



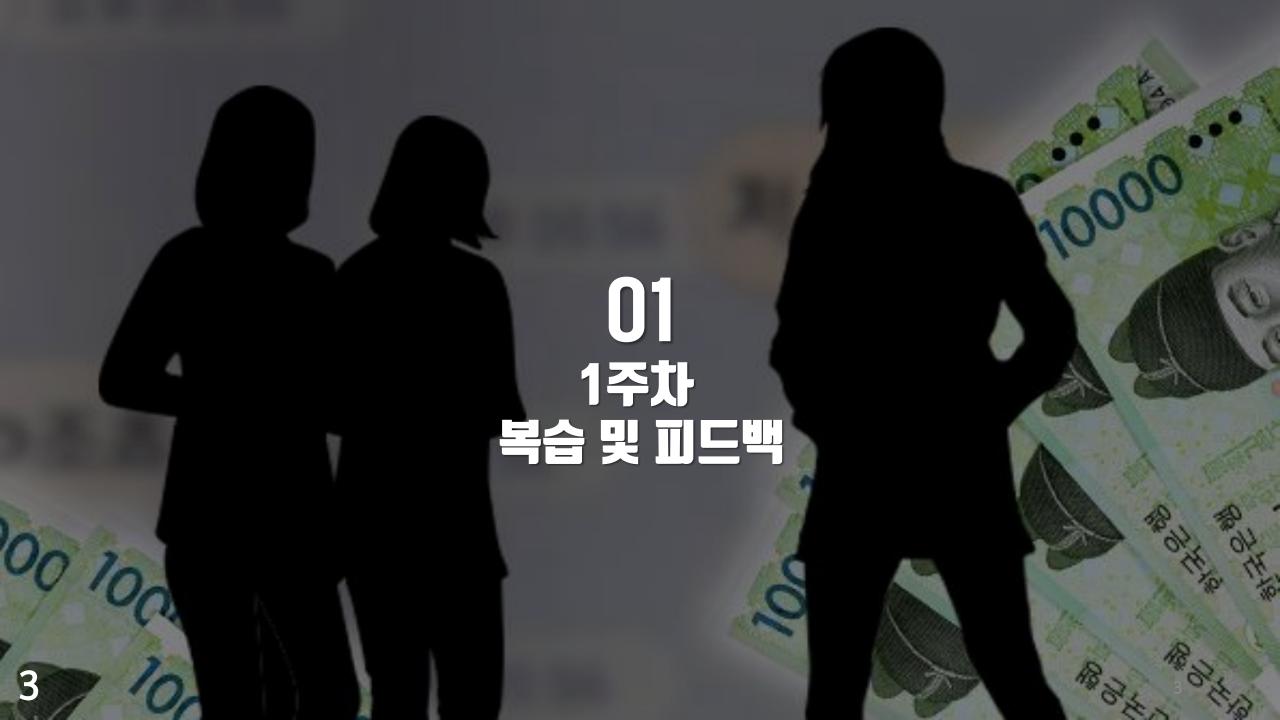
목 차

04 결과해석

05 밀밀 및 한계

02 콕스비레위험모형

> 03 분석과정



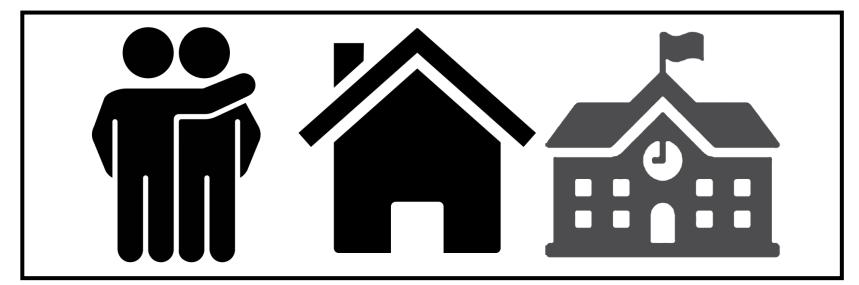
1

지난 주에는..

청소년의 가출은 일시적이고 우발적으로 발생하는 문제가 아니라

개인, 가정, 학교 및 또래 등 다양한 생태체계에 속한 요인들이 복합적으로 작용하여 발생한다.

(박명숙, 2006; 박영호. 김태익, 2002; 배문조. 전귀연)





이러한 생태체계 요인들이 <mark>가출에 미치는 영향력을 실증적으로 검증해보자!</mark>





지난 주에는..

Ⅰ. 일상생활

문3) 학생이 이번 학기(2010-2학기) 중에 하루를 어떻게 보내는지에 대한 질문입니다. 아래 9개 항목을

※ 아래 각 항목에 해당되지 않을 경우에는 '0시간 0분'으로 써 주십시오

	학교 가는 날 (월~금)	학교 가지 않는 날 (놀토, 일, 공휴일)
① 보통 몇 시에 자고 몇 시에 일어나나요?	밤시에 자고 아침시에 일 어난다.	밤시에 자고 아침시에 일 어난다.
② 학원(과외)에서 지내는 시간은 하루 중 얼마나 되나요?	시간 분	시간 분
③ 학교 숙제를 하는 시간은 하루 중 얼마나 되나요?	시간 분	시간 분
④ 학원(과외) 숙제를 하는 시간은 하루 중 얼마나 되나요?	시간 분	시간 분

Ⅱ. 활동과 참여

문7) 지난 일주일 간 학교 체육시간 중 땀을 흘리며 운동한 시간은 몇 시간 입니까? 아래 해당 번호 에 ○표 해 주십시오.

3. 2시간

2. 1시간

5. 4시간 이상

문8) 학생이 중학생이 된 이후, 학교에서 학년 또는 학급 전체가 참가한 수련회 등을 제외하고 가족

과 함께 또는 단체를 통해 한 1박 이상의 여행은 몇 회나 됩니까? 아래에 써 주십시오. * 종교 단체 또는 아동·청소년 단체(예: 보이·걸 스카우트, 누리단, 해양소년단, 우주소년단, RCY) 등을

중학생이 된 이후 _____회

문9) 학생은 중학생이 된 이후, 학교에서 학년 또는 학급 전체가 참가한 것을 제외하고 문화 활동(음 악희, 전시희, 영화, 연극, 뮤지컬 관람 등)을 몇 회나 했습니까? 아래에 써 주십시오.



✓ 중1(2,351명)을 선정하여 7년에 걸쳐 매년 실시한 추 적조사이다.



✓ 가출 항목이 존재하는 2011년-2015년(5년)의 데이터를 이용하여 분석을 진행하였다.



✓ 리커트 척도의 평균을 이용하는 등 다양한 방식으로 N/A값 처리를 하였다.





지난 주에는..

껳생존분석(survival analysis)이란?

:사건이 일어나거나 일어나지 않는 결과와 그러한 사건이 일어날 시점을 예측하고 설명하는 통계방법





생존분석을 위해서는 Censored Data 필요하다!

가출 생존: 식(survival analysis)에던? 사건 이 일어나거나 일어나지 않는 결과와 그러한 사건이 일어날 예측 나고 설명하는 통계방법 Censored data: 관찰 종료

관찰 종료까지 가출x

Censored data: 관찰 도중 탈락

관찰 도중탈락으로 추적불가

주식시장에서 고객 이탁육

취업결정요인 분석



재활 완료까지의 시

관측 시작







2 Censored data Question?!



Q. 절단된 데이터(Censored data)가 많아도 괜찮을까?

(가출한 사람 표본이 작은 게 아닌가?)

A: (Feat. 김동욱 교수님)

생존분석의 목표는 "가출 발생 시점"에 대한 영향력을 파악하는

것이므로, 절단된 데이터가 많은 것 또한 유의미한 정보다!









실제로도 '생존분석'의 다른 분석에선…

표 2.4 업종별 생존율 분석 결과 (단위: 기업)

업종	생존율	부도율	전체
건설업	7,093 (89.4%)	841 (10.6%)	7,934 (100%)
경공업	6,247 (90.2%)	673 (9.8%)	6,920 (100%)
도소매업	18,482 (93.2%)	1,334 (6.8%)	19,816 (100%)
중공업	10,628 (91.4%)	988 (8.6%)	11,616 (100%)
서비스업	4,110 (93.5%)	282 (6.5%)	4,392 (100%)
전체	46560 (91.8%)	4118 (8.2%)	50,678 (100%)

로그-순위 = 150.4681 p<.0001 (df=2) 월콕슨 = 114.1674 p<.0001 (df=2)

(Ⅲ) 연구대상자의 인구사회학적 특성(2015년 기준)

구 분	분류	고졸 이하 (n=181)		전문대졸 (n=171)		대졸 이상 (n=602)	
		빈도	비율(%)	빈도	비율(%)	빈도	비율(%)
MH	남성	139	76.80	98	57.31	285	47.34
성별	여성	42	23.20	73	42.69	317	52.66
연령(세)	평균(표준편차)	24.1	2(2.62)	23.9	1(2.32)	25.4	7(2.30)
거주지역	대도시(특별시, 광역시)	105	58.01	102	59.64	386	64.12
기구시력	중소도시	76	41.99	69	40.36	216	35.88
배우자 유무	유배우	8	4.42	2	1.17	11	1.83
	무배우	173	95.58	169	98.83	591	98.17
생애 첫 취업	예	174	96.13	150	87.72	539	89.53
여부	아니오	7	3.87	21	12.28	63	10.47

〈기업 생존율 분석〉

-부도율이 전체의 8.2%

〈취업 요인 분석〉

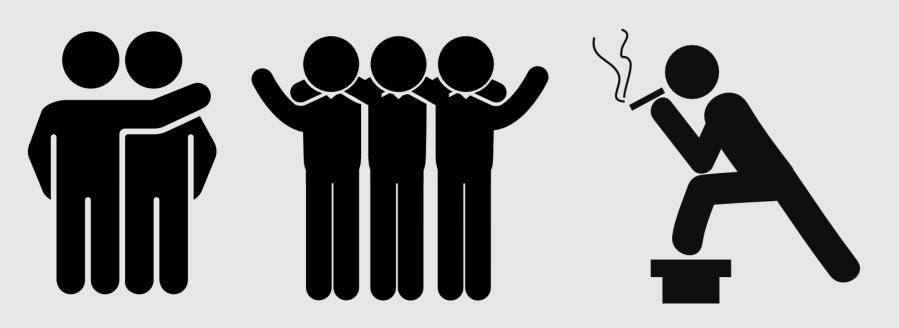
-취업실패율이 전체의 10%

전체 학생의 9%가 가출학생인 우리 데이터에서도 분석이 문제가 없다!

3

변수 추가

(설명력을 좀 높이기 위해) <mark>+새로 추가한 변수</mark>





교외 동아리 유무

비행 행동을 한 친구의 수 (술, 담배, 무단결석 등)



다양한 영역의 체험활동 (ex. 과학, 건강, 문화 등)





4 2주차에는..

1주자

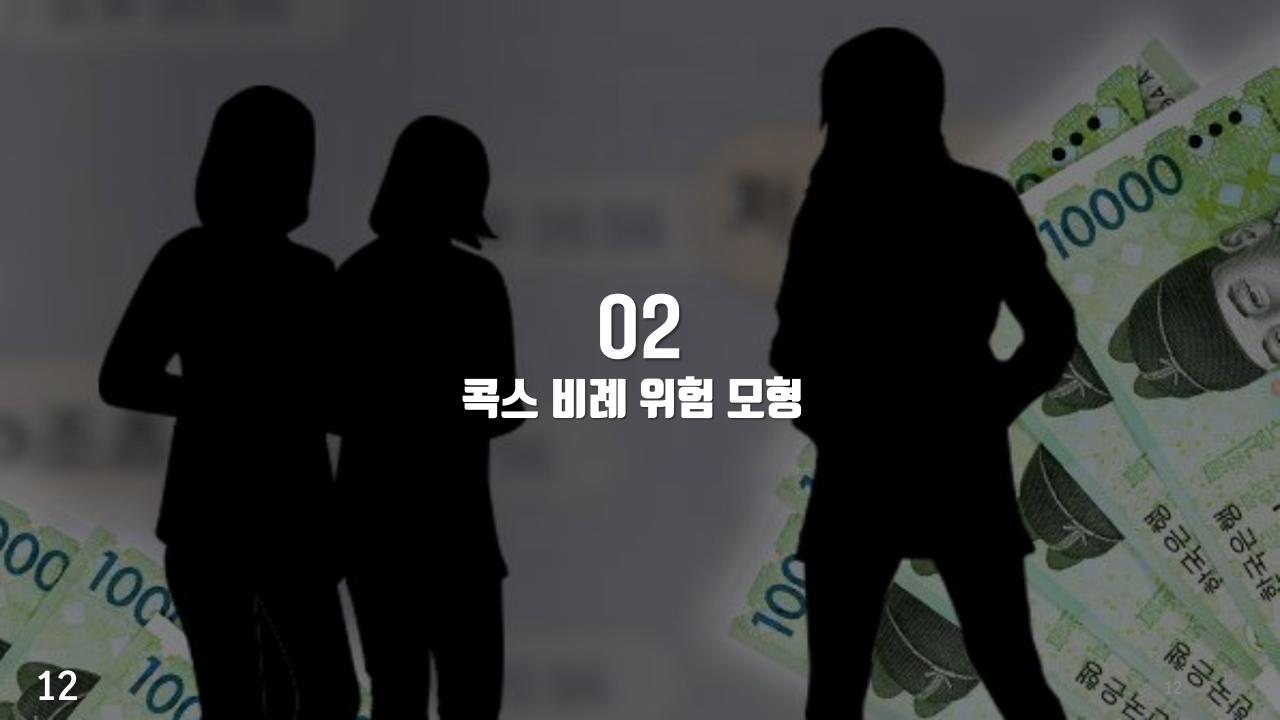
설문조사 데이터를 탐색해보고, 생존 분석에 대해 알아보자.



청소년들의 첫 가출과 가출 시기에 미치는 요인을 분석해보자.







생존 분석

학 다양한 생존 분석 방법

Parametric	Non-parametric	Semi-parametric
분포에 대한 가정 0	분포에 대한 가정 X	분포에 대한 가정 X
AFT 모형	카플란-마이어 추정 방법	Cox 비례위험 모형



콕스 비례 위험 모형

(콕스 비례 위험 모형)

:생존과 관련한 다양한 변수의 영향력을 알아보는 분석 방법







L HI레 의헌 무형 정이

Cox Propotional Hazard Model에서 hazard가 뭐징?

Propotional Hazard Model(곡스 비례 위엄 모영)

:생존과 관련한 다양한 변수의 영향력을 알아보는 분석 방법

해저드(Hazard)

:관측시점(t)까지 생존한 사람이 <mark>관측시점(t) 직후에 순간적으로 사망</mark>할 확률











콕스 비례 위험 모형

🦞 Cox Propotional Hazard Model(콕스 비례 위험 모형)

* X의 영향력이 없을 때 가지는 위험 기저위험(baseline hazard)

공변량의 생존회귀계수

$$h_i(t) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$$

i개체의 hazard

공변량(=설명변수)

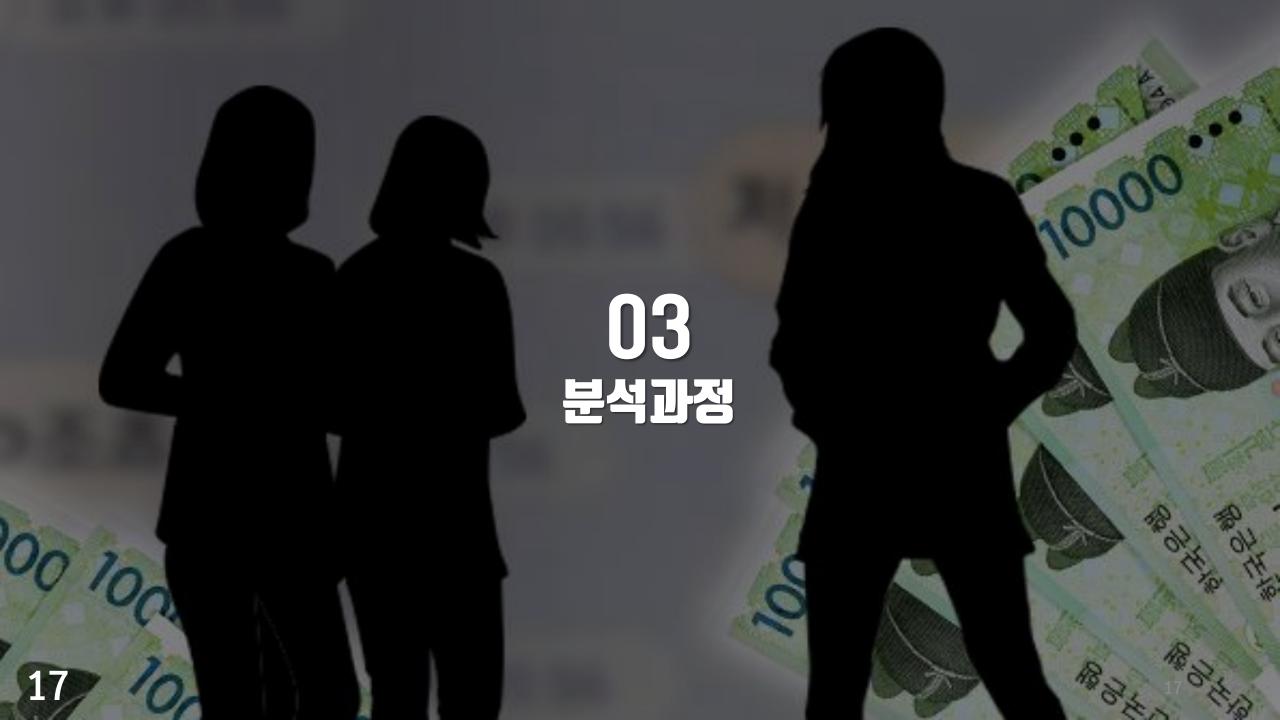


사건이 관측시점(t)에서 발생할 위험!

✓ 사건(가출)이 시점(t)까지 발생하지 않았다는 조건의 회귀분석









0 Modeling



1. Log Linear (가정1)



2. 변수 선택



3. 비례 위험 가정 (가정2)







콕스 비례 위험 모형은,

<mark>공변량과 잔차가 선형 관계</mark>를 가지고 있다고 가정한다.



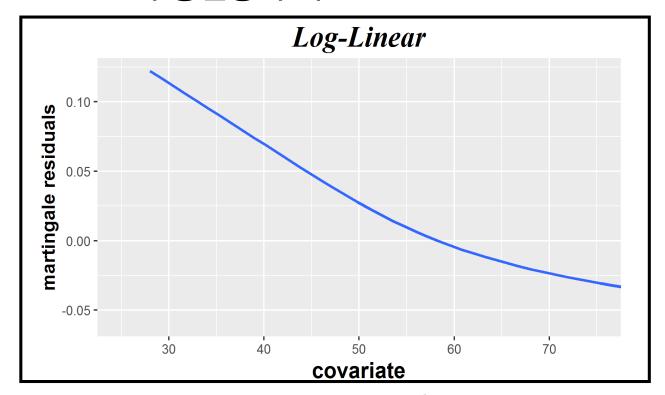
2 Log Linear 가정 확인

목소 비교원데, 모형은, 공변량과 잔차가 선형 관계를 가진다는게 무슨 말일까요…?





Null Cox Model Residual와 공변량의 비교



만약 잔차와 공변량이 선형관계가 있다면? 선형성을 모델에 반영할 수 있다!

* Null Cox Model이란? 공변량을 포함하지

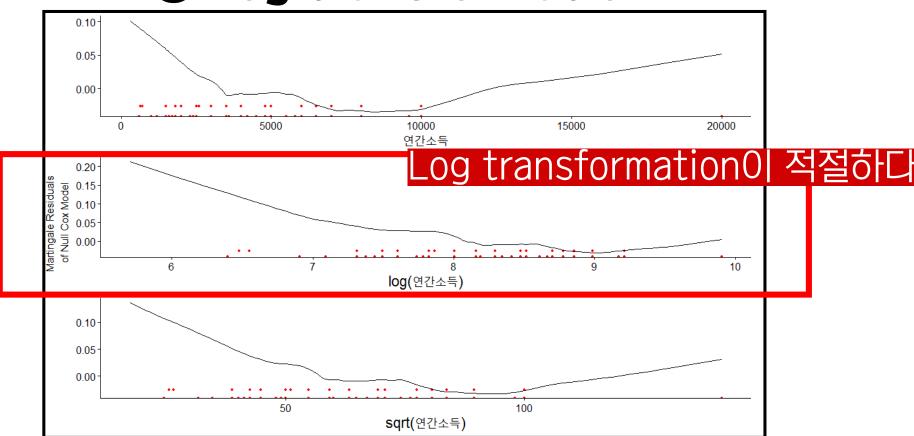








1 Log transformation









과연 Log transformation 효과 있을까?

① log transformation 이전,

```
> summary(coxrq1)
call:
coxph(formula = Surv(age, 가출유무) ~ . . data = runaway3)
  n= 2314, number of events= 167
                                       se(coef)
                           exp(coef)
                           2.388e-01
                                      2.389e-01
                          9.014e-01
                                     2.321e-01 -0.447
                          1.410e+00
                                                       0.49002
                          8.706e-01
                                                      0.55815
                                     2.366e-01 -0.586
                          2.116e-01
                         2.116e+00
                                                      0.28524
보호자건강2
                          7.378e-01
보호자건강3
보호자건강4
                          3.816e-01
                                     8.446e-01 -1.141
                          9.583e-01 1.024e-02 -4.160 3.19e-05
                          7.800e-01
                         1.015e+00
                                     2.795e-01
                                               0.053
                                                       0.95773
                          6.934e-01
                                     8.622e-01 -0.425
                                                       0.67107
```

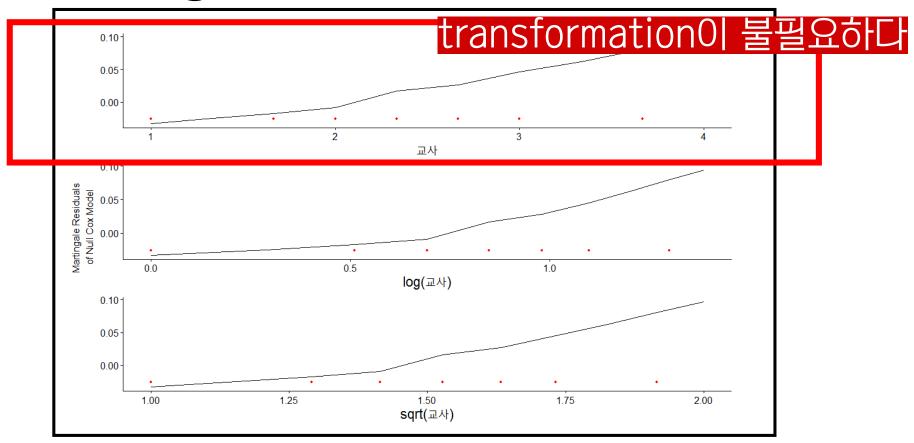
② log transformation 이후,

```
> summary(coxrg2)
call:
coxph(formula = Surv(age, 가출유무) ~ ., data = runaway4)
  n= 2314, number of events= 167
                     coef exp(coef) se(coef)
                                    0.198564
                                   0.225781
                         1.178797
                                   0.492105
                         0.920649
                                   0.229771
                                            -0.360 0.718980
                         0.501804
                        2.931082
                                   0.355082
                                             3.029 0.002458
                         3.370698
                        1.070764
                        1.836598
                         0.701615
                                   0.223558 -1.585 0.112934
보호자건강3
보호자건강4
                         0.764670
                          0.086047
                                    0.509242
                         0.682925
                                            -2.019 0.043503
                         0.920784
                                   0.264104
                                            -0.312 0.754667
건강상태4
                         1.169688
                                   0.756914
                                             0.207 0.835952
```

1

Log Linear

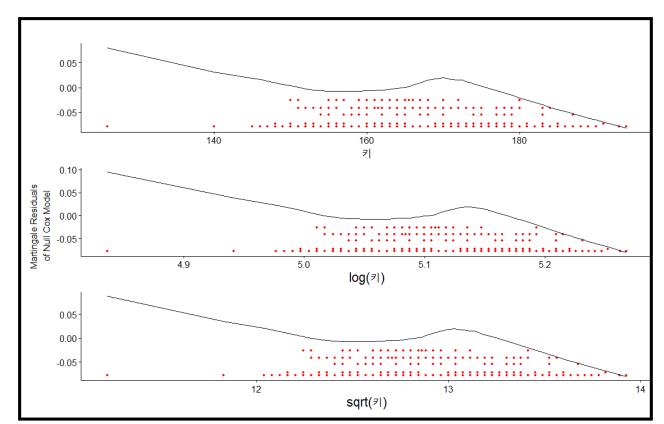
2 No transformation







③ Delete: 변수 삭제



transformation도 효과가 없











변수들의 log linear 가정 결과

1 Log transformation

연간소득 몸무게 가족여행 damage 방임 가족연락 가출친구 비행친구 체험활동 2 No transformation

성별 부 최종학력 모 최종학력 형제자매유무 부 종사상지위 전학경험유무 모 종사상지위 학대 보호자건강 학습 건강상태 교사 성적만족도 교내동아리 가족구성 교외동아리 다문화가정여부

3 Delete

키 sleep 교우 사회성 친구연락 휴대전화 의존도





Log linear 가정 확인을 통해 변수 변환 및 삭제

coxrg2 <- coxph(Surv(age, 가출유무) ~ ., data = runaway4)

연간소득 몸무게 가족여행 damage 방임 가족연락 가출친구 비행친구

Log transformat

```
runaway4 = runaway3
runaway4$연간소득 = log(runaway4$연간소득)
runaway4$9| = NULL
runaway4$몸무게 = log(runaway4$몸무게)
runaway4$가족여행 = log(runaway4$가족여행)
runaway4$damage = log(runaway4$damage)
runaway4$sleep = NULL
runaway4$방임 = log(runaway4$방임)
runaway4$교우 = NULL
runaway4$사회성 = NULL
runaway4$가족연락 = log(runaway4$가족연락)
runaway4$친구연락 = NULL
runaway4$휴대전화의존도 = NULL
runaway4$가출친구 = log(runaway4$가출친구)
runaway4$비행친구수 = log(runaway4$비행친구수)
runaway4 $ 체험활동= log(runaway4 $ 체험활동)
```

Delete

키 sleep 교우 사회성 친구연락 대전화 의존도

이를 바탕으로 <mark>콕스 모델</mark>을 세운다!

새로 만든 모델의 변수선택을 해보자,

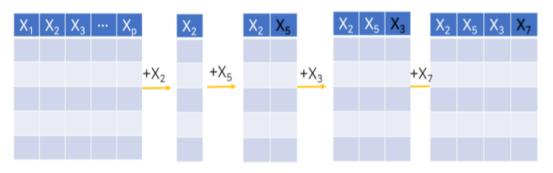
StepAIC 를 사용해서!

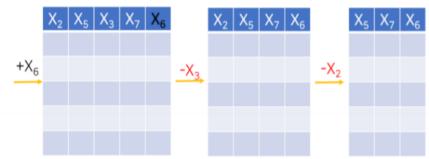
StepAIC란? Stepwise 방식으로 변수를 선택하는데, AIC가 작아지는 방향으로 변수를 선택해낸다!

2

변수 선택

► Hybrid approach[©] MIAI





전진 선택과 유사한 방식으로 모델에 순차적으로 변수가 추가, 새 변수를 추가한 후 모형 적합성을 개선하지 않는 변수를 제거할 수도 있다. 변수선택방법은 데이터에 대한 설명력이 뛰어나며, 동시에 단순한 형태를 가지는 모형을 찾는다.

방법: best, forward, backward, hybrid

Hybrid approach를 이용해서 변수선택을 진행







2

변수 선택

StepAIC를 R로 자세히 알아보자!

ro.step <- stepAIC(coxrg2, trace=FALSE)</pre>



stepAIC function: AIC가 작아지는 방향으 로 변수를 선택한다.

> summary(ro.step)

call:

coxph(formula = Surv(age, 가출유무) ~ 성별 + 부종사상지위 + 모종사상지위 + 연간소득 + 몸무게 + 성적만족도 + 가족구성 + 가족여행 + damage + 학대 + 학습 + 교사 + 가족연락 + 가출친구 + 교외동마리 + 체험활동, data = runaway4)



summary를 통해 선택된 변수 확인

> extractAIC(coxrg1)[[2]] [1] 2200.474 > extractAIC(ro.step)[[2]]
[1] 2145.494



원래 모델보다 AIC 가 확연히 줄었다!







3 비례 위험 가정

비례 위험 가정을 확인해보자!

그래프와 수치적 확인을 통해!

3 비례 위험 가정

비례 위험 가정: 위험비는 시간에 의존하지 않고 항상 일정(비례)하다

위험비
$$\frac{h_i(t)}{h_0(t)} = \exp\left(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p\right)$$
 변하지않는다



16살에 담배를 피는 학생이 피지않는 학생보다 가출할 위험비가 2배



<mark>18살</mark>에 담배를 피는 학생이 피지 않는 학생보다 가출할 위험비도 2배





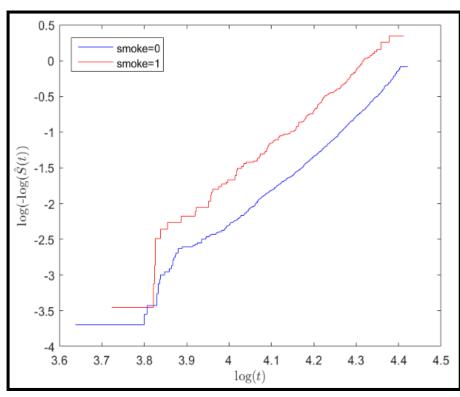




3

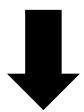
비례 위험 가정

그래프를 통해 이해해보자



log(−log(S(t))) transformation으로 고벼랴 가에 따르 새조하스 항이 가느!

비례 위험 가정: 시간에 관계 없이 위험 비율이 일정하다.



두 그래프가 <mark>시간에 관계없이</mark> <mark>일정한 간격</mark>을 두고 있어야 한다!





3 비례 위험 가정 확인

P

그래프를 통해 이해해보자



비례 위험 가정: 시간에 관계 없이 위험 비율이 일정하다.

청소년 가출 데이터의 변수에서 확인해보자!

두 그래프가 <mark>시간에 관계없이</mark> 일정한 간격을 두고 있어야 한다!

log(−log(S(t)))transformation으로공변량 값에 따른 생존함수확인 가능!



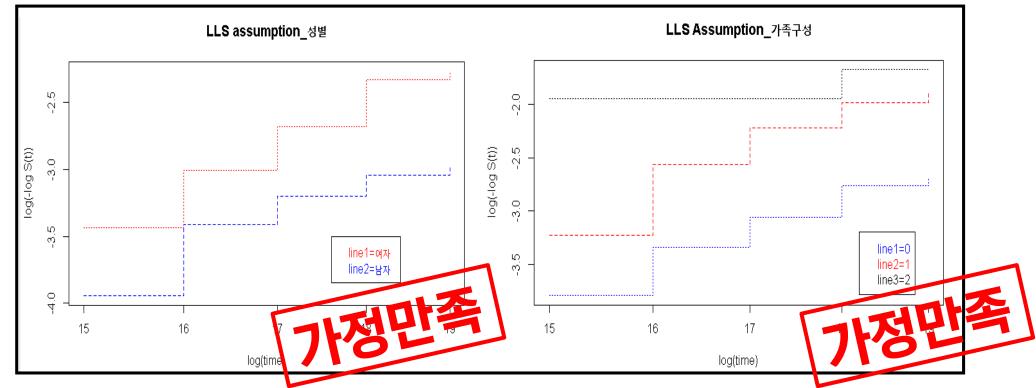


3 비례 위험 가정

STEP1. 그래프로 확인

〈성별〉

〈가족구성〉



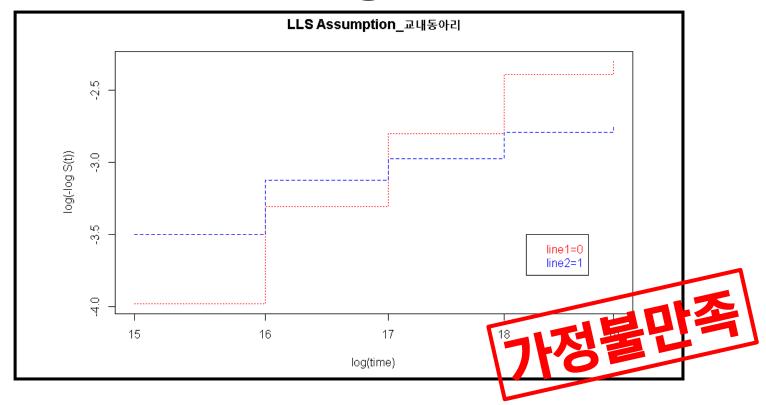




3 비례 위험 가정

STEP1. 그래프로 확인

〈교내동아리〉





외 변수는 부록 참고



비례 위험 가정

STEP2, 통계적 검정 비례위험 모형 가정: 공변량(x)의 효과인 $oldsymbol{eta}(t)$ 가 시간에 따라 변하지 않는다.

```
> cox.zph(ro.step)
                  rho
                          chisq
             -0.07778
                      1.21867 2.70e-01
부종사상지위1 -0.04054
                     0.34578 5.57e-01
부종사상지위2
            0.00630
                     0.00772 9.30e-01
모종사상지위1
           -0.10394
                     1.87937 1.70e-01
모종사상지위2
            0.02023
                     0.06966 7.92e-01
연간소득
             0.09650
                      1.76725 1.84e-01
                      1.49336 2.22e-01
             0.08766
                      0.14447 7.04e-01
            -0.02947
            0.00311
                      0.00163 9.68e-01
            -0.01858
                      0.06102 8.05e-01
             0.04722
                      0.49014 4.84e-01
            -0.02054
                      0.08366 7.72e-01
가족여행
            -0.18362
                      6.48265 1.09e-02
damage
             -0.00732
학대
              0.14063
                       3.84125 5.00e-02
학습
              0.14557
                       3.93640 4.73e-02
교사
             -0.11449
                      2.23809 1.35e-01
가족연락
            -0.07006
                      0.83126 3.62e-01
가출친구
             0.06950
                      0.99128 3.19e-01
교내동아리2
            -0.30295 18.52837 1.67e-05
교외동아리2
            -0.04890
                      0.41739 5.18e-01
체험활동
            -0.10399
                      2.24206 1.34e-01
```

시간 가변 회귀 계수 : $\beta(t) = \beta + \theta g(t)$

귀무가설 H_0 : $\theta = 0$

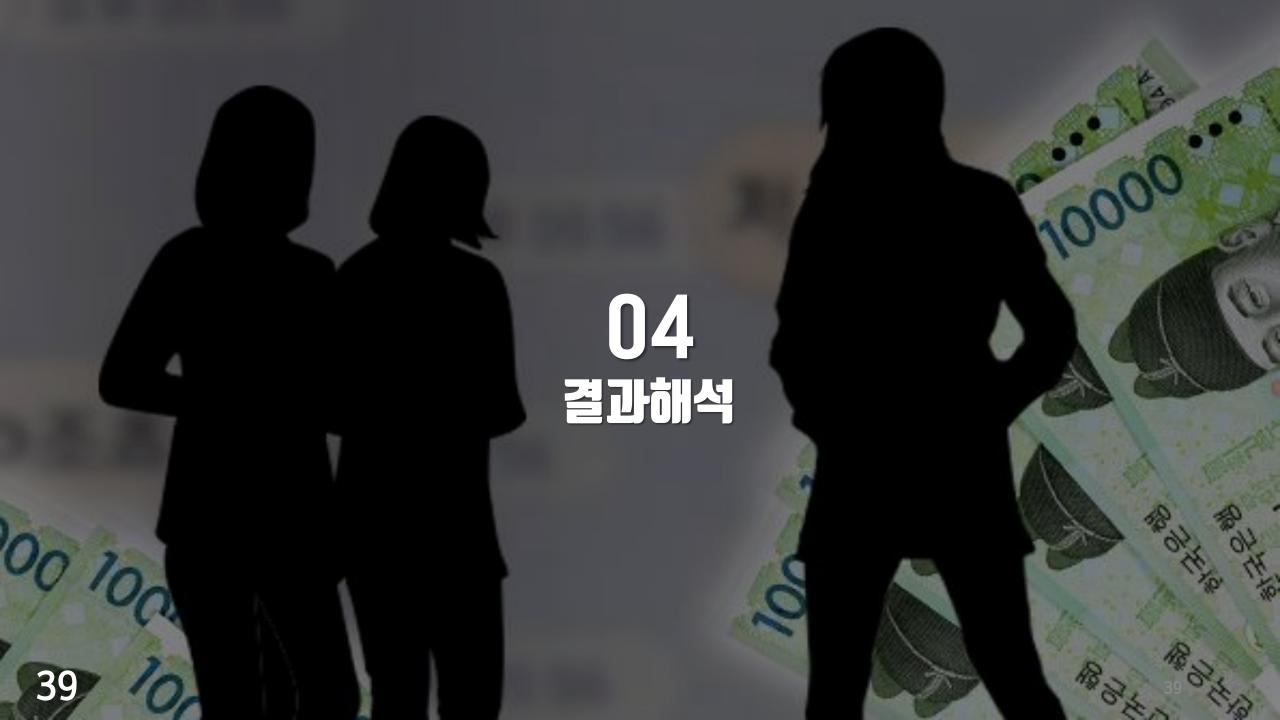












과 해석

최종 선정 모델 결과

```
> summary(ro.step)
call:
coxph(formula = Surv(age, 가출유무) ~ 성별 + 부종사상지위 + 모종사상지위 +
   연간소득 + 몸무게 + 성적만족도 + 가족구성 + 가족여행 + damage +
   학대 + 학습 + 교사 + 가족연락 + 가출친구 + 교외동아리 + 비행친구수 +
   체험활동, data = runaway4)
 n= 2314, number of events= 167
                coef exp(coef) se(coef)
                      0.39758 0.19801 -4.658 3.19e-06
부종사상지위1
           0.98467
                     2.67692 0.33899 2.905 0.003676 **
부종사상지위2
           1.08663
                     2.96428
                             0.34183
모종사상지위1
           0.09170
                                     0.478 0.632847
                    1.09604
모종사상지위2
           0.64496
연간소득
           -0.44339
                     0.64186 0.15144 -2.928 0.003413 **
몸무게
           -2.35795
                     0.09461
                              0.48570 -4.855 1.21e-06 ***
성적만쪽도2
           -0.12859
                     0.87933
                             0.37766 -0.340 0.733485
성적만족도3
            0.20089
                     1.22249 0.37706
석적만족도4
            1.25283
                     3.50022 0.38487
                                      3.255 0.001133 **
가족구성1
            0.63309
                     1.88341 0.27264 2.322 0.020230
가족구성2
            1.42783
                     4.16964 0.51641 2.765 0.005694 **
가족여행
            0.49084
                     1.63369 0.12341 3.977 6.97e-05 ***
damage
             0.96824 2.63329 0.24386 3.970 7.17e-05 ***
학대
             0.63506
                      1.88714 0.10822 5.868 4.41e-09 ***
학습
             0.26107
                      1.29832
                              0.13533
                                      1.929 0.053722
교사
             0.27357
가족연락
            0.85560
                     2.35278
                             0.22746 3.762 0.000169 ***
가출친구
                    2.82698 0.14350 7.242 4.42e-13
            1.03921
교외동아리2
           -0.50635
                     0.60269
                             0.28117 -1.801 0.071727
비행친구수
           -0.25446
                     0.77533 0.11148 -2.283 0.022458 *
체험활동
            0.63201
                             0.13723 4.605 4.12e-06 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Concordance= 0.893 (se = 0.023 )
Rsquare= 0.176
                 (max possible= 0.67 )
Likelihood ratio test= 446.9
                                           p=<2e-16
                               on 22 df.
wald test
                                           p=<2e-16
                     = 556.6
                              on 22 df.
Score (logrank) test = 1009 on 22 df,
                                          p=<2e-16
```

Likelihood ratio test < 0.05 Wald test < 0.05Score test < 0.05





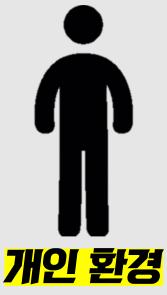




결과 해석



변수들을 분류해 보자!



성별, 소득수준, 부모 종사상지위, 몸무게, 가족구성



가족여행, 가족연락, 학대



학교 환경

폭력, 학습, 성적만족도, 체험, 가출친구수, 비행친구수, 교사관 계, 동아리 및 체험활동









시에 의해 무형 정이

Cox Propotional Hazard Model을 어떻게 해석하지?

Propotional Hazard Model(곡스 비례 위염 모영

Hazard ratio(위험비)로 해석한당!

 $h_i(t) = (h_0(t)) \exp(\beta_1 X_1 + \beta_2 X_2) + \dots + \beta_n X_n$

$$\frac{h_i(t)}{h_0(t)} = \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$$

연속형 자료: X_1 이 한 단위 증가할 때, hazard ratio가 e^{β_1} 배 된다.

범주형 자료: X_2 =1이라면, X_2 =0일 때보다 hazard ratio가 e^{β_2} 배 된다.



결과 해석

🍟<mark>개인 환경</mark> 계수 해석

*log 변환 취해준 변수들은 1단위가 아닌 1%변화 할 때로 해석한다

부종사상지위 $1 (\beta = 0.9846)$

부의 지위가 0이 아니라 1일 경우

*근로를 안 하는 경우보다 하는 경우

가출을 할 위험비가 2.676(e^{0.9846})배

∴ 가출을 할 확률이 증가 성별 (-0.9223)

남자(1)가 아니라 여자(2)일 경우

가출을 할 위험비가 $0.397(e^{-0.9223})$ 배

∴가출을 할 확률이 감소 연간소득 (-0.4433)

연간소득이 1% 증가할수록

가출을 할 위험비가 $0.995(e^{-0.4433*0.01})$ 배

∴ 가출을 할 확률이 감소



99 兴計湖



<u> 1주차피드백 콕스비레위험모형 분석과정 분석결과</u>

결과 해석

TO 환경 계수 해석

학대 (0.635)

학대가 1% 증가할 때

*학대 정도가 심할수록

가출을 할 위험비가 $1.006(e^{0.635*0.01})$ 배

∴가출을 할 확률이 증가 가족여행 (0.4908)

가족여행이 1% 증가할 때

가출을 할 위험비가 1.004(e^{0.4908*0.01})배

∴ 가출을 할 확률이 증가 가족연락 (0.8556)

가족연락이 1% 증가할 때

*가족과의 연락이 적을 수록

가출을 할 위험비가 1.008(e^{0.8556*0.01})배

∴ 가출을 할 확률이 증가







결과 해석

🍟<mark>학교 환경</mark> 계수 해석

*변수별 정확한 의미는 부록 참고

학교폭력 (0.9682)

damage가 1% 증가할 때

*비행 피해가 많을 수록

가출을 할 위험비가 1.009(e^{0.9682*0.01})배

∴가출을 할 확률이 증가 성적만족도4 (1.2528)

성적만족도가 매우 낮은 경우는 매우 좋을 때보다 가출을 할 위험비가 3.5($e^{1.2528}$)배

∴가출을 할 확률이 증가 가출친구 (1.039)

가출친구가 1% *증가할 때*

가출을 할 위험비가 1.01(e^{1.039*0.01})배

∴ 가출을 할 확률이 증가







결과 해석

유의미한 정도가 큰 변수(***)

성별, 몸무게, 가족여행, damage(비행피해 경험), 학대, 가족연락, 가출친구, 체험활동



학대에 대한 노출, 가족과 연락 빈도, 가출친구의 수, 가족여행 수와 같이 주변사람과의 관계가 가출에 매우 유의미하게 나타났다.

범주형 변수 중 계수가 큰 변수

부종사상지위1,2: 2.67692

(기준 범주 – 근로하지 않는 경우)

성적만족도4: 3.50022

가족구성2: 4.16964

(기준 범주 – 부모 모두 계신 경우)



부종사상지위나 가족구성과 같이 환경적요소가 매우 열악한 경우는 가출에 큰 영향을 미쳤다.

2 Cox vs Logistic

Logistic과 결과 비교

*비교를 위해 같은 변수들로 logistic을 돌려보았다

<Logistic regression>

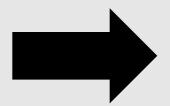
+종사상지위(모), 가족구성, 교사, 동아리, 비행친구

<Cox regression> > summary(ro.step)

data = runaway4)

la = Surv(age, 가출유무) ~ 성별 + 부종사상지위 + 모종사상지위 + 몸무게 + 성적만족도 + 가족구성 + 가족여행 + damage + 다습 + 교사 + 가족연락 + 가출친구 + 교외동마리 + 비행친구수 +

```
glm(formula = 가출유무 ~ 성별 + 부종사상지위 + 모종사상지위
    연간소득 + 몸무게 + 성적만족도 + 가족구성 + 가족여행 +
학대 + 학습 + 교사 + 가족연락 + 가출친구 + 교외동아리
    체험활동, family = "binomial", data = runaway4)
Deviance Residuals:
                  Median
        -0.2897
                 -0.1847
-2.4385
                          -0.1124
                                    3.9417
Coefficients:
               Estimate Std. Error z value Pr(>|z|)
               6.150793
(Intercept)
                         3.055732
                                   2.013 0.044128
성별2 -1.06826
부종사상지위1 1.192397
                        0.242596
                                  -4.403 1.07e-05 ***
             -1.068267
                                  2.662 0.007774 **
                       0.447975
부종사상지위2
            1.192191
                       0.456681
                                  2.611 0.009040 **
모종사상지위1
            0.034133
                       0.230362
                                  0.148 0.882207
모종사상지위2 0.488343
                       0.340975
                                  1.432 0.152087
연간소득
몸무게
            -0.460549
                        0.195523
                                  -2.355 0.018499
                                  -4.286 1.82e-05
             -2.614122
                        0.609956
성적만족도2
            -0.148428
                        0.453020
                                 -0.328 0.743182
성적만족도3
             0.008216
                        0.453641
                                   0.018 0.985551
정적만족도4
             1.035370
                        0.472880
                                   2.190 0.028560
가족구성1
             0.566271
                        0.341268
                                   1.659 0.097053
가족구성2
             0.991025
                        0.751038
                                             6989
가족여행
             0.442745
                        0.151351
                                             B441 **
                                              7e-06 ***
damage
              1.989075
                         0.4
                                              e-10 ***
학대
              0.880510
                         0.14
학습
              0.660665
                         0.18
                                              0247
교사
              0.256775
                         0.164
                                             19019
가족연락
             0.938714
                        0.275
                                      13 0.000642
가출친구
             1.890845
                        0.240528
                                   7.861 3.80e-15
교외동아리2
            -0.489025
                        0.411304 -1.189 0.234454
비행친구수
            -0.192521
                                 -1.466 0.142691
                        0.131338
체험활동
             0.672730
                                   3.790 0.000151
```



number of events= 167 coef exp(coef) se(coef) z Pr (> |z|) -0.92235 0.39758 0.19801 -4.658 3.19e-06 부종사상지위1 2.67692 0.33899 2.905 0.003676 ** 부종사상지위2 2.96428 0.34183 3.179 0.001479 모종사상지위1 0.09170 1.09604 0.19196 0.478 0.632847 모종사상지위2 0.64496 1.90591 0.26997 2.389 0.016894 0.64186 0.15144 -2.928 0.003413 ** 0.09461 0.48570 -4.855 1.21e-06 *** 성적만족도2 0.87933 0.37766 -0.340 0.733485 석적만족도3 1.22249 0.37706 0.533 0.594192 석적만족도4 3.50022 0.38487 가족구성1 1.88341 0.27264 가족구성2 4.16964 0.51641 2.765 0.005694 가족여행 0.49084 1.63369 0.12341 3.977 6.97e-05 damage 0.96824 2.63329 0.24386 3.970 7.17e-05 1.88714 0.10822 5.868 4 41e-09 0.63506 학습 0.26107 1.29832 0.13533 0.27357 1.31465 0491 0.85560 169 교외동아리2 -0.50 비행친구수 -0.25 체험활동 4.605 4.12e-06 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1 Signif. codes: 0

+시간이라는 추가적인 정보를 고려함으로써 유의미한 변수를 더 생성해냈다











결과 시각화

청소년의 가출 시기에 영향을 미치는 요인을 봤으니, 우리가 만든 모델로 <mark>"가출 시점" 위험도</mark>를 시뮬레이션 해보자!

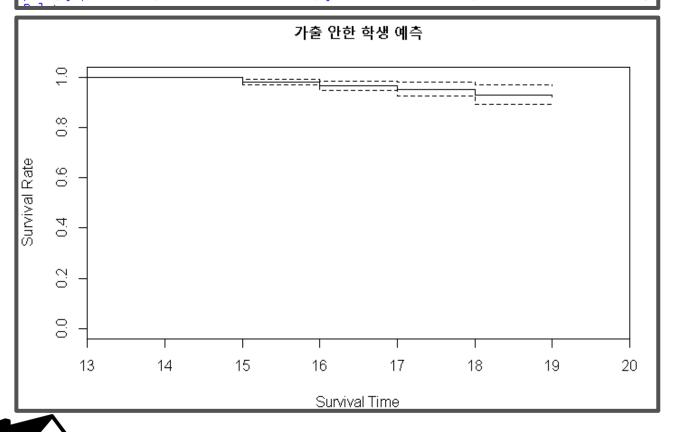


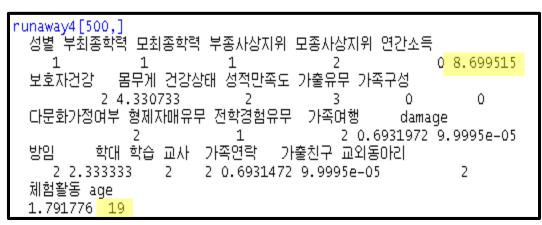




생존분석 결과 시각화

```
# runaway4[500,] : 가출x
my_prediction <- survfit(ro.step, newdata = runaway4[500,])
summary(my_prediction)
plot(my_prediction, xlab="survival Time", ylab="survival Rate",main="가출 안한 학생 예측")
```



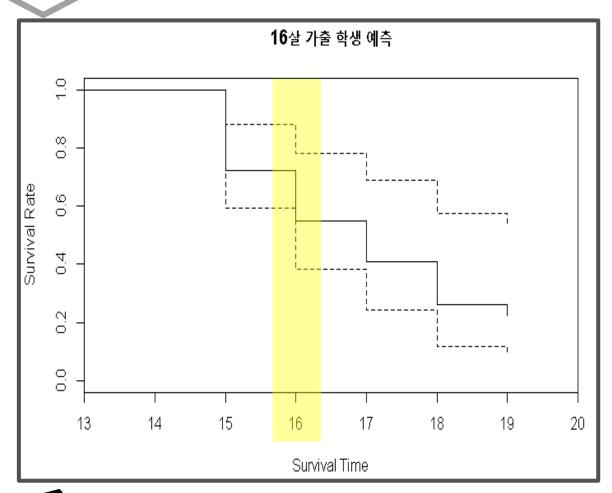


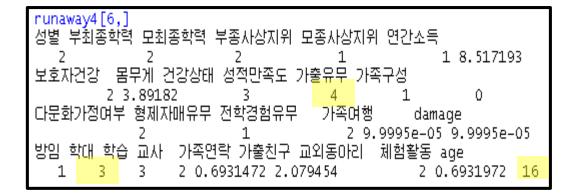


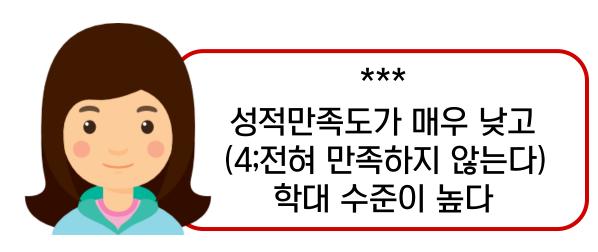




생존분석 결과 시각화







학생2



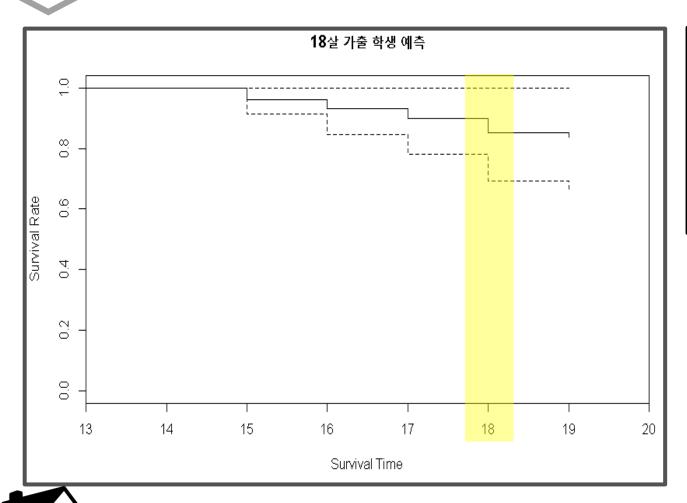


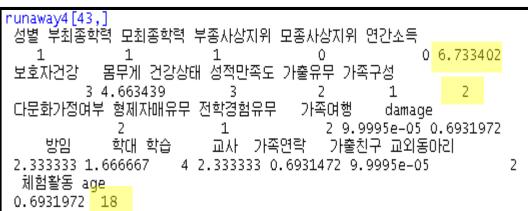


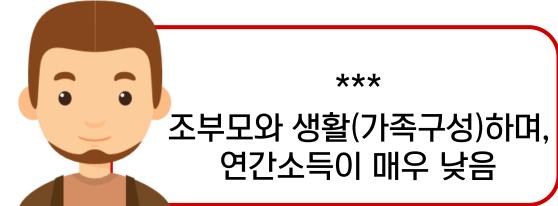




생존분석 결과 시각화





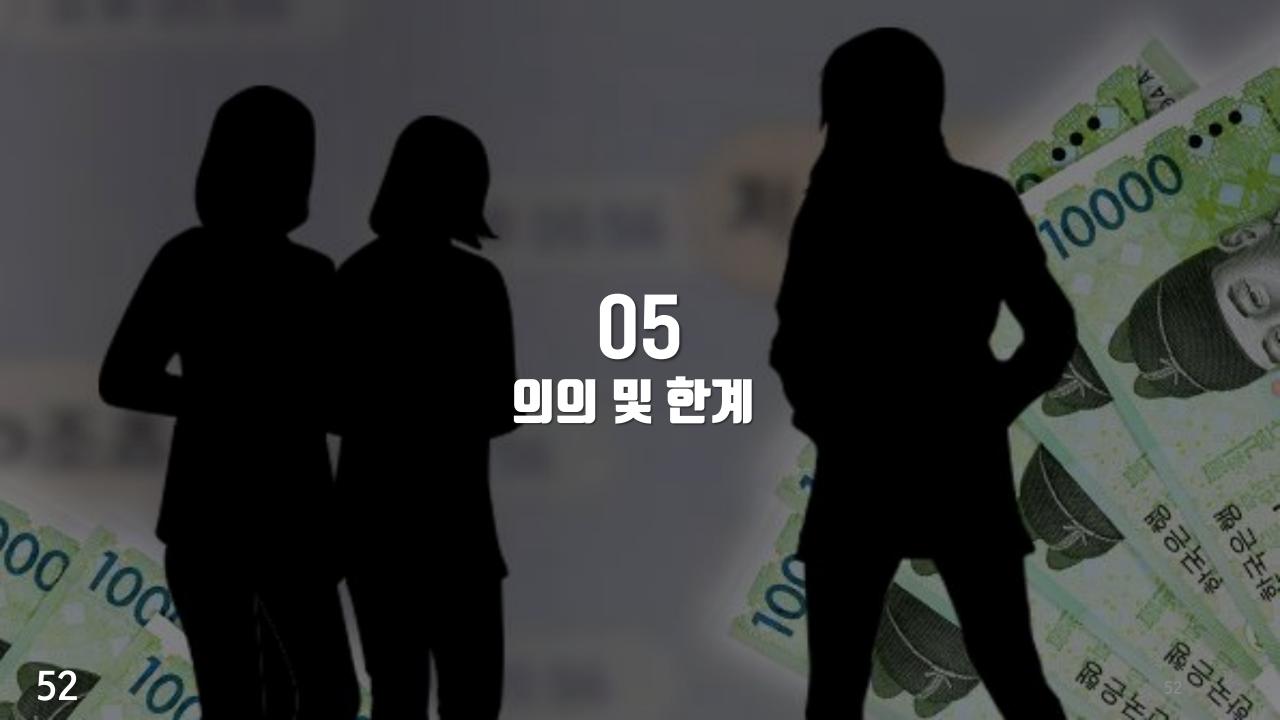


학생3







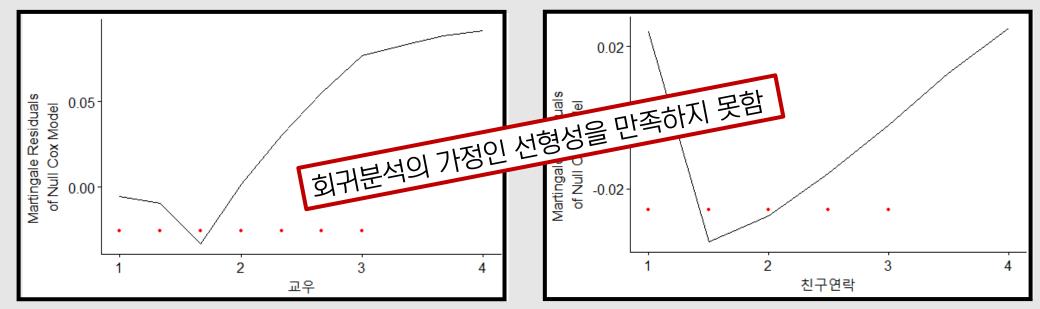


분석의 한계점



가정을 만족하지 못하는 <mark>변수 제외</mark>

: 유의미해 보이는 변수를 모두 사용하지 못했다.



교우관계가 원만하지 못한 경우와 교우관계가 활발한 경우를 가출 요인으로 반영하지 못했다.





분석의 한계점

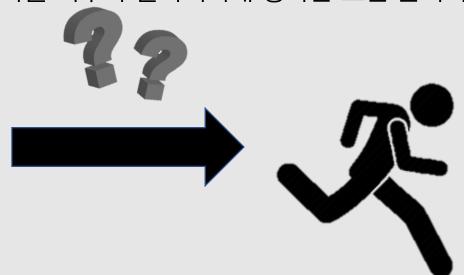


가출 요인에 대한 명확한 선후 관계 파악 불가

:설문 기점에 따라 가출 여부가 달라지기에 명확한 요인 분석이 어렵다.







-설문데이터의 한계로, 선후관계가 불명확하기에 정확한 분석이 어려웠다.





석의 의의



설문데이터를 최대한 이용 노력

: N/A 처리, 5개년의 정보를 정리하여 모두 사용하는 등 설문데이터를 최대한으로 이용하려고 했다.

연간 피해경험 유무



학생 1

학생 2

2회 리커트 척도의 0회

추적조사 특성 상 대체 불가한 행 삭제

전후 년도 데이터 활용









분석의 의의



<mark>사회적인 문제</mark>를 수치적으로 분석

: 사회적 문제를 단순히 판단하기 보다는 수치적으로 청소년 가출시기에 영향을 끼치는 요인을 분석해보았다.









활용 방안



<mark>가출 관련된 유의미한 요인</mark>을 통해 도움 방안 제시

:유의미하게 도출된 13개의 변수를 통해 가출을 예방할 수 있는 방안을 제시할 수 있다.

개인요소	성별, 부/모 종사상지위,연간소득, 가족구성
<mark>가족요소</mark>	가족여행, 부모님의 학대, 가족연락 횟수
학교요소	성적만족도, 가출 친구 수, 동아리, 체험활동 횟수,
	교사와의 관계, damage(비행 피해 종류 수)



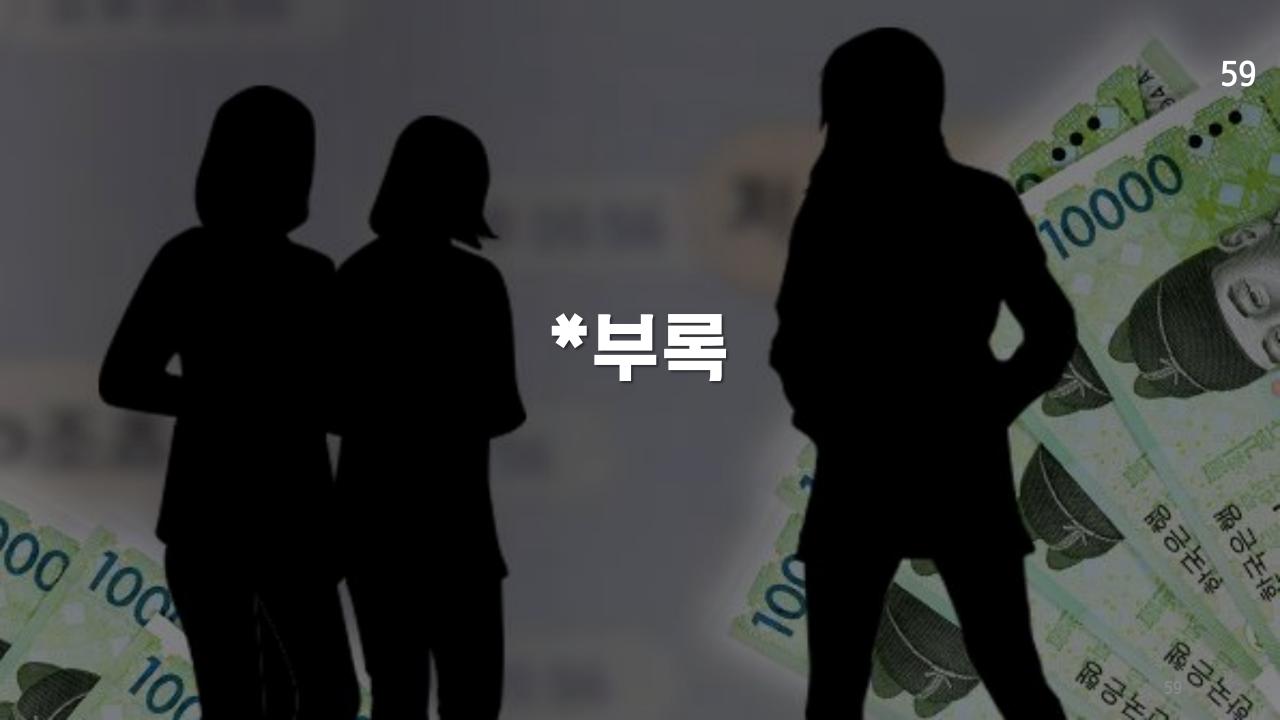
가족 기능을 강화하기 위해 <mark>가족 회복 운동</mark>실시

동아리, 체험활동 등의 프로그램 적극적 실시









Log 변환 해석 방법!

level-log model

$$y = \alpha + \beta \log x$$

• 앞서와 마찬가지로 Δx , Δy 를 적어보자.

$$y + \Delta y = lpha + eta \log(x + \Delta x)$$

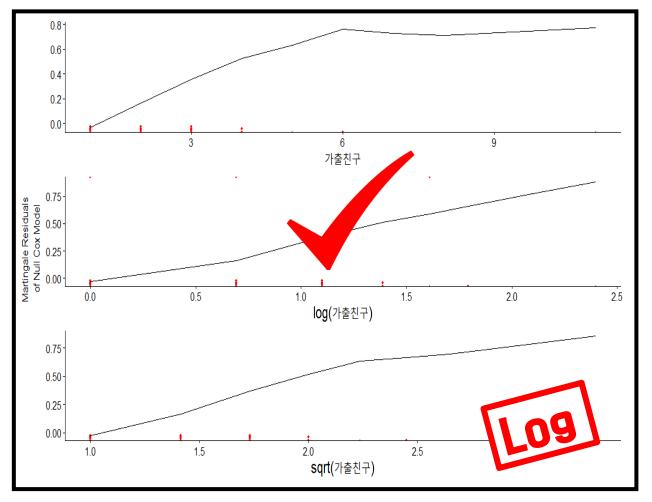
ullet 양변에서 y를 빼주면 아래와 같이 전개된다.

$$\Delta y = eta(\log(x+\Delta x) - \log x) = eta\log(1+rac{\Delta x}{x})$$

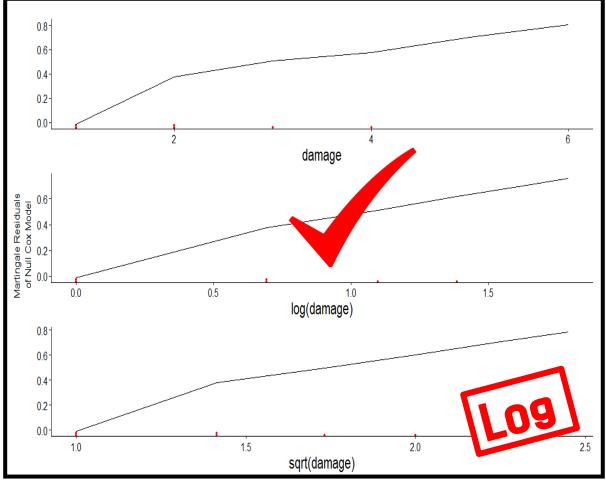
- 만일 $\frac{\Delta x}{x}$ 의 값이 작다면, $\log(1+\frac{\Delta x}{x}) pprox \frac{\Delta x}{x}$ 를 사용할 수 있다. $\frac{\Delta x}{x}$ 를 %로 표현하고 싶다면 100을 곱하면 된다. 따라서 $\Delta y pprox \frac{\beta}{100} (100 imes \Delta x/x)$ 라고 볼 수 있다.
- ullet 같은 결과를 편미분을 통해 도출할 수 있다. 양변을 x에 대해서 미분하면 $dy=etarac{dx}{x}$
- 결국 x가 1% 만큼 증가할 때 $y \in \frac{\beta}{100}$ 만큼 변한다는 것을 의미한다.

부록 - Log Linear

〈가출 친구수〉

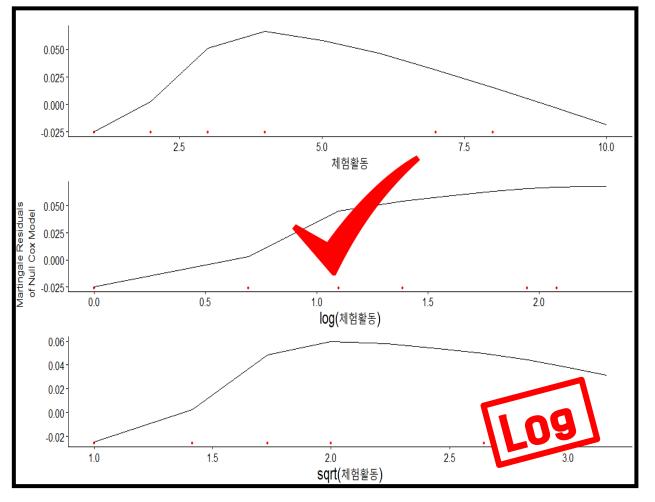


〈학교폭력〉

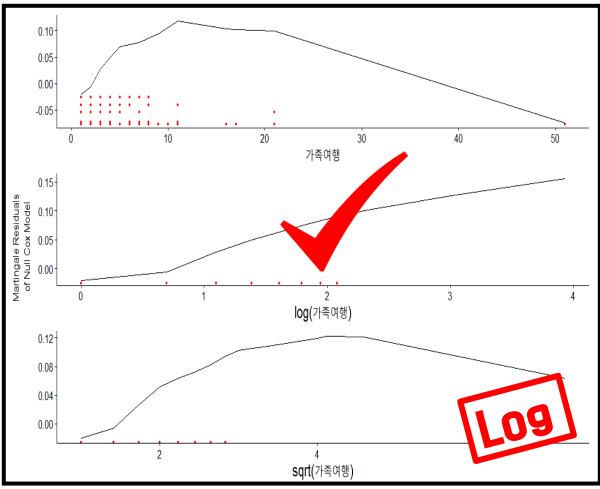


부록 - Log Linear

〈체험활동 참여〉

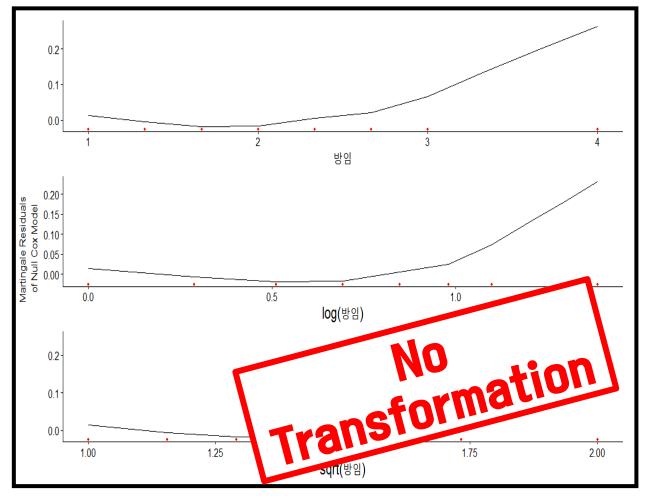


〈가족여행〉

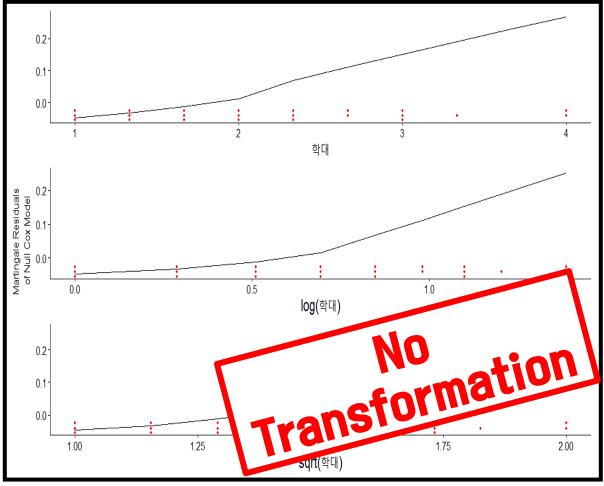


부록 - Log Linear

〈방임〉



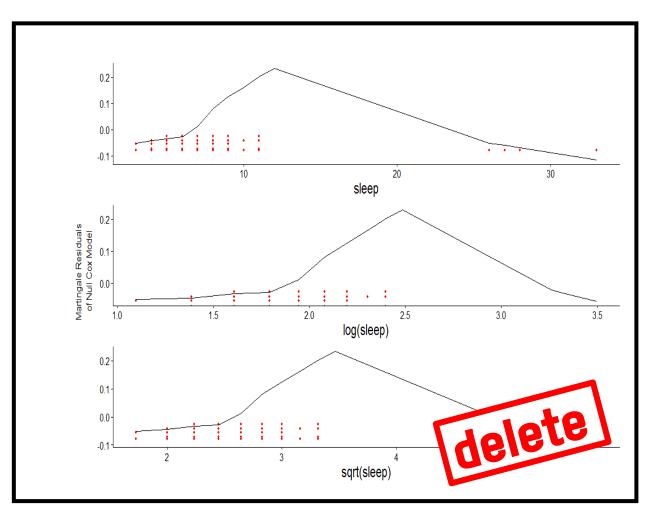
<학대>

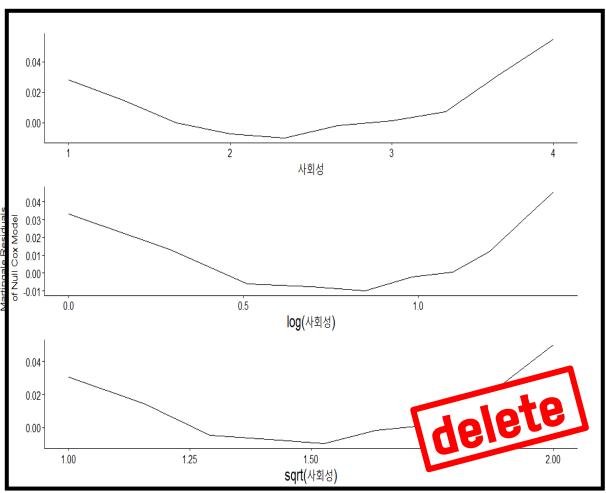


부록 - Log Linear

〈평균수면시간〉

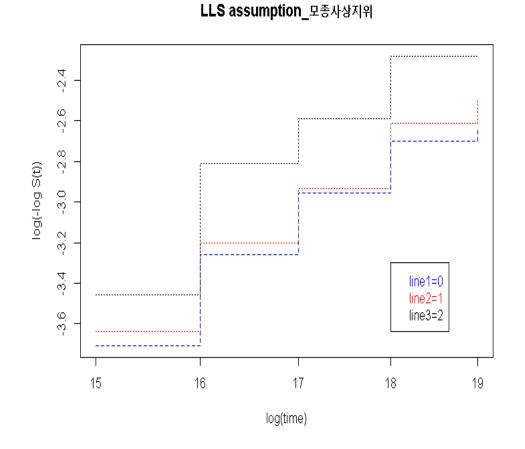
〈사회성〉

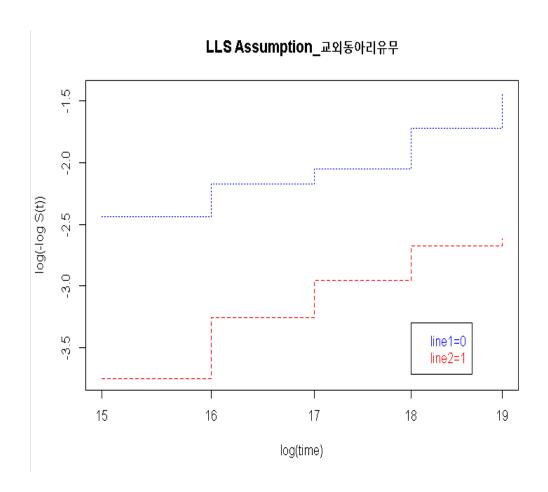




비례 위험 가정 그래프 확인







부록(변수 설명)

범주형 질문들

- I. 성별: ① 남자 ② 여자
- Ⅱ. (부,모)종사상 지위: ① 근로하지않음 ② 임금 가족 근로자 ③ 사업주
- Ⅲ. 성적만족도: ① 매우 만족한다 ② 만족하는 편이다. ③ 만족하지 않는 편이다.
 - ④전혀 만족하지 않는다.
- l. 가족 구성: ① (조부모)+부모+자녀 ② (조부모)+한 부모+자녀 ③ (한)조부모+자녀
- II. 교외동아리: ① 있다 ② 없다

수치형 질문들

- 연간소득
- 몸무게

- Damage : 비행 피해 종류 횟수
- 가출친구: 가출 친구 수
- 체험활동 : 경험한 체험활동 종류 수

부록(변수 설명)

리커트척도의 질문들 (점수가 높을수록 부정적임)

- I. 방임: 점수가 높을수록 부모님의 관심이 적은 것.
- Ⅱ. 학습: 점수가 높을수록 학교수업에 대한 적응도가 낮은 것.
- Ⅲ. 교우: 점수가 높을수록 학교에서 교우관계가 나쁜 것.
- IV. 교사: 점수가 높을수록 선생님과의 관계가 나쁜 것.
- V. 가족연락: 점수가 높을수록 가족과 연락을 적은 것.
- VI. 학대: 점수가 높을수록 부모님의 학대가 심한 것.