범주형 자료분석을 통한 고혈압의 주원인 분석

|  |  |
| --- | --- |
| 201932100 김수진: | r코드, 발표 |
| 201932125 손지은: | r코드, 레포트 |
| 201932139 이윤진: | r코드, ppt |
| 201935240 김지우: | r코드, 발표 |

저희는 알려진 고혈압의 주요원인들이 실제로 영향력을 미치는지, 영향력을 가진다면 어느 정도 영향력을 가지는지 알아보기 위해 이 주제를 선정하였습니다.

분석에 사용한 데이터는 공공데이터포털의 2020년도 ‘국민건강보험공단\_건강검진정보’를 이용했습니다. 약 100만명에 대한 기본정보와 건강검진결과가 들어있는 다음 데이터 중 시도코드를 이용해 서울에서 수집된 데이터 74208개 중 결측치를 제외한 74188개와 7개의 변수(sex, age, height, weight, waist, alcohol, blood)를 이용했습니다.

성별, 연령대, 키, 무게, 허리둘레, 음주여부를 설명변수로 설정하고 고혈압 여부에 따라 고혈압이 없다면 0, 있다면 1을 부여한 범주형 변수 blood를 종속변수로 설정하였습니다.

음주여부와 고혈압, 성별과 고혈압에 대한 이원분할표를 wald, relative risk, odds ratio 세가지 방법으로 분석하여 변수간 관계를 파악하고, height, weight, waist를 설명변수로, blood를 종속변수로 각각 로지스틱 회귀분석을 실시해 회귀선을 추정하고 모형의 유의성검정, 적합성검정을 실시하였습니다.

또, weight, age, blood 간의 다중 로지스틱 회귀분석을 실시하고, stepwise을 통해 최종적으로 어떤 모형이 가장 적합한 모형인지 추정했습니다.

1. **주제선정 이유**

고혈압은 수축기혈압이 140 mmHg 이상이거나 이완기혈압이 90 mmHg 이상일 때로 정의한다. 현재 우리나라 고혈압 환자는 약 1200만 명, 20세 이상 성인 중 유병률은 29%에 이르고 있다고 알려져 있다. 고혈압 자체가 사망원인이 되는 경우는 많지 않지만 고혈압을 치료하지 않을 경우 심근경색, 뇌졸중, 신부전 등 신체 각 부위에 여러 합병증이 발생할 수 있다. 고혈압의 주요원인은 나이, 가족력, 비만, 식습관 등이 있다. 이러한 각각의 원인들이 고혈압에 실제로 영향을 미치는지, 영향을 미친다면 어느정도로 영향을 미치는지 알아보고자 하였다. 또한 고혈압은 일반적으로 증상이 없기 때문에 혈압을 측정해보기 전까지는 진단이 되지 않는다. 이에 대한 방안으로, 로지스틱 회귀모형을 통해 직접 혈압을 재보지 않더라도 특정한 조건에서 고혈압일 확률을 구해볼 수 있을 것이라고 생각한다.

1. **데이터 설명**

**1) 분석 데이터 설명**

분석에 사용할 데이터는 2020년도 “국민건강보험공단\_건강검진정보”로 공공데이터포털에서 얻을 수 있었다. 건강검진정보란 국민건강보험의 직장가입자와 40세 이상의 피부양자, 세대주인 지역가입자와 40세 이상의 지역가입자의 일반검강검진 결과와 이들 일반건강검진 대상자 중에 만40세와 만66세에 도달한 이들이 받게 되는 생애전환기건강진단의 결과이다. 원본데이터에는 기준년도, 시도코드, 성별코드, 연령대코드, 신장, 체중, 허리둘레 등 100만명에 대한 기본정보와 건강검진결과가 들어있다.

텍스트, 캐비닛이(가) 표시된 사진

자동 생성된 설명

**2) 변수 전처리**

이중에서 시도코드를 이용해 서울에서 수집된 데이터 74208개 중 결측치를 제외한 74188개를 이용하기로 하였다. 또한 수축기혈압 변수와 이완기혈압 변수를 이용해 고혈압 여부를 나타내는 변수를 새로 만들었다. 마지막으로 5세 단위로 되어있던 연령대 코드의 범주를 재구성하여 10세 단위로 바꾸고 변수들의 이름을 간략하게 바꾸었다.

테이블이(가) 표시된 사진

자동 생성된 설명

**3) 변수 설명**

다음은 변수에 대한 설명이다.

|  |  |  |  |
| --- | --- | --- | --- |
| 변수명 | 변수 의미 | 변수명 | 변수 의미 |
| sex | 성별 | waist | 허리둘레 |
| age | 연령대 | alcohol | 음주 여부 |
| height | 키 | blood | 고혈압 여부 |
| weight | 몸무게 |  | |

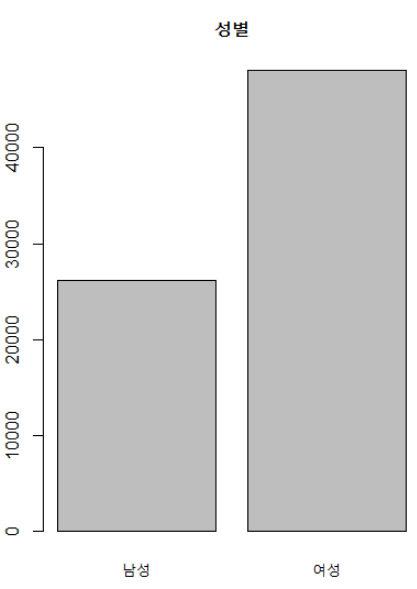
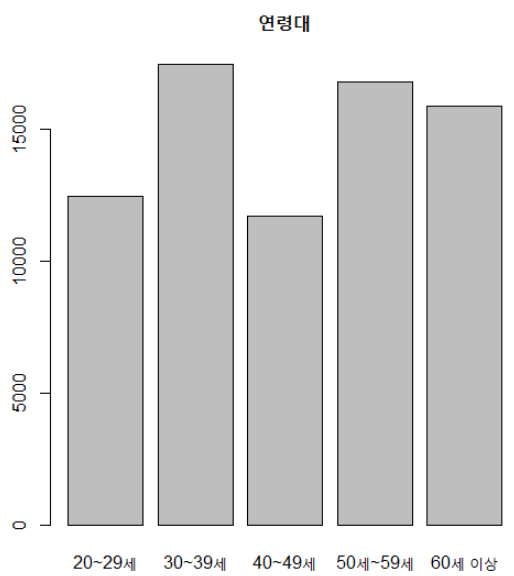
|  |  |
| --- | --- |
| sex: 성별 | |
| 코드 | 코드 의미 |
| 1 | 남성 |
| 2 | 여성 |

|  |  |
| --- | --- |
| age: 연령대 | |
| 코드 | 코드 의미 |
| 1 | 20~29세 |
| 2 | 30~39세 |
| 3 | 40~49세 |
| 4 | 50~59세 |
| 5 | 60세 이상 |

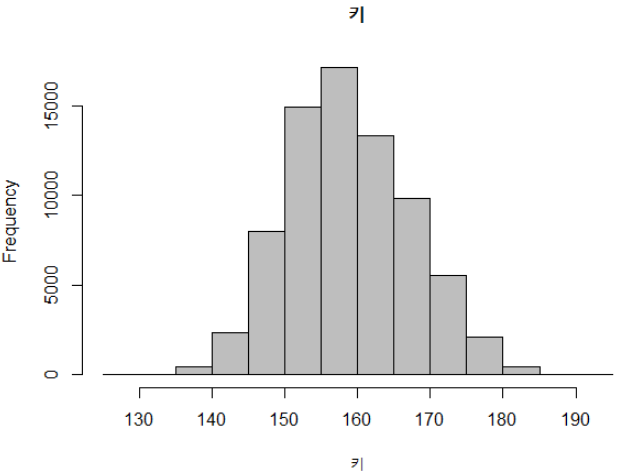
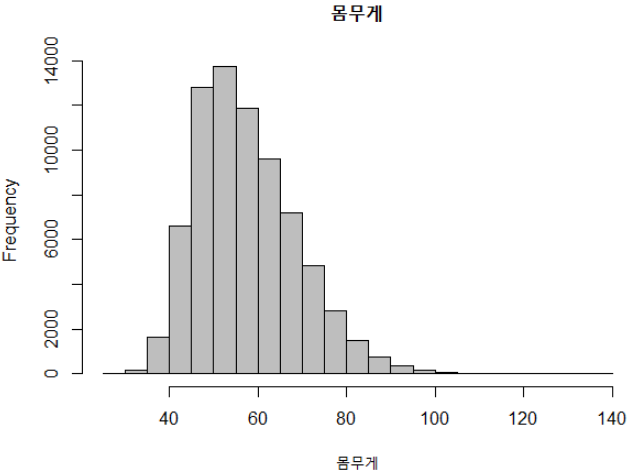
|  |  |
| --- | --- |
| alcohol: 음주 여부 | |
| 코드 | 코드 의미 |
| 0 | 마시지 않음 |
| 1 | 마신다 |

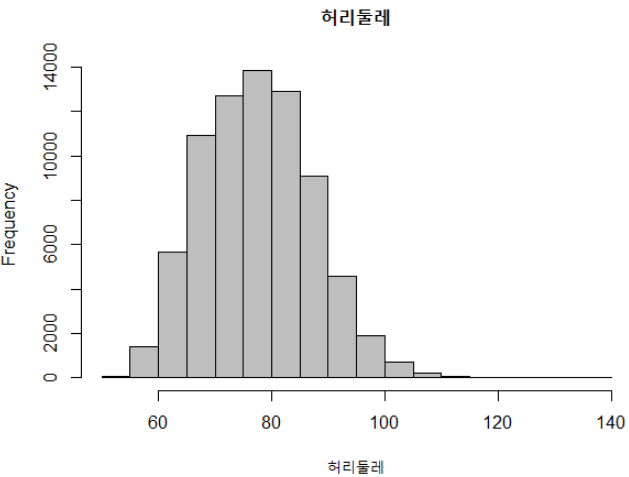
|  |  |
| --- | --- |
| blood: 고혈압 여부 | |
| 코드 | 코드 의미 |
| 0 | 고혈압 없음 |
| 1 | 고혈압 있음 |

1. **기초분석**

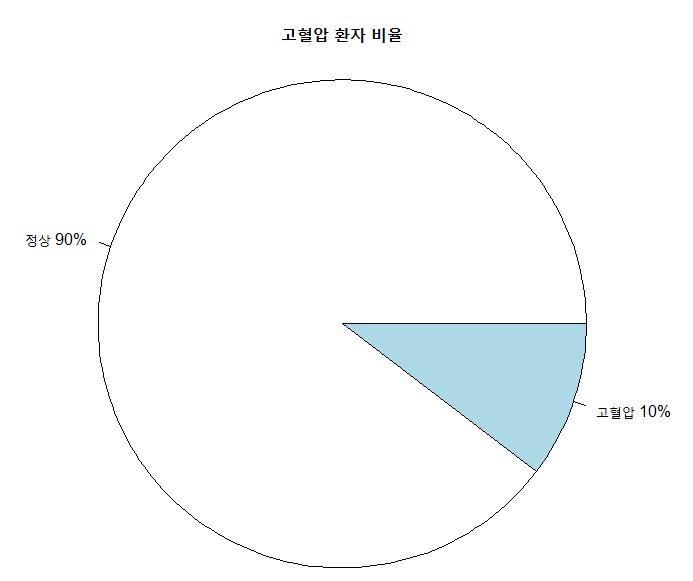
 

성별과 연령대에 대한 막대그래프를 통해 수집된 데이터에 성별은 남성보다 여성이 더 많으며, 연령대는 30대가 가장 많고 40대가 가장 적음을 확인하였다.



키와 몸무게, 허리둘레에 대한 히스토그램을 보면 키는 고르게 분포되어 있고, 몸무게와 허리둘레는 왼쪽으로 치우친 것을 확인할 수 있다.

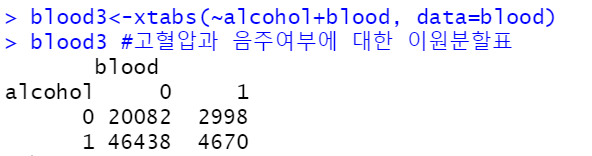


마지막으로 고혈압 여부에 대한 파이차트를 그려보았다. 결과는 그림과 같이 고혈압 환자는 전체 데이터에서 10%, 정상인 사람은 90%인 것을 볼 수 있다.

**4. 이원분할표**

각각 음주여부와 고혈압, 성별과 고혈압에 대한 이원분할표를 생성하여 신뢰구간을 확인해보았다.

텍스트이(가) 표시된 사진

자동 생성된 설명

**1) 음주여부와 고혈압 간의 관계**

**a. difference of proportion**

Wald의 방법을 이용했을 때, = P(고혈압O | 음주X) = 2998/(20082+2998) = 0.129896

= P(고혈압O | 음주O) = 4670/(46438+4670) = 0.09137513

= 0.03852089

과 의 차이가 약 3.85%로 음주를 하지 않았을 때가 음주를 했을 때보다 고혈압에 걸릴 확률이 3.85% 더 높다.

**b. relative risk**

/ = 1.421569

relative risk가 약 1.42로 음주하지 않았을 때 고혈압에 걸릴 확률이 음주했을 때보다 약 42% 높다.

**c. odds ratio**

음주를 하지 않았을 때, 고혈압에 걸릴 오즈값은 = 2998/20082 = 0.1492879이며, 음주를 했을 때, 고혈압에 걸릴 오즈값은 = 4670/46438 = 0.1005642이다. 두 집단의 오즈비( = / )를 구해보면 1.484504임을 확인할 수 있다. 따라서, 음주하지 않을 때 고혈압에 걸릴 오즈값이 음주할 때보다 약 48% 높다.

**d. confidence interval**

앞에서 구했던 difference of proportion, relative risk, odds ratio에 대한 신뢰구간을 구하였다.

텍스트이(가) 표시된 사진

자동 생성된 설명

먼저 difference of proportion에 대한 95% 신뢰구간을 구해보았다. 음주여부에 따른 두 집단의 고혈압에 걸릴 확률에 대한 95% 신뢰구간은 (0.034, 0.044)이다. 즉 음주를 할 때와 안할 때 고혈압에 걸릴 확률의 차이는 0.034에서 0.044정도임을 알 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

두번째로 relative risk의 95% 신뢰구간을 구해보았다. 그 결과 relative risk에 대한 95% 신뢰구간은 (1.36, 1.48)이다.

텍스트이(가) 표시된 사진

자동 생성된 설명

마지막으로 odds ratio에 대한 95% 신뢰구간을 구하였다. score 방식으로 측정한 오즈비를 이용해 구한 95% 신뢰구간은 (1.43, 1.73)이다.

**e.카이제곱 검정**

텍스트이(가) 표시된 사진

자동 생성된 설명

마지막으로 카이제곱 검정을 통해 음주여부와 고혈압 간에 연관성이 있는지 알아보았다. 으로 귀무가설 : ‘음주여부과 고혈압은 독립이다.’를 기각한다. 따라서 음주여부와 고혈압은 연관이 있음을 알 수 있다.

**2) 성별과 고혈압 간의 관계**

**a. difference of proportions(Wald)**

= P(고혈압O | 남자) = 3649/(22489+3649) = 0.1396052

= P(고혈압O | 여자) = 4670/(46438+4670) = 0.08317467

= 0.0564305

과 의 차이가 약 5.64%로 남자가 여자보다 고혈압에 걸릴 확률이 5.64% 더 높다.

**b. relative risk**

/ = 1.678458

relative risk가 약 1.68로 남자가 고혈압에 걸릴 확률이 여자보다 상대적으로 68% 높다.

**c.odds ratio**

성별이 남자일 때, 고혈압에 걸릴 오즈값은 = 3649/22489 = 0.1622571이다.

성별이 여자일 때, 고혈압에 걸릴 오즈값은 = 4019/44031 = 0.0912766이다.

두 집단의 오즈비를 구해보면 / = 1.777642임을 확인할 수 있다. 따라서, 성별이 남성일 때 고혈압에 걸릴 오즈값이 여성일 때보다 약 78% 높다.

**d. confidence interval**

앞에서 구했던 difference of proportion, relative risk, odds ratio에 대한 신뢰구간을 구하였다.

텍스트이(가) 표시된 사진

자동 생성된 설명

먼저 difference of proportion에 대한 95% 신뢰구간을 구해보았다. 성별에 따른 두 집단의 고혈압에 걸릴 확률에 대한 95% 신뢰구간은 (0.051, 0.061)이다. 즉 성별간 고혈압에 걸릴 확률의 차이는 0.051에서 0.061정도임을 알 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

두번째로 relative risk의 95% 신뢰구간을 구해보았다. 그 결과 relative risk에 대한 95% 신뢰구간은 (1.60, 1.74)이다.

텍스트이(가) 표시된 사진

자동 생성된 설명

마지막으로 odds ratio에 대한 95% 신뢰구간을 구했다. score 방식으로 측정한 오즈비의 95% 신뢰구간은 (1.75, 2.12)이다.

**e. 카이제곱 검정**

텍스트이(가) 표시된 사진

자동 생성된 설명

마지막으로 카이제곱 검정을 통해 성별과 고혈압 간에 연관이 있는지 알아보았다. 으로 귀무가설 : ‘성별과 고혈압은 독립이다.’를 기각한다. 따라서 성별과 고혈압은 연관이 있음을 알 수 있다.

**5. 연령대와 고혈압 간의 독립성 검정**

대부분의 질병이 그렇듯 고혈압도 나이가 많을수록 고혈압이 나타날 확률이 높아진다고 알려져 있다. 이번에는 독립성 검정을 통해 연령대와 고혈압 간에 실제로 연관이 있는지 독립성 검정을 통해 알아보고자 하였다.

먼저 독립성 검정을 위해 교차표를 그려보았다.

|  |  |  |
| --- | --- | --- |
|  | **blood (고혈압 여부)** | |
| **age (연령대)** | 0 (고혈압X) | 1 (고혈압O) |
| 1 (20~29세) | 12195 | 234 |
| 2 (30~39세) | 16730 | 698 |
| 3 (40~49세) | 10607 | 1099 |
| 4 (50~59세) | 14575 | 2189 |
| 5 (60세이상) | 12413 | 3448 |

다음으로 연령대와 고혈압 간에 연관이 있는지 알아보기 위해 chisq.test()함수를 이용하여 카이제곱 검정을 해보았다.

텍스트이(가) 표시된 사진

자동 생성된 설명

결과는 다음과 같이 에서 으로 “ 연령대와 고혈압은 독립이다”를 기각한다. 따라서 연령대와 고혈압은 서로 연관이 있음을 알 수 있다. 하지만 age(연령대) 변수는 순서형 변수인 것과 다르게 카이제곱 검정에서는 순서형 변수도 명목형 변수처럼 다루기 때문에 연령대의 순서에 대한 정보를 전혀 활용할 수 없다. 따라서 CMHtest()함수를 이용해 순서에 대한 정보도 활용할 수 있는 독립성 검정을 시행하였다.

텍스트이(가) 표시된 사진

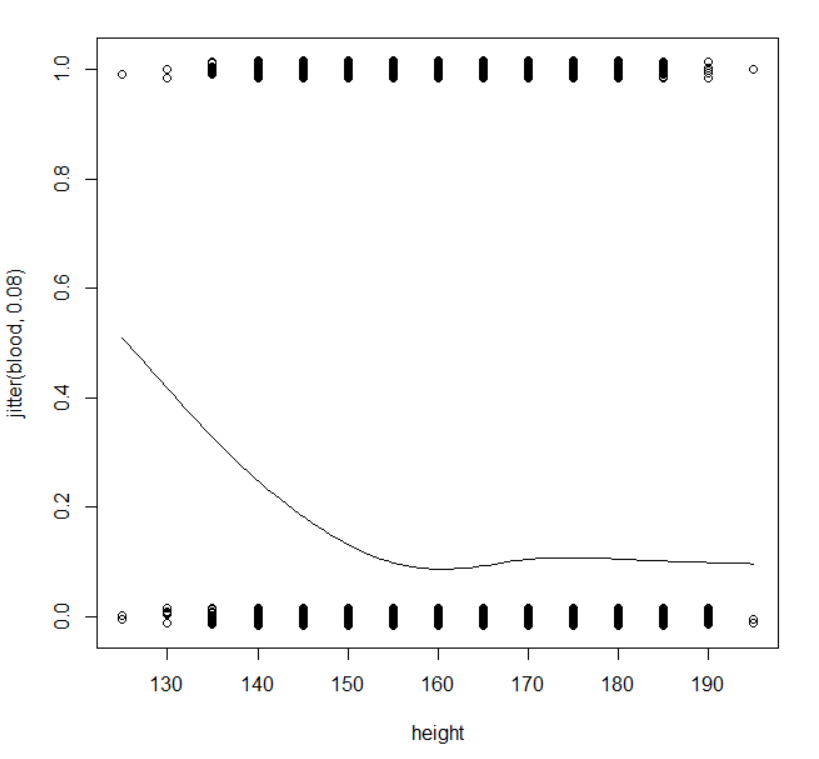
자동 생성된 설명

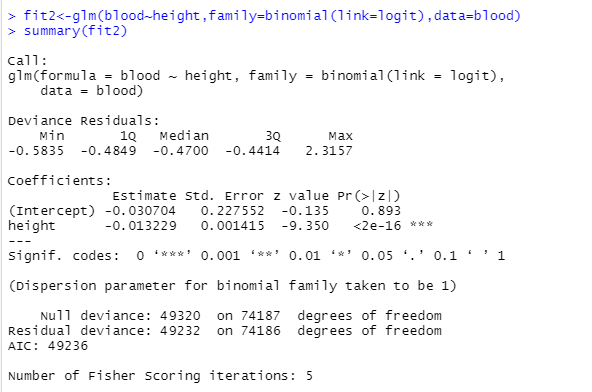
에서 가 0에 가깝게 나왔기 때문에 순서정보를 활용하는 독립성 검정에서도 “ 연령대와 고혈압은 독립이다”를 기각한다. 결과적으로 연령대와 고혈압 여부는 알려진 것과 같이 서로 연관이 있는 것으로 나타났다.

**6. 로지스틱 회귀분석**

**1) 설명변수를 height으로 둔 로지스틱 회귀분석**

**a. 로지스틱 회귀분석**

****



키와 고혈압 여부로 회귀선을 추정했을 때, 다음과 같은 결과를 얻을 수 있다.

이때, 이므로, 키가 클수록 고혈압을 얻을 확률이 증가한다고 볼 수 있다. 즉, 키가 5cm 증가하면 고혈압을 가질 오즈가 , 약 0.987배 증가한다.

**b. 모형의 유의성 검정**

1) wald:

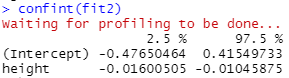
2) likelihood: with df=1

p-value가 매우 작으므로 귀무가설 ‘H0: 회귀모형은 유의하지 않다.’를 기각한다. 따라서 주어진 모형은 유의하다고 볼 수 있다.

**c. 모형의 적합성 검정**

로지스틱 회귀모형의 deviance는 49232이고 df=74186를 갖는다. 이때 우도비를 통해 deviance 차이를 구해봤을 때, 그 차이는 49320-49232=88이며 df=1을 갖는다. 이 경우 p-value가 매우 크므로 귀무가설 ‘H0: 모형이 적합하다.’을 기각할 수 없다. 따라서 주어진 모형은 적합하다. 두 deviance 간의 차이가 크므로 더 복잡한 모형인 M1=를 사용해야한다.

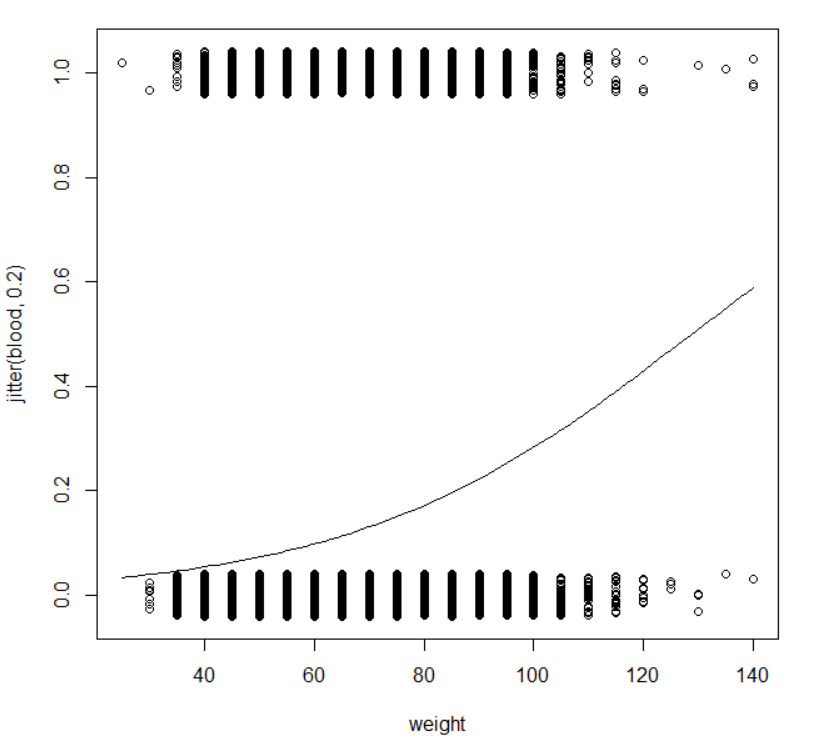
**d. confidence interval**

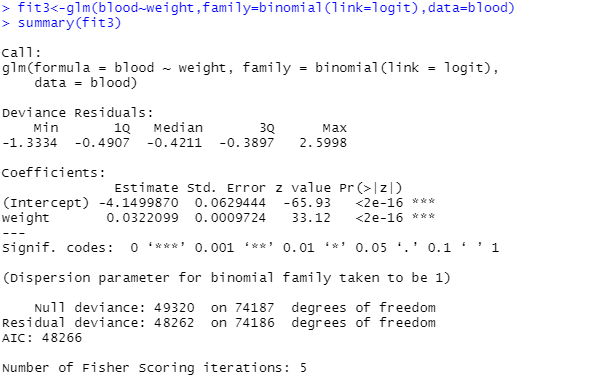


우도비를 이용한 height의 신뢰구간은 (-0.016, -0.010)이다.

**2) 설명변수를 weight으로 둔 로지스틱 회귀분석**

**a. 로지스틱 회귀분석**





몸무게와 고혈압 여부로 회귀선을 추정했을 때, 다음과 같은 결과를 얻을 수 있다.

>0이므로 몸무게가 많이 나갈수록, 고혈압을 얻을 확률이 증가한다. 즉, weight가 5kg 증가하면 고혈압을 가질 오즈가 배 증가한다.

**b. 모형의 유의성 검정**

1) wald:

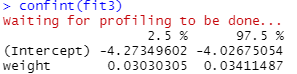
2)likelihood: with df=1

p-value가 매우 작으므로 귀무가설 ‘H0: 모형은 유의하지 않다.’을 기각한다. 따라서 주어진 모형은 유의하다고 볼 수 있다.

**c. 모형의 적합성 검정**

로지스틱 회귀모형의 deviance는 48262 이고 df=74186를 갖는다. 이때 우도비를 통해 deviance 차이를 구해봤을 때, 그 차이는 49320-48262=1058 이며 df=1을 갖는다. 이 경우 p-value가 매우 크므로 귀무가설 ‘H0: 모형이 적합하다.’을 기각할 수 없다. 따라서 주어진 모형은 적합하다. 두 deviance 간의 차이가 크므로 더 복잡한 모형인 M1=를 사용해야한다.

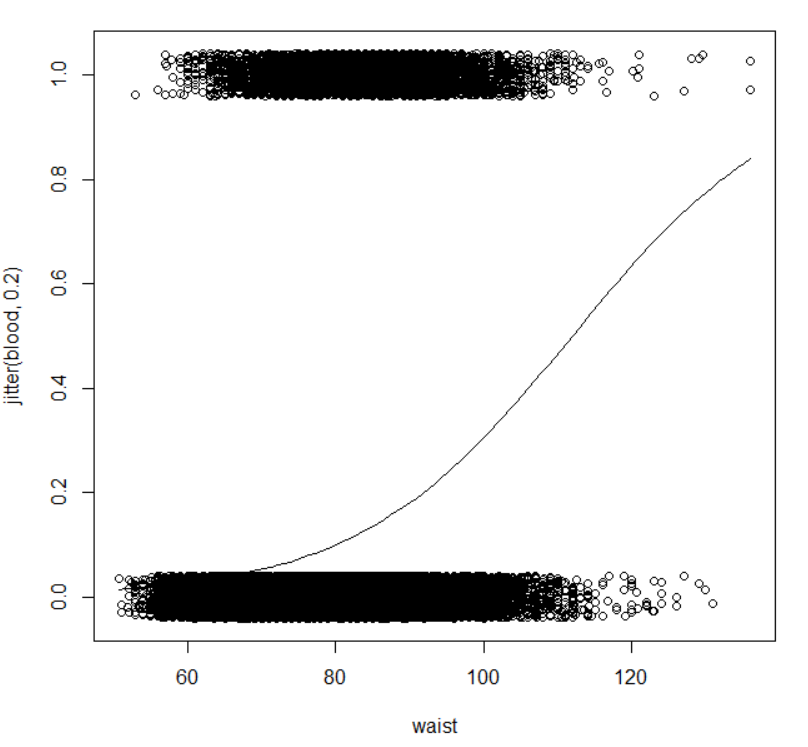
**d. 신뢰구간**

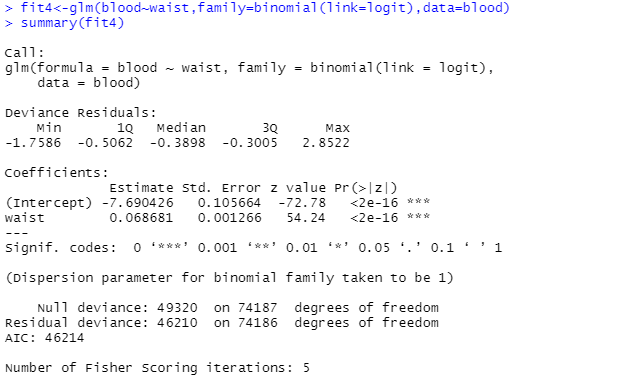


우도비를 이용한 신뢰구간은 (0.030, 0.034)이다.

**3) 설명변수를 waist으로 둔 로지스틱 회귀분석**

**a. 로지스틱 회귀분석**





허리둘레와 고혈압 여부로 회귀선을 추정했을 때, 다음과 같은 결과를 얻을 수 있다.

>0이므로, 허리둘레가 클수록, 고혈압을 얻을 확률이 증가한다.

exp()=exp(0.068681)=1.07109, waist가 1cm 증가하면 고혈압을 가질 오즈가 1.07109배 증가한다.

**b. 모형의 유의성 검정**

1) wald:

2)likelihood: with df=1

p-value가 매우 작으므로 귀무가설’ H0: 모형은 유의하지 않다.’를 기각한다. 따라서 주어진 모형은 유의하다고 볼 수 있다.

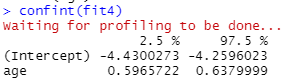
**c. 모형의 적합성 검정**

로지스틱 회귀모형의 deviance는 46210 이고 df=74186를 갖는다.

이때 우도비를 통해 deviance 차이를 구해봤을 때, 그 차이는 49320-46210=3110이며 df=1을 갖는다.

이 경우 p-value가 매우 크므로 귀무가설 ‘H0: 모형이 적합하다.’을 기각할 수 없다. 따라서 주어진 모형은 적합하다. 두 deviance 간의 차이가 크므로 더 복잡한 모형인 M1=를 사용해야한다.

**d. 신뢰구간**



우도비를 이용한 신뢰구간은 (0.597, 0.638)이다.

**7. 다중 로지스틱 회귀분석**

**1) weight, age와 blood간의 로지스틱 회귀분석**

텍스트이(가) 표시된 사진

자동 생성된 설명**a. 나이(age): 명목형 변수**

age가 고정일 때, weight가 5kg이 증가할 때마다 고혈압이 생길 오즈가 배 증가한다. age가 1일 때보다 2, 3, 4, 5일때 고혈압을 가질 확률이 증가하며, 증가할 확률은 5>4>3>2>1 순으로 높다. 각 연령대별 차이를 구하면 연령대가 60세 이상일 경우 고혈압이 생길 오즈가 20~29세일 때 보다 배 더 높고 30~39세 보다는 배 더 높다는 것을 알 수 있다.

weight의 경우 가 2e-16으로 아주 작은 값을 가지므로 유의하다. age의 경우도 마찬가지임을 알 수 있다. Anova검정을 통해 확인해주면 가 2.2e-16으로 age가 유의한 설명변수임을 확인할 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

**b. 나이(age): 순서형 변수**

텍스트이(가) 표시된 사진

자동 생성된 설명

같은 age일 때, 몸무게가 5kg 늘어날수록 고혈압이 생길 오즈가 배 증가하고, 연령대가 한 단계 높아지면 고혈압이 생길 오즈가 배 증가한다.

앞서 나이를 명목형 변수로 두었을 때와 같은 결과를 나타냄을 알 수 있다. 또한 P-value가 2e-16로 아주작은 값을 가지므로 weight는 유의한 설명변수임을 알 수 있다. 연령대별로 차이를 알아보면 20대보다 60대일 경우 배 증가함을 알 수 있다.

**c. 나이변수의 분류(순서형, 명목형)에 따른 모형의 적합성 비교**

텍스트이(가) 표시된 사진

자동 생성된 설명

p-value가 4.222e-06으로 아주 작으므로 두 모형 간에 유의미한 차이는 없다. 따라서 나이를 순서형 변수로 두었을 때 더 단순하고 해석이 용이한 모형이 나온다고 할 수 있다.

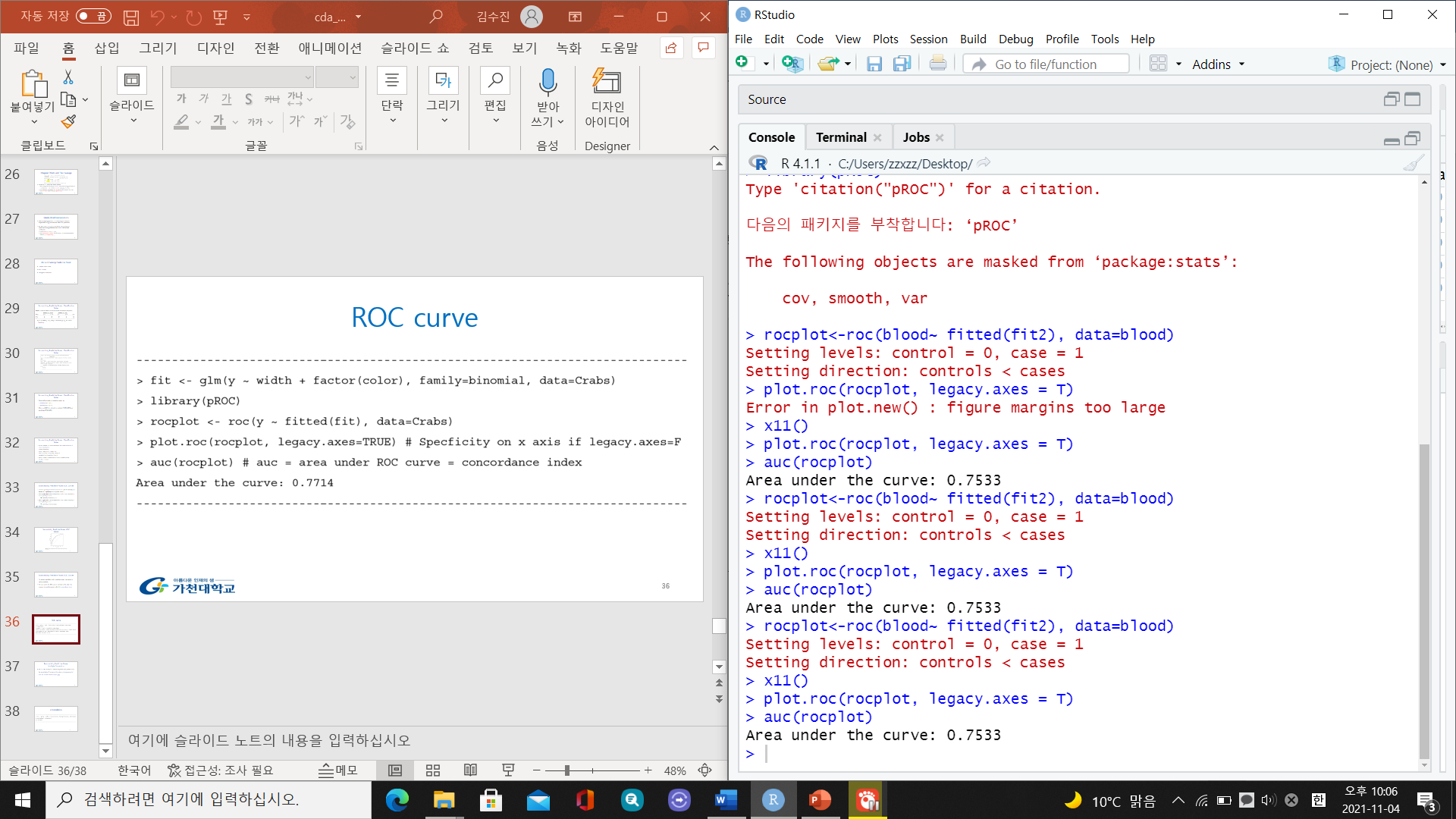
**d. 예측력을 통한 모형의 적합성 비교**

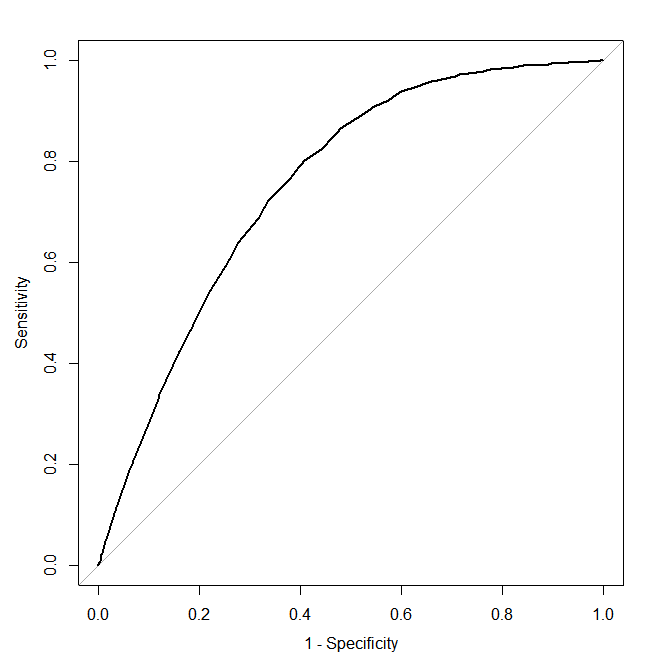
텍스트이(가) 표시된 사진

자동 생성된 설명

분할표를 만들어서 보았을 때, Sensitivity = 0.765, Specificity = 0.620, 총 예측력은 0.635이다.

다음으로 roc curve를 그려보았다.





면적이 0.7533으로 1에 가까운 면적을 가지고 있어 예측력이 높다는 것을 확인할 수 있다.

**8. 최적 모형 찾기**

**1) model selection**

Blood 데이터가 unbalanced 데이터이고 셀 값이 74188으로 매우 크므로 최대한 많은 변수를 사용해 모형을 적합시켜야 한다. 그러므로 데이터에 있는 모든 변수를 넣어 총 9개의 변수를 만들어 주자.

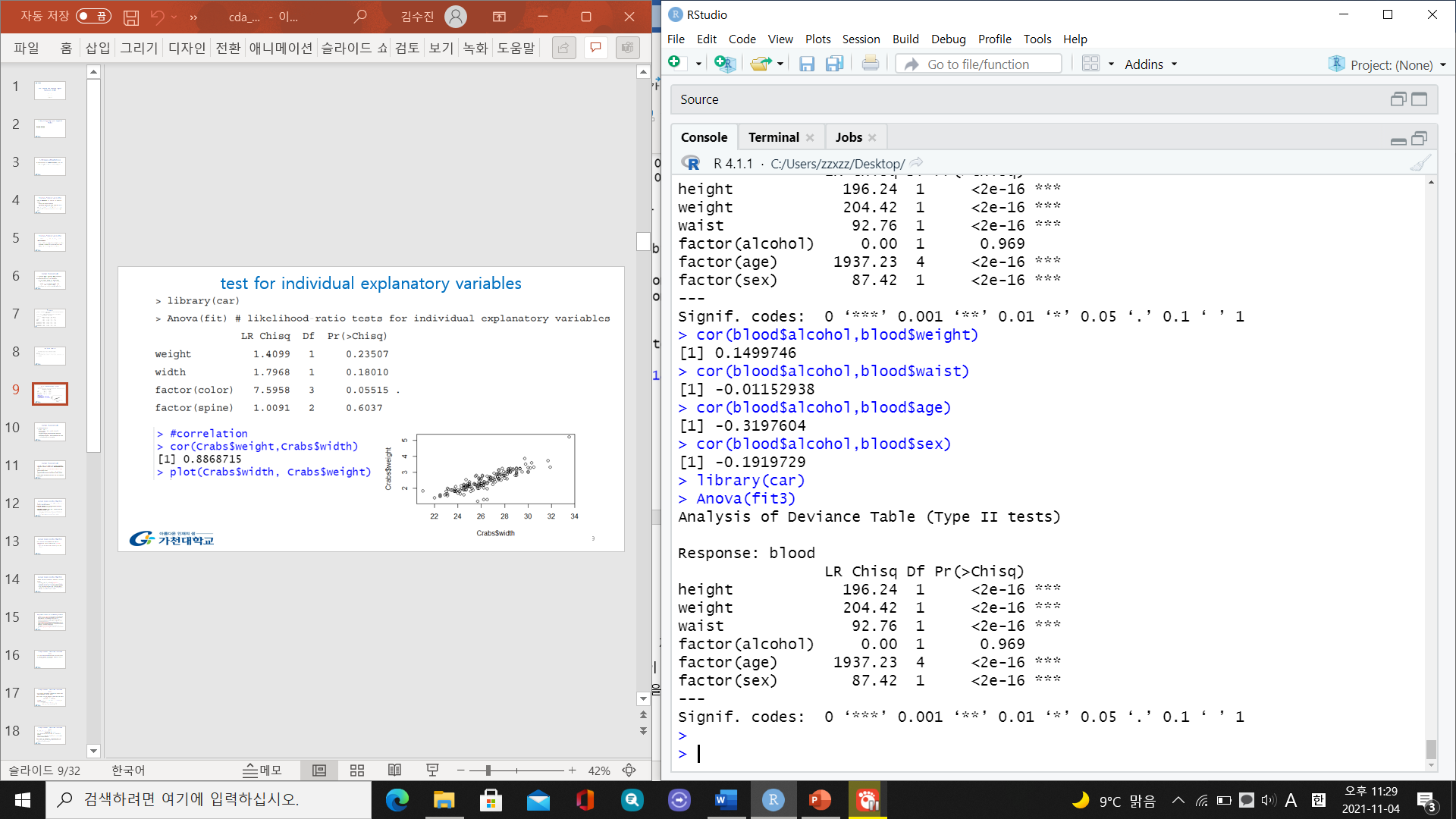
텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

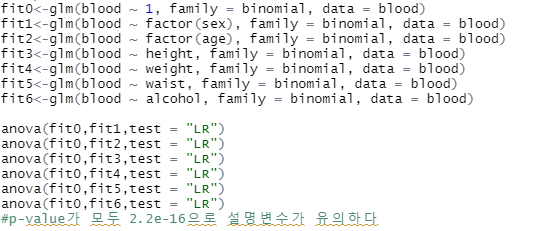
모든 값들의 p값을 보면 alcohol을 제외한 모든 변수가 유의함을 알 수 있다. 모든 베타가 0인지에 대한 검정을 해보면 P값이 0이 나오므로 귀무가설(===…==0)을 기각한다. 따라서 적어도 하나는 유의한 변수값을 가짐을 알 수 있다.



각각의 변수들로 우도비 검정을 해준 경우에도 alcohol만 적합하지 않다는 것을 알 수 있다. 그렇다면 다음으로 어떤 모형이 가장 적합한 모형인지 stepwise방법을 통해 알아보았다.

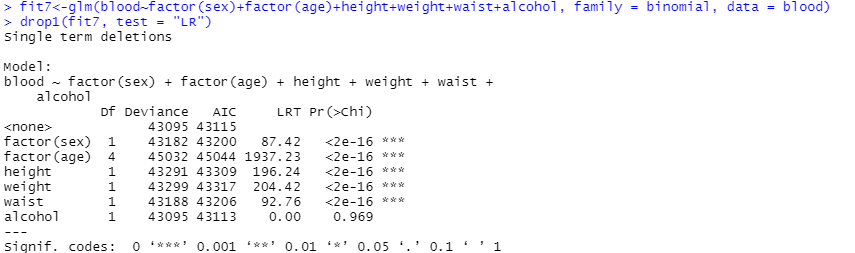
**2) stepwise**

Step1, 각각의 변수가 유의한지 알아보자

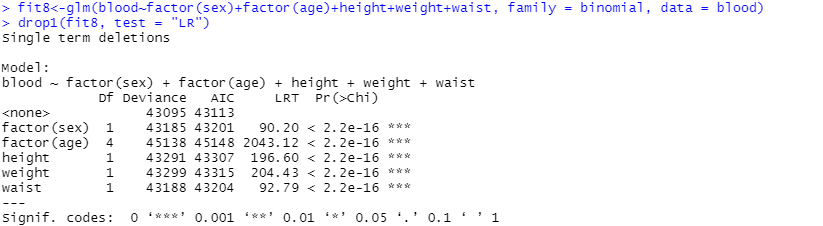
****

Anova 검정 결과 모든 변수의 p-value가 2.2e-16으로 아주 작게 나옴을 알 수 있다. 따라서 모든 변수를 더해주기로 한다.

Step2, drop1함수를 이용해 불필요한 변수를 제거해주었다.



Alcohol의 p값이 0.969로 0.05보다 크므로 제거해 주고 alcohol을 제외한 나머지 변수들이 유의한지 적용해보면 모두 유의함을 알 수 있다.



Step3, 더해줄 변수가 step2에서 제거해준 alcohol변수밖에 없으므로 건너뛴다.

Step4, 교호작용항을 더해주었다.

텍스트이(가) 표시된 사진

자동 생성된 설명

교호작용을 해주면 다음과 같은 결과가 나온다. Fit8과 Fit9가 유의하게 나오며 Fit11의 p값이 0.7068으로 유의하지 않게 나온다. 그러므로 weight:waist, height:waist을 더해준 fit12를 만들어줘 검정을 해보면 유의함을 알 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

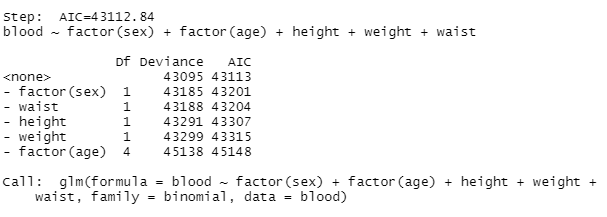
anova검정을 해주어 deviance의 차이를 알아보면 23.268으 차이로 Fit12의 모형이 가장 적절하게 나타난다.

텍스트이(가) 표시된 사진

자동 생성된 설명

다음으로 stepAIC와 bestglm을 이용해 stepwise를 통해 얻은 모형과 같은지 확인해보았다. stepAIC와 bestglm함수를 적용한 결과는 다음과 같다.





텍스트이(가) 표시된 사진

자동 생성된 설명

두개의 결과 교호작용을 해주기 전 모형인 Fit8의 모형이 가장 유의하게 나옴을 알 수 있다.

**3) 최종모형**

**a. 최종모형**

**텍스트이(가) 표시된 사진

자동 생성된 설명**

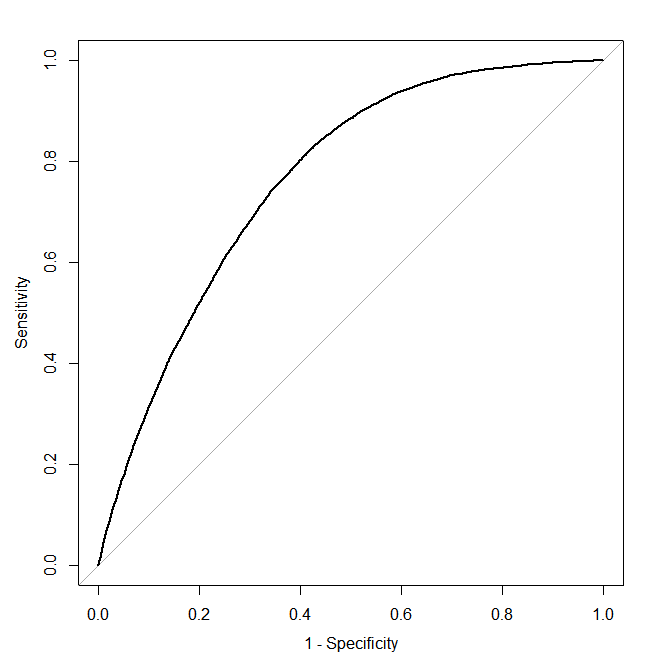
**b. 최종모형의 예측력**

**텍스트이(가) 표시된 사진

자동 생성된 설명**

최종모형의 예측력을 확인하기위해 분할표를 그려보았다. Sensitivity = 0.636, Specificity = 0.764, 총 예측력 = 0.649 이다.

다음으로 roc curve를 그려보았다.



텍스트이(가) 표시된 사진

자동 생성된 설명

면적이 0.7629으로 1에 가까운 면적을 가지고 있어 예측력이 높다는 것을 확인할 수 있다.

**9. 결론**

대체로 증상이 뚜렷한 다른 질병에 비해 일반적으로 측정해보기 전엔 진단이 힘든 고혈압은 여러 합병증을 유발하며, 최근 들어 그 위험성이 더욱 부각되고 있다.

그에 따라 본 연구에서는 성별, 키, 몸무게, 허리둘레, 음주여부와 고혈압의 상관성에 대해 알아보았다.우리는 어떤 조건에서 고혈압일 확률이 증가하는지에 중점을 두고 그 원인을 분할표 분석, 로지스틱 회귀분석, 다중 로지스틱 회귀분석 등을 통해 분석해보았다.

최종적으로 분할표분석을 통해 고혈압과 음주여부는 연관이 있으며 성별, 연령대와도 연관이 있음을 알 수 있었다. 그리고 다중 로지스틱 회귀분석을 통해 고혈압은 age5와 가장 큰 관련성을 가지는 것을 확인할 수 있었다. 또 고혈압은 키와 음의 상관을, 몸무게와 허리둘레와 양의 상관관계를 가진다. 따라서 연령대가 높을수록, 몸무게와 허리둘레가 클수록, 키가 작을수록, 성별이 남성일수록 고혈압에 걸릴 위험이 높다는 결론을 얻을 수 있었다.

그러므로 나이가 많을수록 특히 60대 이상일 경우에 고혈압의 위험성을 인지하고 정기적인 건강검진을 받는 것을 추천한다. 나이가 어릴 경우에도 식단조절과 꾸준한 운동으로 생활습관을 개선해 적정체중을 유지하는 것이 고혈압을 예방하는 것에 도움이 된다.

또한, 몸무게, 허리둘레와 양의 상관관계를, 키와 음의 상관관계를 갖는 것으로 보아 체내 체지방률을 변수로 추후 추가적으로 연구를 실시해보는 것도 좋을 것으로 생각한다.