

Finance Analytics

Chapter3. Linear Regression Model

Part 5. Model checking

권태연

한국외대 국제금융학과

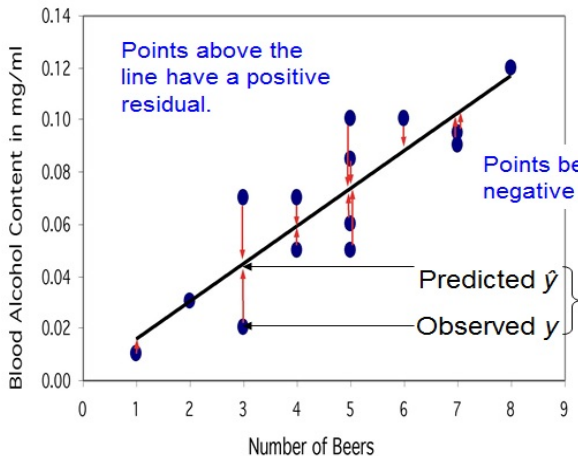
선형회귀모형의 가정에 대한 진단 - residual plot 을 이용

선형회귀모형의 가정

- 가정 1. 변수 Y와 X의 관계는 선형(Linear)이다. ✓
- 가정 2. X는 확률변수가 아닌 주어진 상수값이다.
- 가정 3. X값이 주어져 있을 때, 오차항의 평균은 0이다. ✓
- 가정 4. X값이 주어져 있을 때, 오차항의 분산은 σ^2 로 모든 개체 i 에 대해 동일하다 ✓
- 가정 5. 서로 다른 개체간 오차항들은 상관되어 있지 않다. ✓
- 가정 6. X변수들이 여러개 있을때, X변수들 사이에는 선형관계가 없다. = 다중공선성(Multicollinearity) 문제가심하지 않다 로 완화
- 가정 7. 모형 설정 오류가 없음
- 가정 8. 오차항은 정규분포를 따름을 가정 ✓

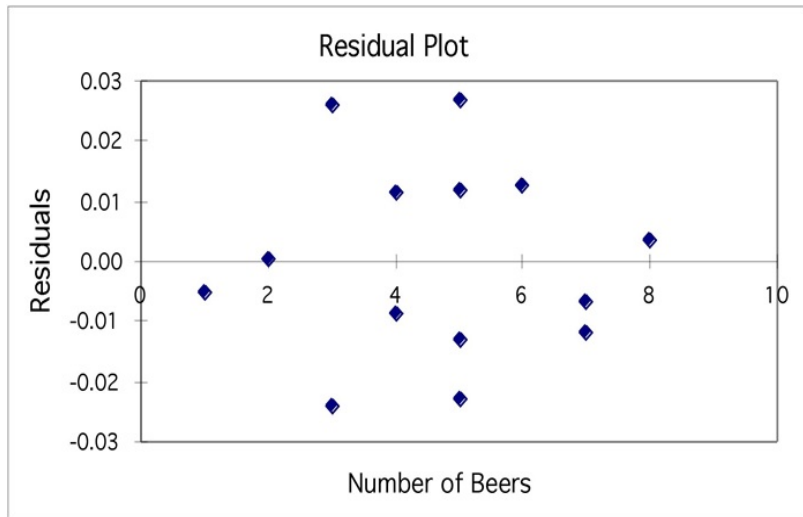
Residual

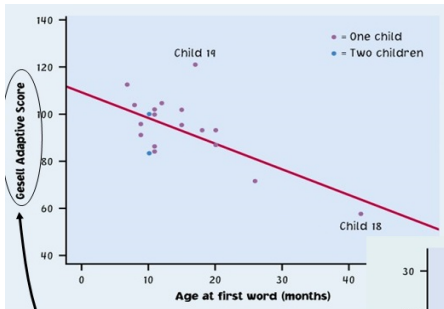
Blood Alcohol Content as a function of Number of Beers



Residual plot

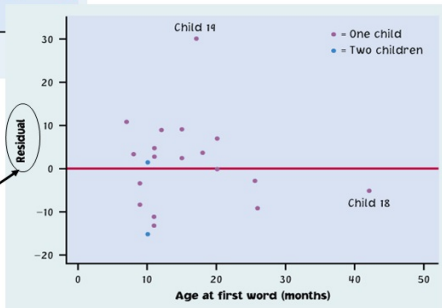
Residual plot : scatter plot of residuals versus x values





The x-axis in a residual plot is the same as on the scatterplot.

Only the y-axis is different.



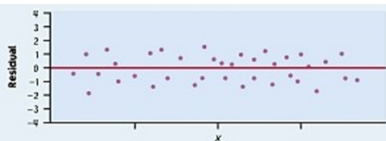
Ideal Residual plot

- Residuals are close to zero for all values of x
- Residuals are randomly distributed with mean 0.
- Have no pattern when plotted against the explanatory variable x (random scatter)
- Variance (spread) of the residuals is constant over the x values.

Ideal Residual plot

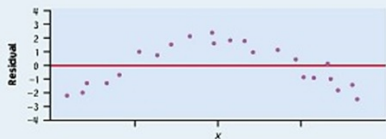
- Residuals are close to zero for all values of x
→ 가정 3
- Residuals are randomly distributed with mean 0.
→ 가정 3
- Have no pattern when plotted against the explanatory variable x (random scatter)
→ 가정 1, 가정 5
- Variance (spread) of the residuals is constant over the x values.
→ 가정 4

Residual plot



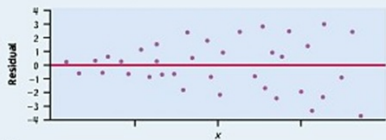
(a)

Residuals are randomly scattered—good!



(b)

Curved pattern—means the relationship you are looking at is not linear.



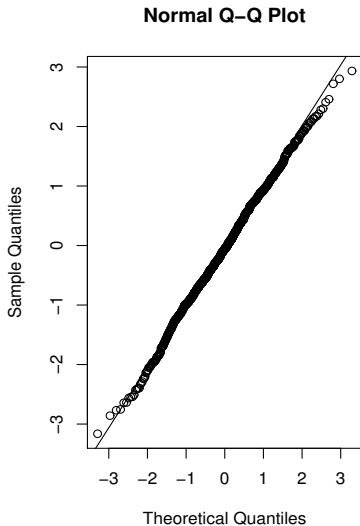
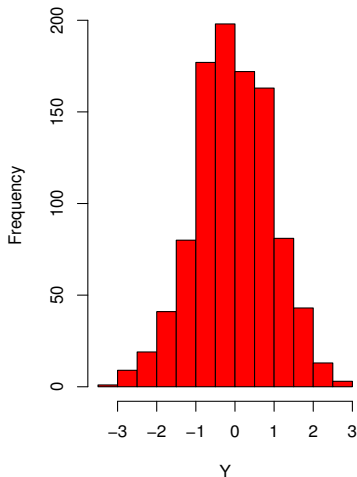
(c)

A change in variability across a plot is a warning sign. You need to find out why it is, and remember that predictions made in areas of larger variability will not be as good.

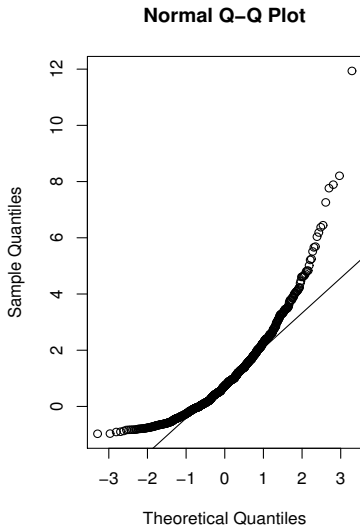
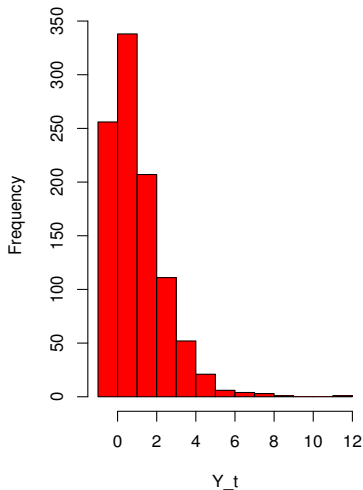
가정의 확인

- 잔차 plot을 통해 확인할 수 있는 가정은?
 - : 오차항과 관련한 가정들..
 - : 가정 3, 가정4, + 가정 1 가정 5
 - : 가정 8. 정규분포에 대한 확인은?
 - : → 잔차의 히스토그램 or 잔차에 대한 Normal q-q plot

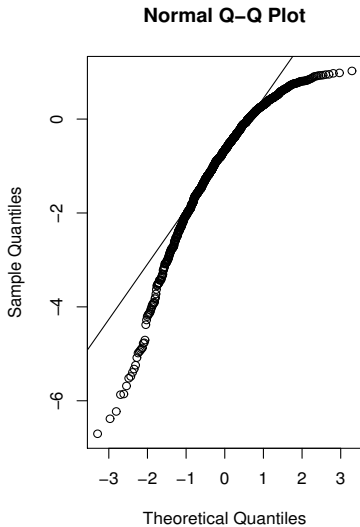
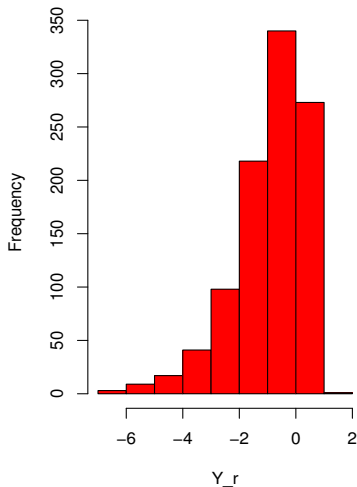
Skewness and Normal Quantile Plot 1. Symmetric



Skewness and Normal Quantile Plot 2 . Right Skewed



Skewness and Normal Quantile Plot 3. Left Skewed



Normal Quantile Quantile Plot of Residuals (잔차의 Normal qqplot)

가정8 확인: 잔차의 Histogram과 Normal qq plot이 Plot1과 같음을 확인

R을 이용한 잔차플롯

- wage education, 단순회귀모형을 적합하고, 시간당 임금함수 예제에서 잔차 plot을 그려보자.

```
> model2 <- lm(wage ~ education, data = data)
> names(model2)
```

- lm()함수 실행 결과를 저장한 R객체에 "residuals" 라는 이름으로 잔차가 저장되어 있다.
- "fitted.value" 는 \hat{y}

다중회귀모형에서의 잔차플롯

- Residual plot in multiple regression model
: scatter plot of residual versus \hat{y}
- 임금함수 예제에서 education과 experience를 사용하여
다중회귀모형을 적합하고 잔차플롯을 그려보자.

R을 이용한 qqplot

- wage education, 단순회귀모형을 적합하고, 시간당 임금함수 예제에서 잔차 plot을 그려보자.

```
> model2 <- lm(wage ~ education, data = data)
> names(model2)
```

- lm() 함수 실행 결과를 저장한 R객체에 "residuals" 라는 이름으로 잔차가 저장되어 있다.
- "residuals"에 대한 qqplot
- qqnorm(), qqline() 함수를 사용한다.