* 참고

- 오차 error
- 잔차 residual
- 편차 deviation
- 표준편차 standard deviation
- 표준오차 standard error

- RMSE root-mean-square error = standard error of regression, 회귀의 표준오차

VII. 회귀분석 심화

1. 회귀분석의 추정

- Basic format: $y_i = \beta_0 + \beta_1 x_i + u_i$
- SSE (sum of the squared residuals)=$\sum \varepsilon_i{}^2$=$\sum(y_i$-$\hat{y}_i)^2$=$\sum(y_i$-$(\beta_0 + \beta_1 x_i))^2$

1. 회귀분석의 추정

- Random variable: $y_i = \beta_0 + \beta_1 x_i + u_i$

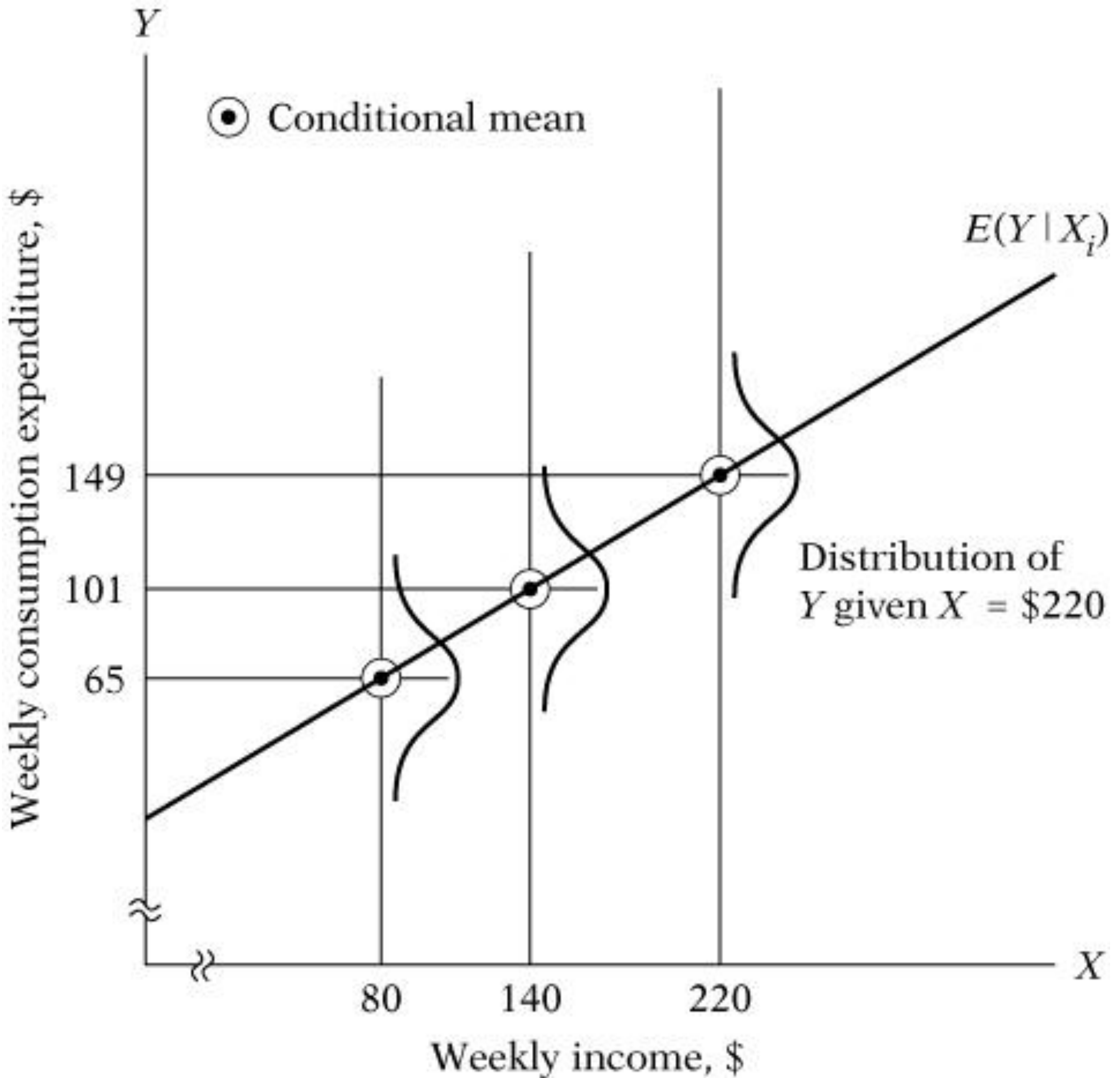
1. 회귀분석의 추정

- E(Y|X): X에 대한 조건부 평균/기댓값 conditional mean
- Unconditional mean

TABLE 2.1
Weekly Family Income X, \$

$Y \downarrow$ \ $X \rightarrow$	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y, \$	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y, $E(Y \mid X)$	65	77	89	101	113	125	137	149	161	173

1. 회귀분석의 추정

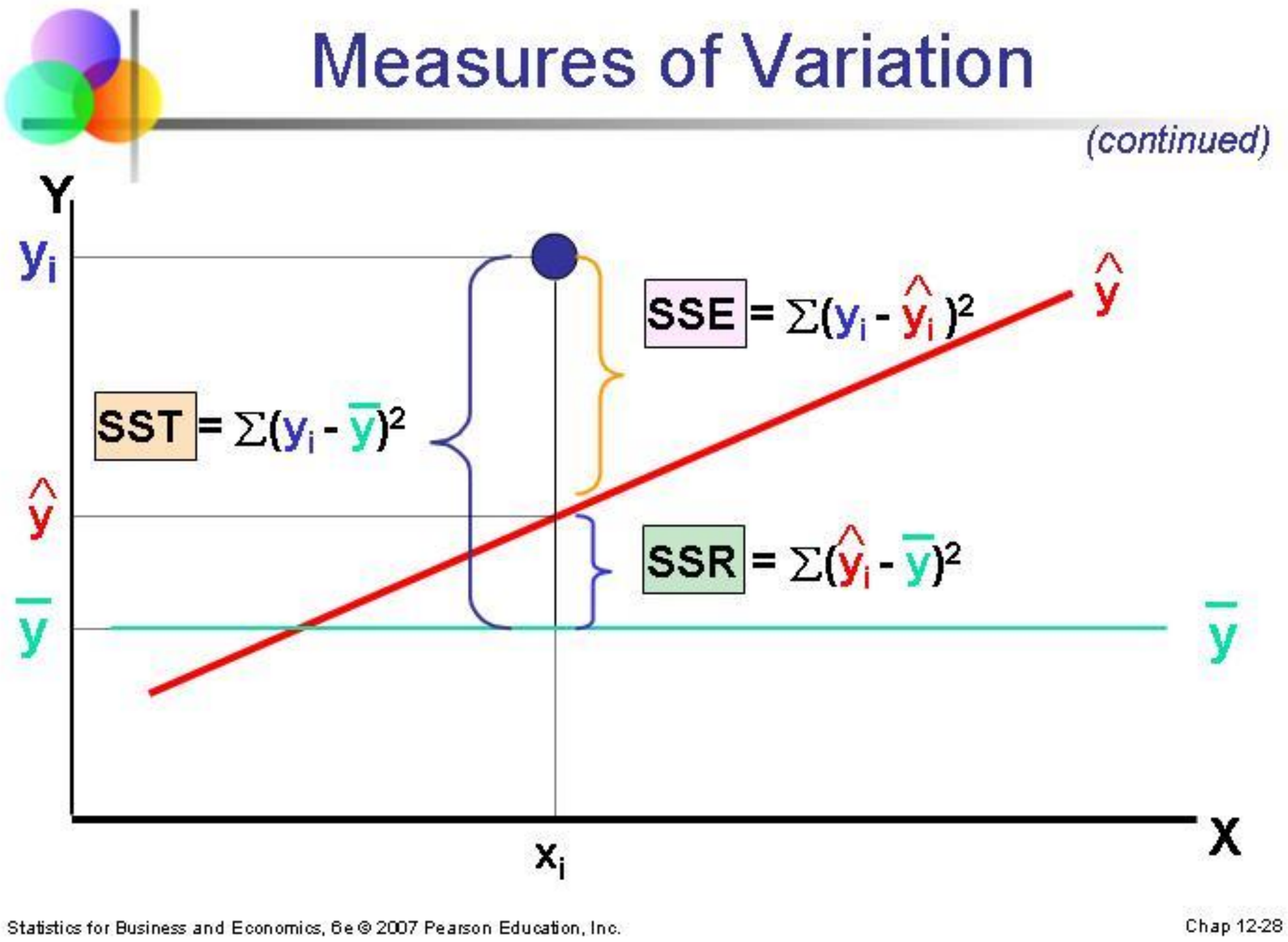


Y
Conditional mean
E(Y | X_i)
149
101
65
Distribution of
Y given X = $220
X
80
140
220
Weekly income, $
Weekly consumption expenditure, $

2. 적합도$_{\text{goodness of fit}}$의 측정

- 총제곱합 total sum of squares $SST = \sum_{i=1}^{n}(y_i - \bar{Y})^2$
- 오차제곱합 error sum of squares $SSE = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_i))^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$
- 회귀제곱합 regression sum of squares $SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2 = b_1^2 \sum_{i=1}^{n}(x_i - \bar{X})^2$

2. 적합도$_{\text{goodness of fit}}$의 측정



2. 적합도goodness of fit의 측정

- R^2=SSR/SST = 1-SSE/SST
  = 회귀제곱합/총제곱합 = 1-오차제곱합/총제곱합

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \overline{Y})^2 = b_1^2 \sum_{i=1}^{n} (x_i - \overline{X})^2$$

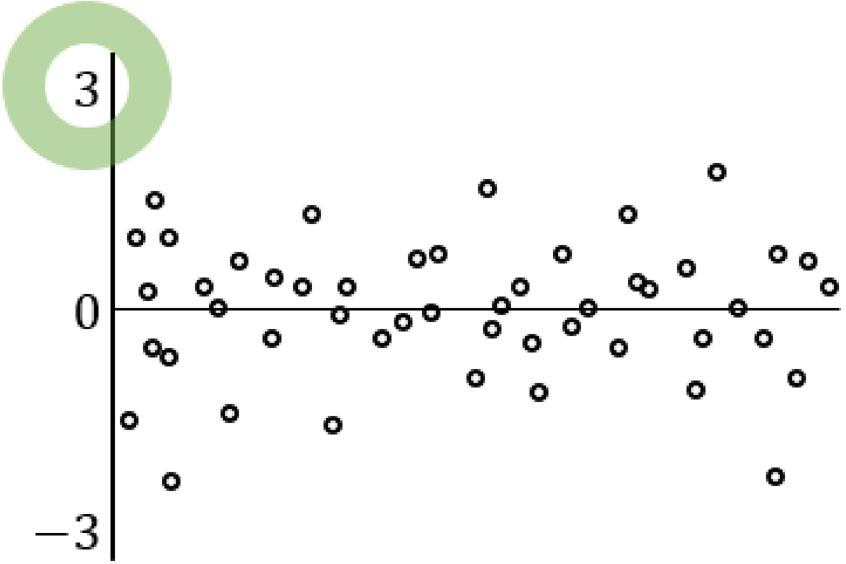
3. 회귀분석의 가정

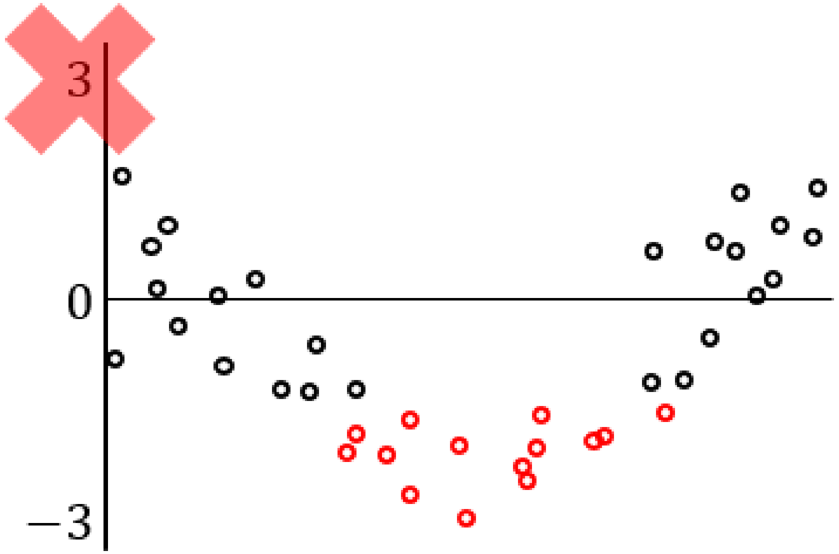
- OLS의 추정치 estimator
  - 불편성(unbiasedness)
  - 효율성(efficiency)
  - 선형성(linearity)
  - 일관성(consistency)

  - Best Linear Unbiased Estimator: BLUE

3. 회귀분석의 가정

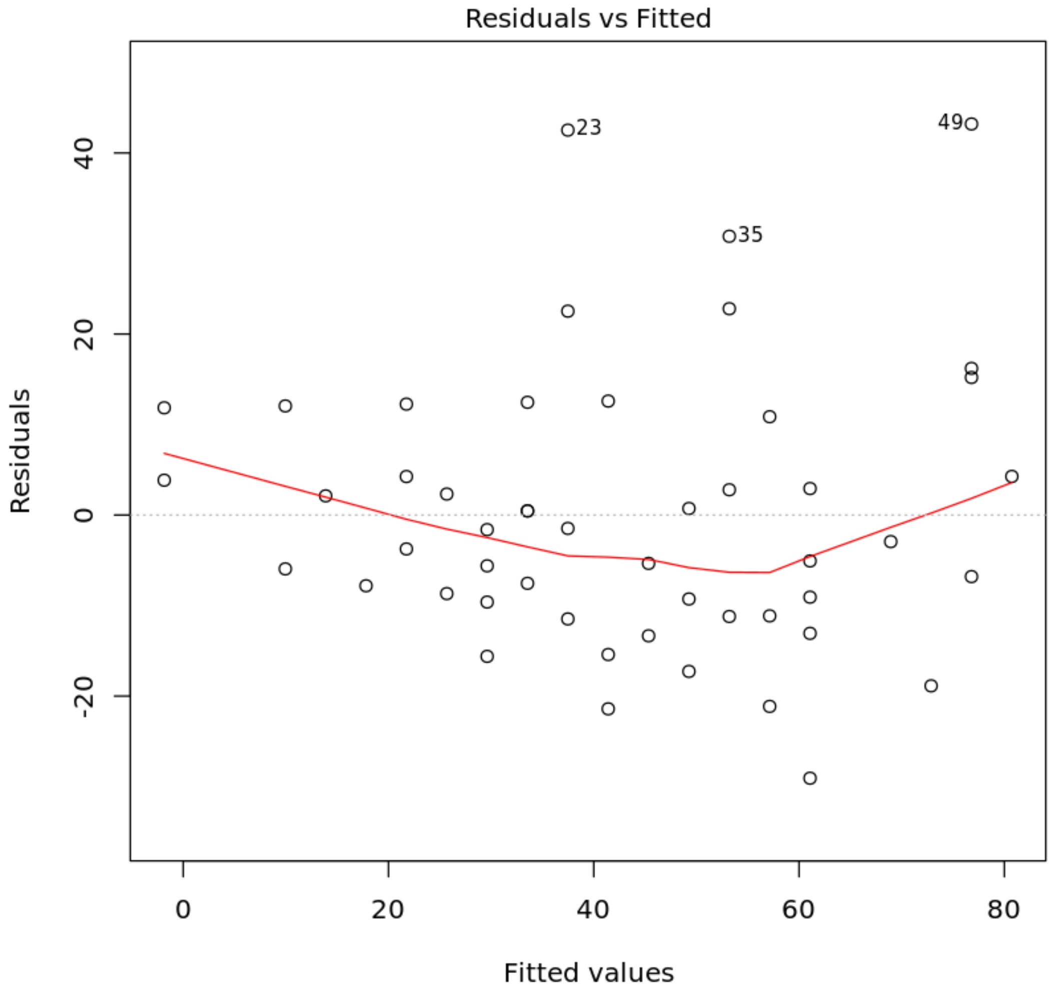
1) 선형성 $y_i = \beta_0 + \beta_1 x_i + u_i$





3. 회귀분석의 가정

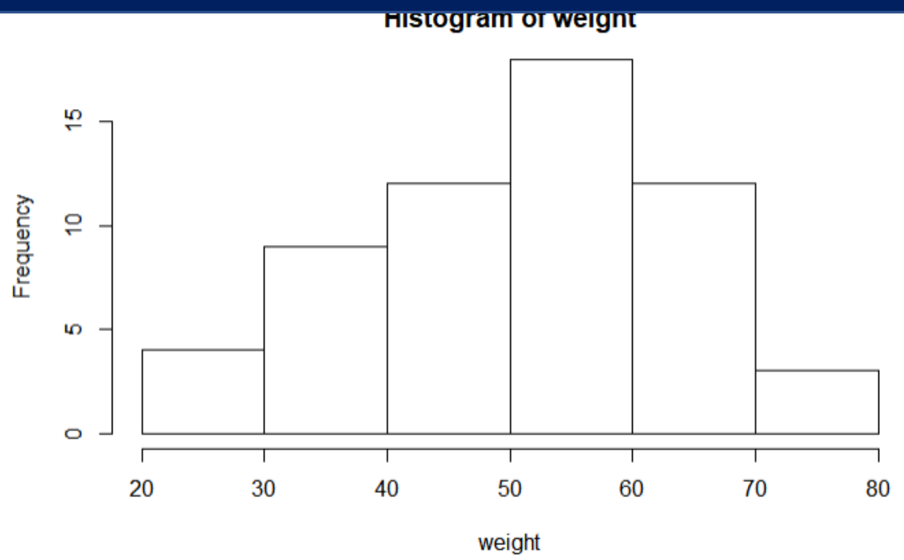
1) 선형성 $y_i = \beta_0 + \beta_1 x_i + u_i$

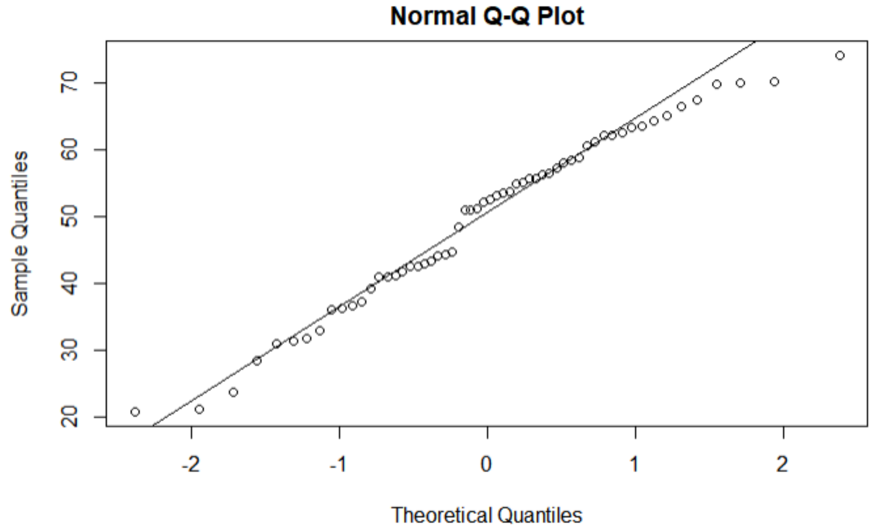


https://mindscale.kr/course/basic-stat-r/residuals

3. 회귀분석의 가정

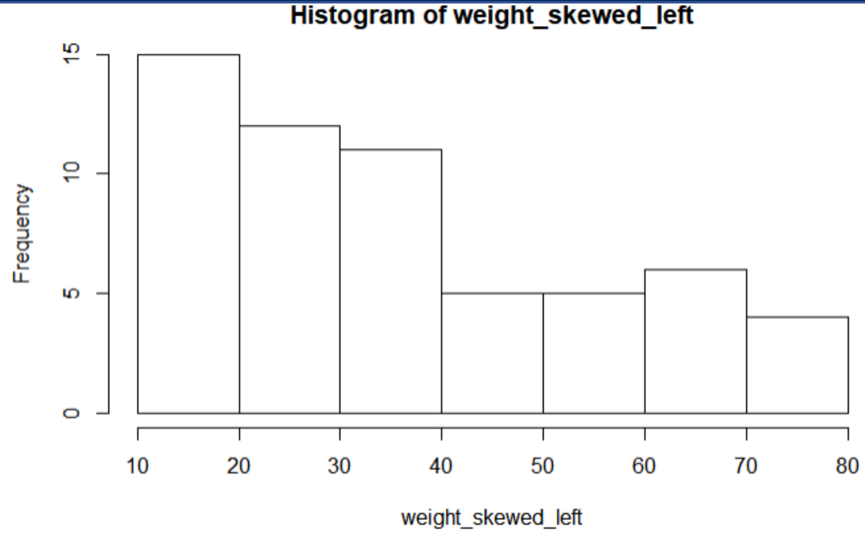
2) 정규성 $\varepsilon_i \sim N(0, \delta^2)$

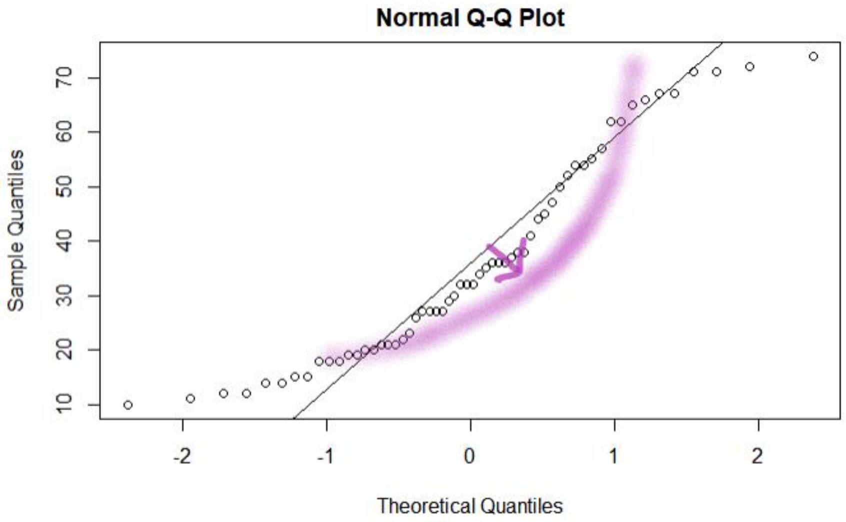




3. 회귀분석의 가정

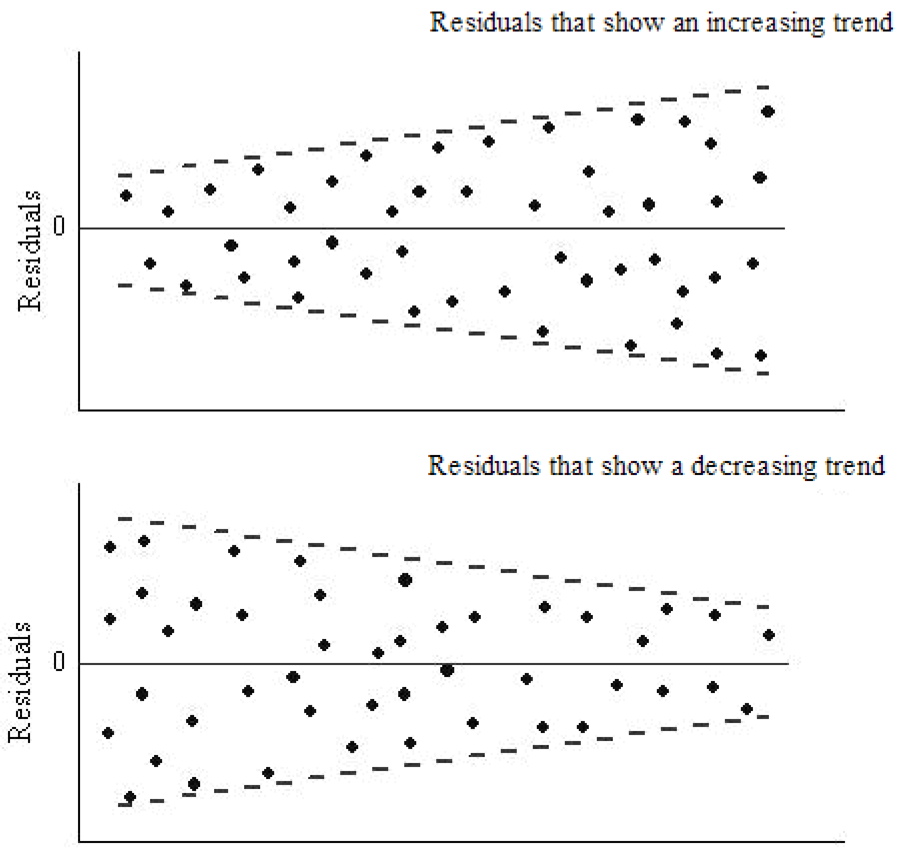
2) 정규성 $\varepsilon_i \sim N(0, \delta^2)$





3. 회귀분석의 가정

3) 등분산성 $var(\varepsilon_i)=\delta^2$



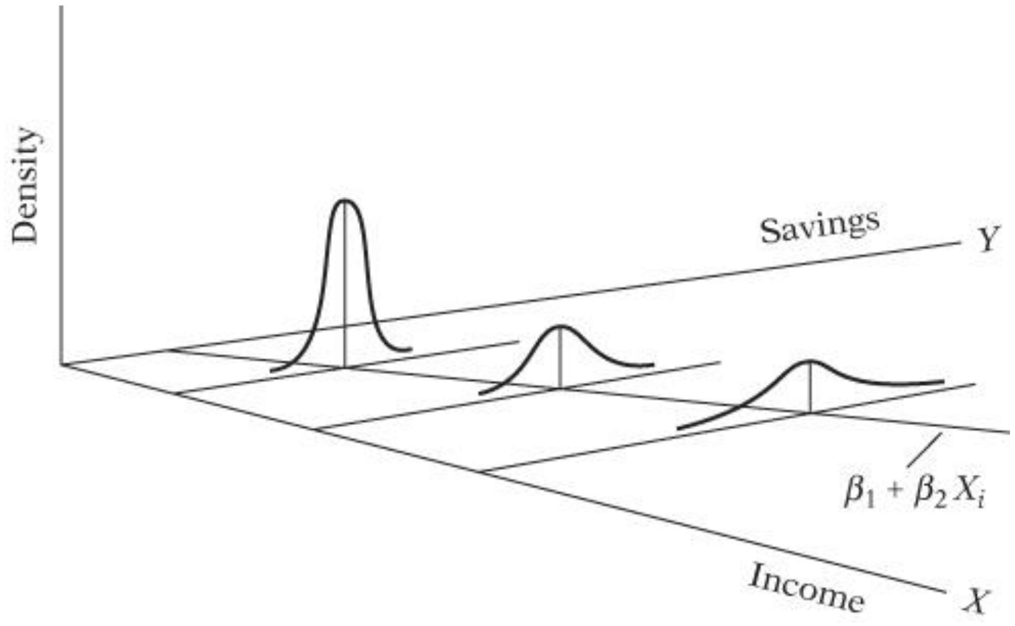
http://piramvill2.org/wp/?p=3137

3. 회귀분석의 가정

3) 등분산성 var(ε_i)=δ^2

- Breusch-Pagan test

- weighted least square
- Transformation of variables



3. 회귀분석의 가정

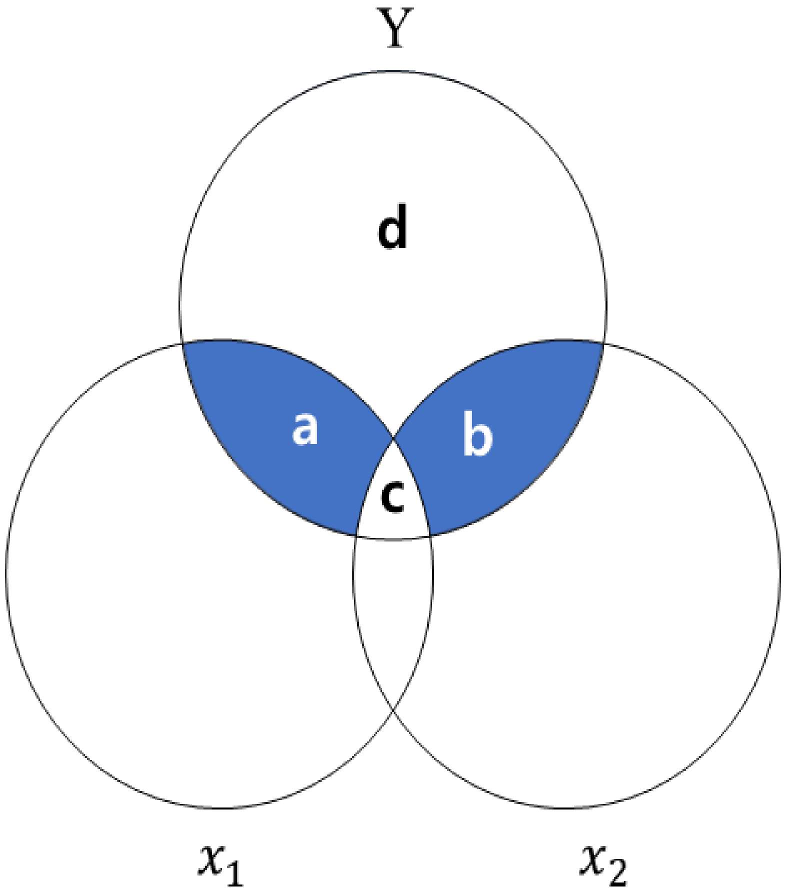
4) 잔차의 독립성 E(ε_i)=0

3. 회귀분석의 가정

5) 비상관성 Cov($\varepsilon_i, \varepsilon_j$)=0

- Durbin-watson test
- generalized least square

4. 다중공선성 multicollinearity



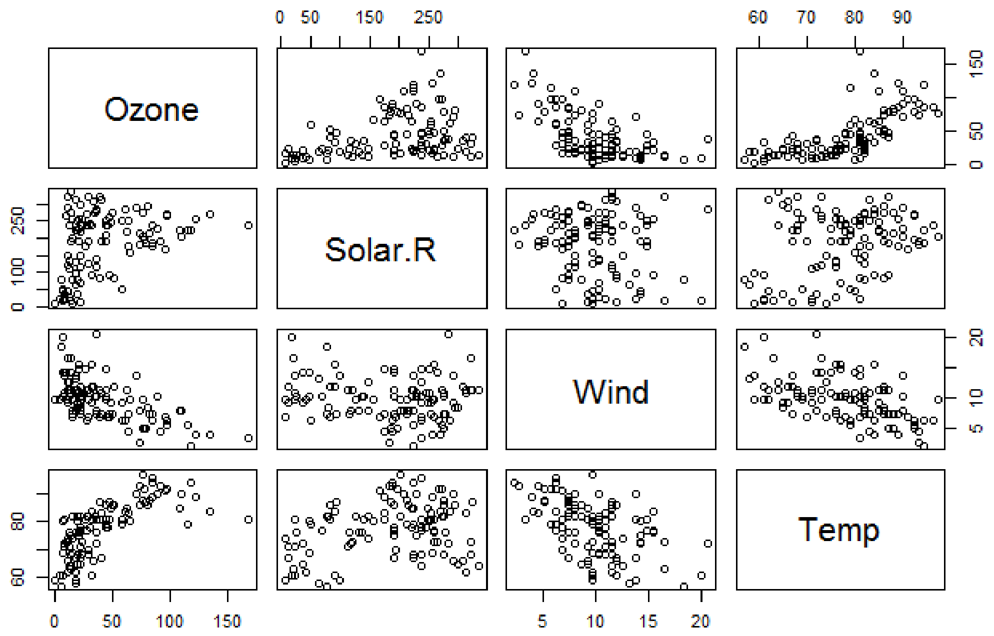
a: x_1에 의해 설명되는 고유 분산

b: x_2에 의해 설명되는 고유 분산

c: x_1, x_2에 의해 설명되는 공유분산

d: x_1, x_2 의해 설명되지 않는 분산

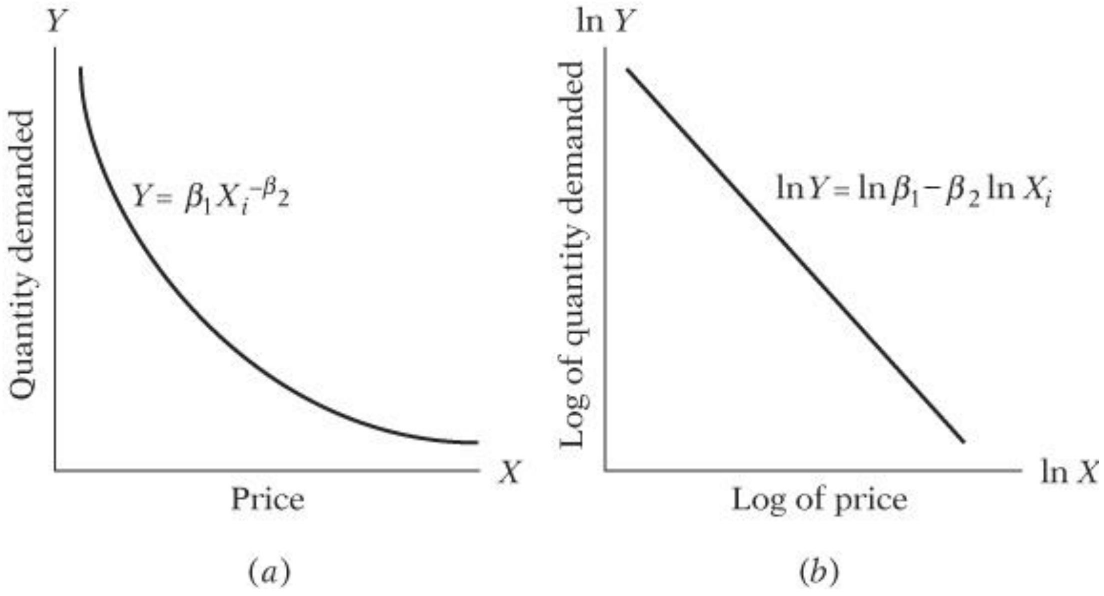
4. 다중공선성 multicollinearity



- Tolerance 공차한계 = $1-R^2$
- VIF 분산팽창계수 = $1/ 1-R^2$

5. 비선형관계

- Log-log model: 탄력성 모델
- $Y=b1X^{b2}e^{u}$
- $\ln(Y)=\ln(b1)+b2\ln(X)+u$



5. 비선형관계

- Log-lin model: growth rate model
- $Yt=Y0(1+r)^t$
- $\ln(Yt)=\ln(Y0)+t\ln(1+r)$.

5. 비선형관계

- Lin-log model: absolute change model
- $Y=b1+b2\ln(X)+u$