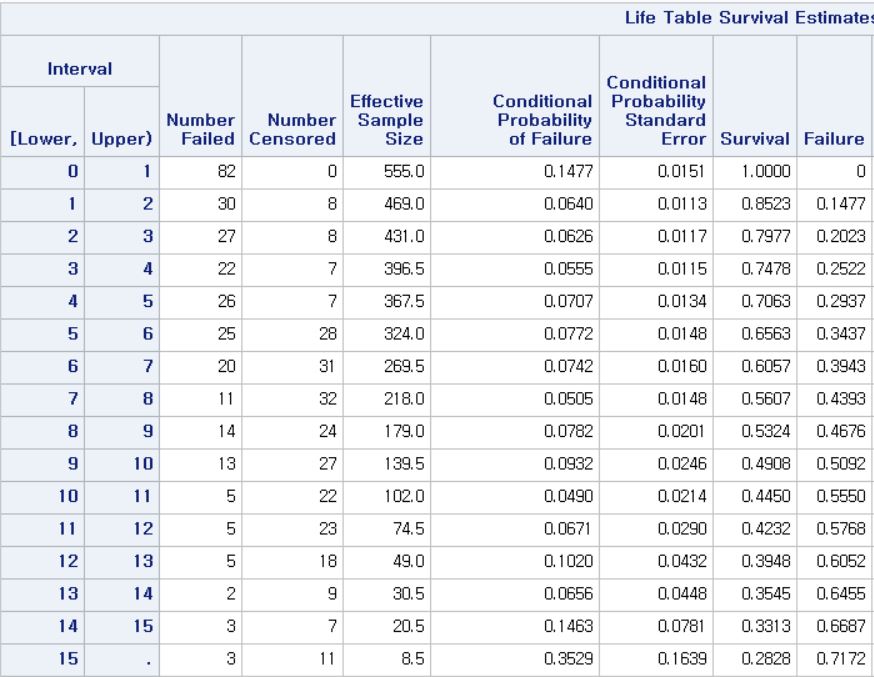
Not

P-value가 0.0376으로 유의수준 =0.05하에서 귀무가설을 기각한다.

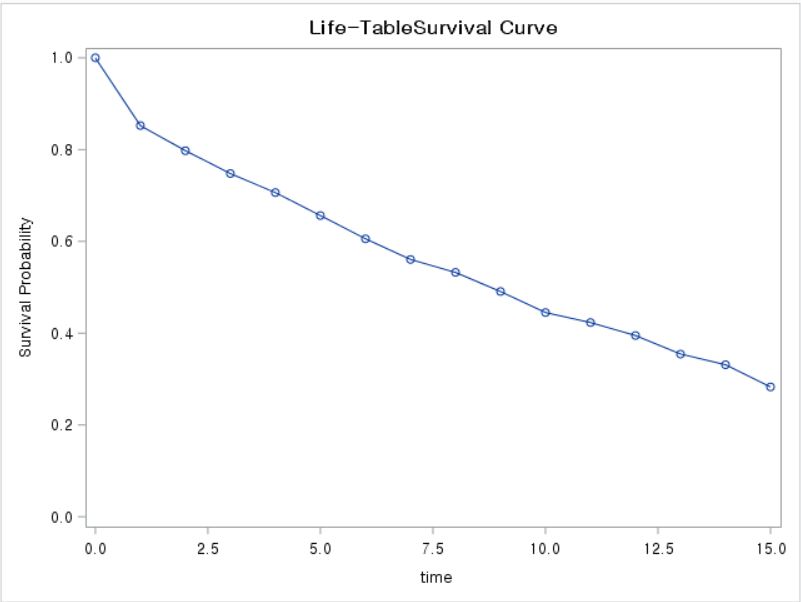
즉, 시간에 따라 약의 종류 사이에 유의한 교호작용이 있다.

두 분석 방법 모두에서 유의한 교호작용이 있음을 알 수 있다.

7-2

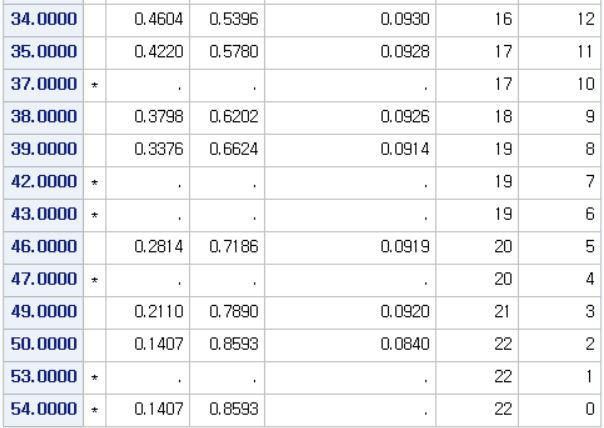


<추정된 각 구간별 유효인원수, 사망률, 생존율, 생존함수>

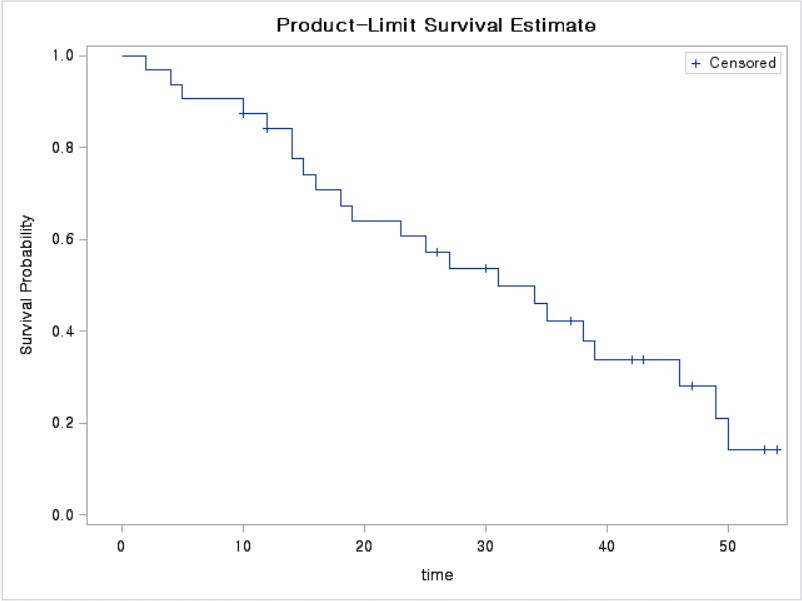


<생존 함수 그래프>(실제로는 계단식 모양이어야 함)

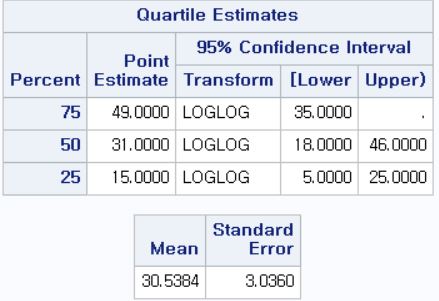
7-3



각 시점에서의 분산은 Survival Standard Error 행을 제곱함으로써 얻을 수 있다.

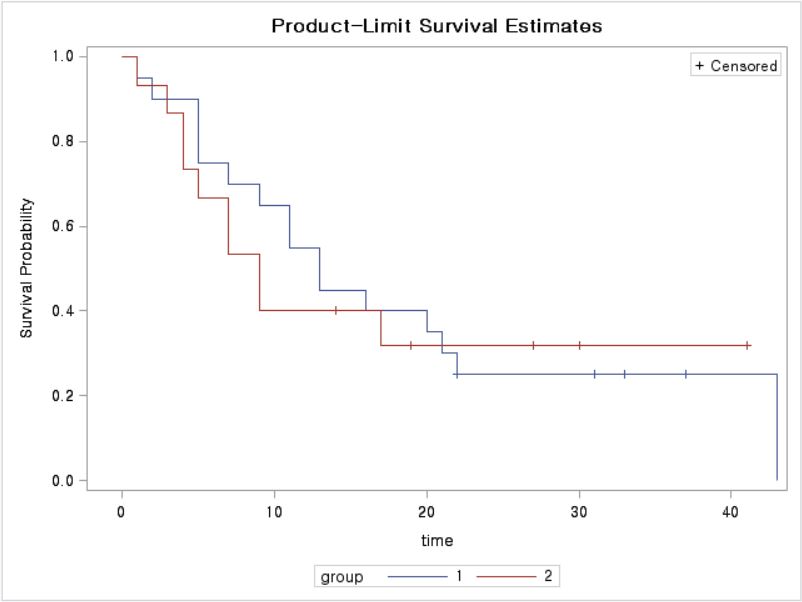


<생존함수의 그래프>



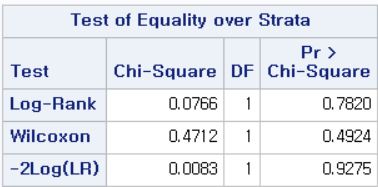
환자의 50%가 생존하는 시간은 31로 추정된다.

7-4



<생존함수 그래프>

생존 함수 그래프가 교차하므로, 로그-순위 검정법을 사용하기 적합하지는 않음을 알 수 있지만, 문제에서 요구한대로 검정을 수행한다.



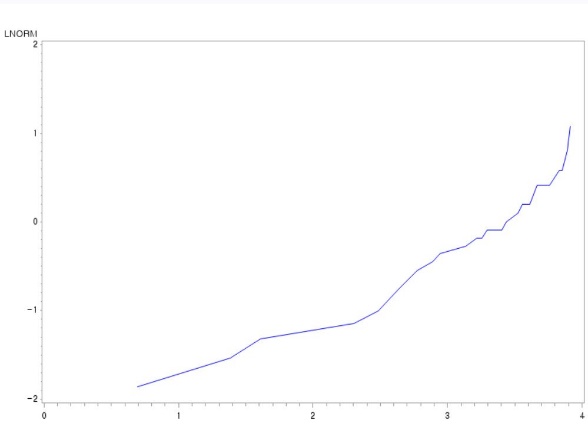
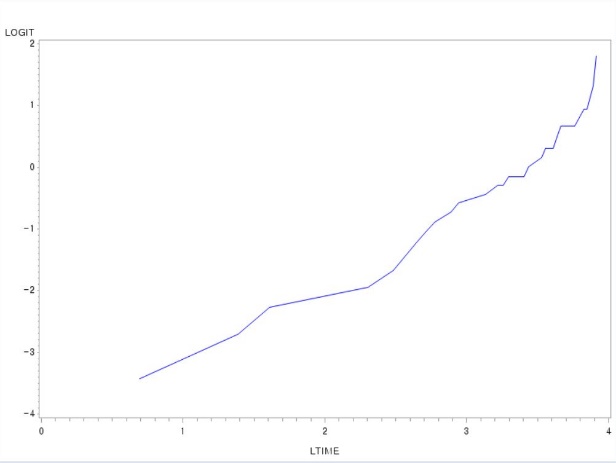
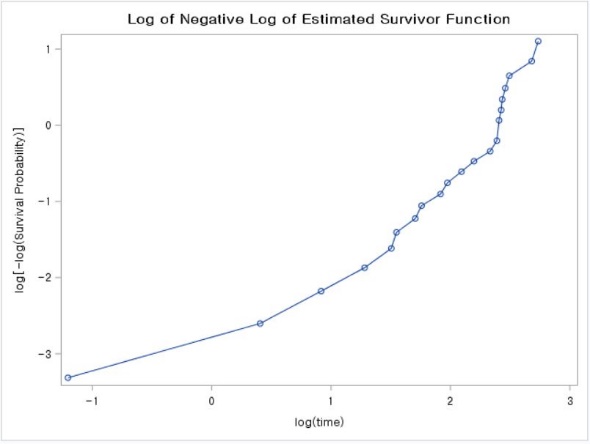
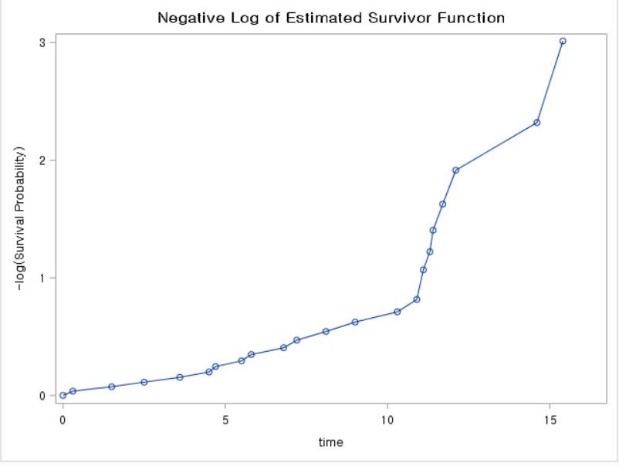
Not

로그 순위 검정 통계량의 P-value가 0.7820으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉, 두 그룹간 생존 함수는 유의미한 차이를 보이지 않는다.

Gehan 검정 통계량의 P-value는 0.4924로 역시 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

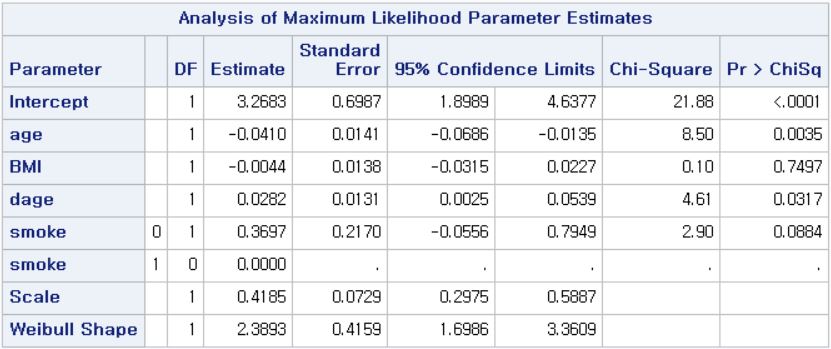
7-5



왼쪽 위부터 순서대로, 지수분포, 와이블 분포, 로그 로지스틱, 로그 정규분포의 그림.

지수분포를 제외하고는 대체로 직선 형태를 띠고 있다.

1. 와이블 분포를 이용하여 생존시간을 분석하였다.



분석하려는 모델은,

이다.

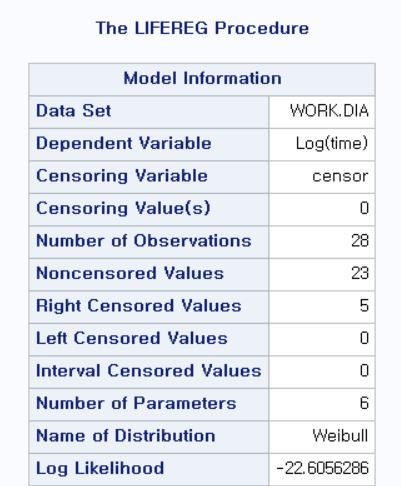
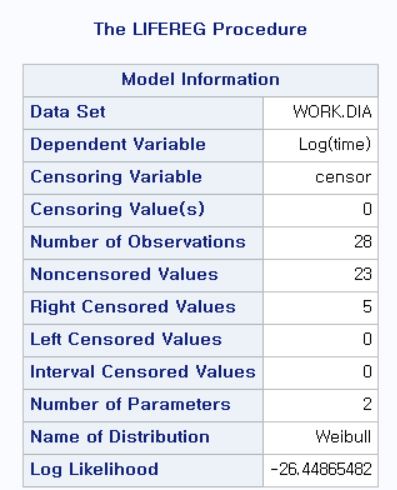
각 변수들에 대하여,

Not

를 유의수준 =0.05하에서 검정한 결과,

BMI와 흡연 상태는 귀무가설을 기각하지 못하여, 유의한 변수가 아닌 것으로 생각되며,

나이와 진단시 나이는 같은 유의수준에서 귀무가설을 기각하여 유의한 변수로 생각된다.



Not

의 우도비 검정 통계량의 경우,

-2(log L – log L0) = -2(-26.4487+22.6056)=7.69 이고 이는, 를 따른다. =9.84 이므로 유의수준 =0.05에서 귀무가설을 기각하지 못한다. 다만 P-value = 0.1로 유의수준과 큰 차이가 나지는 않는다.

7-6



* 나이

Not

카이스퀘어 검정 통계량의 P-value가 0.0268으로 유의수준 =0.05하에서 귀무가설을 기각한다. 즉, 나이에 대한 회귀 계수 1.01317은 유의하다고 할 수 있다.

* 골수 응고여부

Not

카이스퀘어 검정 통계량의 P-value가 0.4252로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다. 즉, 골수 응고여부에 대한 회귀 계수 0.35025는 유의하지 않다고 할 수 있다.

다른 변인이 동일할 때, 나이가 50세 이상인 경우, 50세 미만인 경우보다 위험률이 2.754배 높다.

다른 변인이 동일할 때, 골수 응고가 100%인 경우, 그렇지 않은 경우보다 위험률이 1.419배 높다.

<SAS code>

/\*7-2\*/

**DATA** cancer1;

INPUT time censor freq @@;

CARDS;

0.5 1 82 0.5 0 0 1.5 1 30 1.5 0 8 2.5 1 27 2.5 0 8 3.5 1 22 3.5 0 7

4.5 1 26 4.5 0 7 5.5 1 25 5.5 0 28 6.5 1 20 6.5 0 31

7.5 1 11 7.5 0 32 8.5 1 14 8.5 0 24 9.5 1 13 9.5 0 27 10.5 1 5 10.5 0 22

11.5 1 5 11.5 0 23 12.5 1 5 12.5 0 18 13.5 1 2 13.5 0 9 14.5 1 3 14.5 0 7

15.5 1 3 15.5 0 11

;

**RUN**;

**PROC** **LIFETEST** METHOD=LIFE

INTERVALS = **1** **2** **3** **4** **5** **6** **7** **8** **9** **10** **11** **12** **13** **14** **15**

PLOTS = (S, H)

GRAPHICS;

TIME time\*censor(**0**);

FREQ freq;

**RUN**;

/\*7-3\*/

**DATA** cancer2;

INPUT time censor @@;

CARDS;

2 1 4 1 5 1 10 1 10 0 12 1 12 0 14 1 14 1 15 1 16 1 18 1 19 1 23 1 25 1 26 0 27 1 30 0

31 1 34 1 35 1 37 0 38 1 39 1 42 0 43 0 46 1 47 0 49 1 50 1 53 0 54 0

;

**RUN**;

**PROC** **LIFETEST** METHOD=KM PLOTS=(S) GRAPHICS OUTSURV=a;

TIME time\*censor(**0**);

**PROC** **PRINT** DATA=a;

**RUN**;

/\*7-4\*/

**DATA** Hod;

INPUT time censor group @@;

CARDS;

1 1 1 2 1 1 5 1 1 5 1 1 5 1 1 7 1 1 9 1 1 11 1 1 11 1 1 13 1 1 13 1 1 16 1 1 20 1 1 21 1 1 22 0 1 22 1 1 31 0 1 33 0 1 37 0 1 43 1 1

1 1 2 3 1 2 4 1 2 4 1 2 5 1 2 7 1 2 7 1 2 9 1 2 9 1 2 14 0 2 17 1 2 19 0 2 27 0 2 30 0 2 41 0 2

;

**RUN**;

**proc** **lifetest** data = Hod plots = (s) graphics;

time time\*censor(**0**);

strata group;

symbol1 v = none color=b line=**1**;

symbol2 v = none color=b line=**2**;

**run**;

/\*7-5\*/

**DATA** dia;

INPUT censor time age BMI dage smoke @@;

CARDS;

1 3.6 63 25.1 46 1 1 15.4 71 26.0 59 0 1 11.3 51 32.0 49 1 1 10.3 59 28.1 57 1

1 5.8 50 26.1 49 1 0 8.0 66 45.3 49 0 1 14.6 42 30.0 41 1 1 11.4 40 35.7 36 1

1 7.2 67 28.1 61 0 1 5.5 86 32.9 61 0 1 11.1 52 37.6 46 1 1 16.5 42 43.4 37 0

1 10.9 60 25.4 60 0 1 2.5 75 49.7 57 1 0 10.8 81 35.2 81 0 1 4.7 60 37.3 39 0

0 5.5 60 26.0 42 0 1 4.5 63 21.8 60 1 1 9.0 62 18.2 43 0 1 6.8 57 34.1 41 1

0 3.6 71 25.6 54 1 1 12.1 58 35.1 45 0 1 8.1 42 32.5 28 1 1 11.1 45 44.1 40 0

0 7.0 66 29.7 59 1 1 1.5 61 29.2 54 0 1 11.7 48 25.2 30 1 1 0.3 82 25.3 50 0

;

**RUN**;

**PROC** **LIFEREG** DATA=dia;

CLASS smoke;

MODEL time\*censor(**0**) = age BMI dage smoke/DIST=WEIBULL;

**RUN**;

**PROC** **LIFETEST** DATA=dia PLOTS=(LS, LLS) GRAPHICS;

TIME time\*censor(**0**);

**RUN**;

**PROC** **LIFETEST** DATA=dia OUTSURV = a;

TIME time\*censor(**0**);

**RUN**;

**DATA** b; SET a;

S = SURVIVAL;

LOGIT = LOG((**1**-S)/S);

LNORM = PROBIT(**1**-S);

LTIME=LOG(time);

**PROC** **GPLOT** DATA=b;

SYMBOL1 VALUE=NONE I=JOIN;

PLOT LOGIT\*LTIME LNORM\*LTIME;

**RUN**;

/\*7-6\*/

**DATA** Leukemia;

INPUT time censor age x @@;

CARDS;

18 1 0 0 9 1 0 1 28 0 0 0 31 1 0 1 39 0 0 1 19 0 0 1 45 0 0 1 6 1 0 1 8 1 0 1 15 1 0 1 23 1 0 0 28 0 0 0 7 1 0 1 12 1 1 0 9 1 1 0

8 1 1 0 2 1 1 1 26 0 1 0 10 1 1 1 4 1 1 0 3 1 1 0 4 1 1 0 18 1 1 1 8 1 1 1 3 1 1 1 14 1 1 1 3 1 1 0 13 1 1 1 13 1 1 1 35 0 1 0

;

**RUN**;

**proc** **phreg** data = Leukemia;

model time\*censor(**0**) = age x;

**run**;

**proc** **export** data=cancer1

outfile='C:\Users\pkmon\Desktop\cancer1.csv'

dbms=csv

replace;

**run**;

**proc** **export** data=cancer2

outfile='C:\Users\pkmon\Desktop\cancer2.csv'

dbms=csv

replace;

**run**;

**proc** **export** data=Hod

outfile='C:\Users\pkmon\Desktop\Hod.csv'

dbms=csv

replace;

**run**;

**proc** **export** data=dia

outfile='C:\Users\pkmon\Desktop\dia.csv'

dbms=csv

replace;

**run**;

**proc** **export** data=Leukemia

outfile='C:\Users\pkmon\Desktop\Leukemia.csv'

dbms=csv

replace;

**run**;