11 장

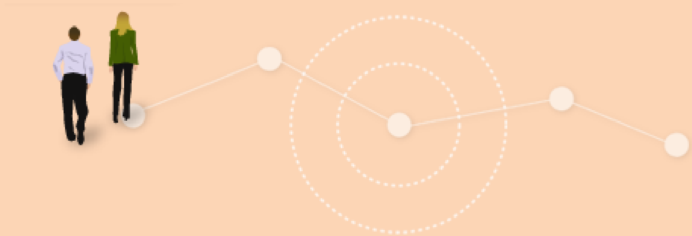
목 차

01 단순회귀분석

02 다중회귀분석

01 단순회귀분석

:: Keywords 단순회귀분석 | 독립변수 | 종속변수 | 최소자승법



회귀분석

광고비와 매출액 간에 인과관계를 알 수 있을까?

[표 11–1]은 10장에서 살펴본 A주식회사의 매출액과 광고비에 대한 자료이다. 상관관계를 살펴본 결과, A주식회사의 광고비와 매출액은 양의 상관관계가 있는 것으로 나타났다. 그렇다면 과연 2016년도에 광고비에 따른 매출액을 미리 예상해볼 수 있을까?

[표 11–1] **광고비와 매출액**

연도	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
광고비	13	8	10	15	12	15	14	15	17	19	20	21	22	21	25
매출액	94	70	90	100	95	100	85	95	105	105	110	105	104	105	121

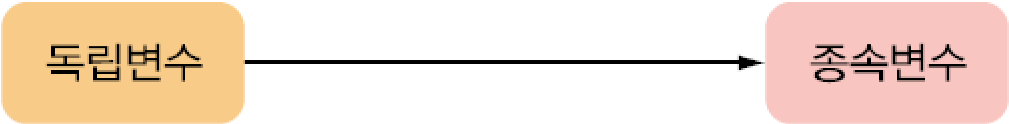
회귀분석

■ 독립변수(dependent variable)

연구나 조사의 연구모델에서 변수에 일어나는 현상을 설명하거나 원인이 되어 다른 변수에 영향을 주는 변수

■ 종속변수(independent variable)

연구나 조사의 연구모델에서 설명되거나 결과가 되어 다른 변수로부터 영향을 받는 변수



단순회귀분석의 개념과 특징

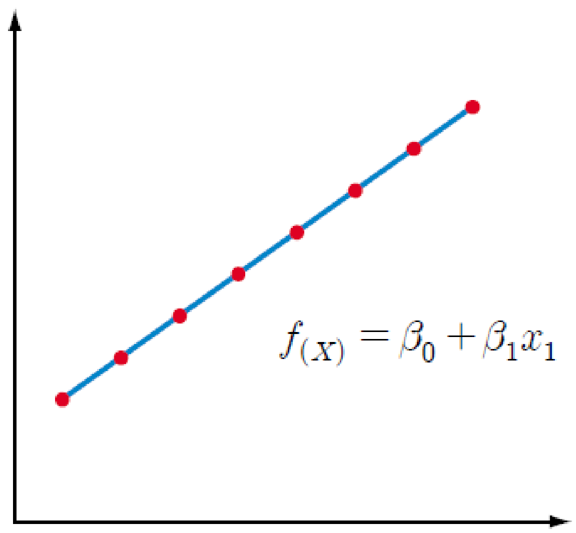
■ 단순회귀분석(simple regression analysis)

독립변수 X 가 종속변수 Y 에 미치는 영향을 회귀식(회귀방정식)을 이용하여 분석하는 방법

→ 회귀식을 이용하여 변수 X 를 원인으로 Y 가 어떻게 될지를 추정하거나 설명하는 것을 목적으로 함.

단순회귀분석의 개념과 특징

자연과학에서는 X의 변화로 인해 Y가 받는 영향이 1:1로 짝을 이루어 계산되므로 회귀식을 특정한 함수식 $f(x)=\beta_0 + \beta_1 x_1$으로 도출



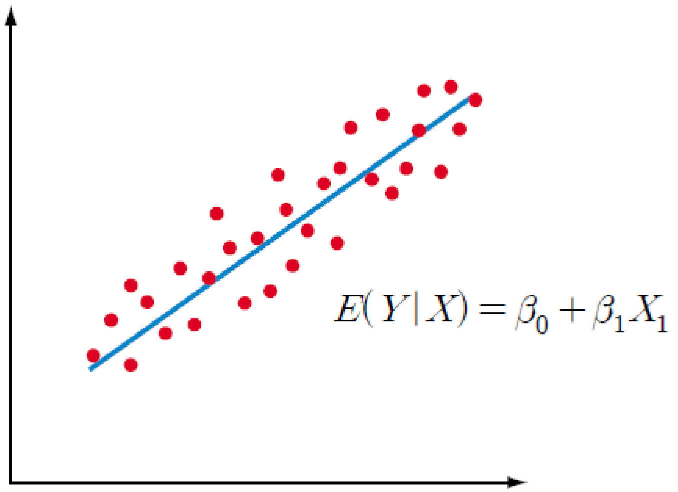
(a) 1 : 1 함수식

단순회귀분석의 개념과 특징

인문/사회과학에서는 고려해야 할 변수가 많다.

[표11-1]의 광고비와 매출액을 보더라도 지난 15년간 국내 경기 상황도 변화무쌍했고, A 주식회사의 기술력이나 업무 능력도 매년 동일하지 않았다.

이로 인해 두 변수 X와 Y를 정확한 식으로 도출하기 어려우므로, 독립변수 X와 종속변수 Y의 관계는 [그림 $11-2(b)$]와 같이 조건부 평균 $E(Y|X)=\beta_0+\beta_1 X_1$ 이 된다.



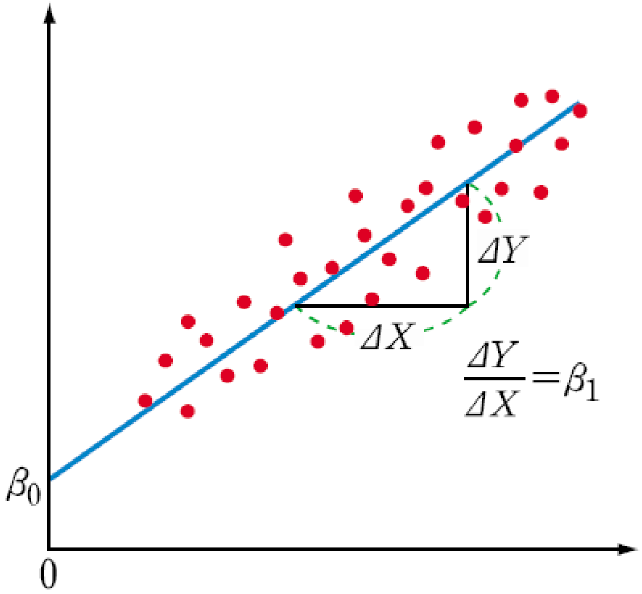
(b) 조건부 평균

단순회귀분석의 개념과 특징

β_0는 x_i = 0 일 때 Y의 값, 즉 β_0는 상수값이며, Y축의 절편

β_1은 X가 증가할 때 Y의 변화량

그래프에서 β_1은 $\dfrac{\Delta Y}{\Delta X}$ 이므로 회귀선의 기울기를 나타낸다.



단순회귀분석의 개념과 특징

■ 잔차(residual)

실제 데이터를 측정하다 보면 독립변수에 따라 종속변수의 변화하는 정도가 다르게 나타나는 경우가 있는데, 이러한 개별 측정치들 간에 차이

잔차를 포함하는 확률적 회귀모형은

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

Note **잔차의 특징**

잔차(ϵ)는 조건부 평균 $E(Y|X) = \beta_0 + \beta_1 X_1$이 실제로 측정된 Y와 다르게 나타나는 부분을 의미하지만, 특정한 패턴(규칙성)을 나타내지는 않는다. 종속변수에 미치는 독립변수들 간에 특정한 패턴을 보인다는 것은 회귀식에 β_2라는 새로운 변수를 넣어야 한다는 의미가 되기 때문이다. 즉 변수를 더 고려해야 한다는 의미이므로 회귀식이 더 복잡해지는 결과로 나타난다.

최소자승법

■ 최소자승법(method of least squares) 혹은 최소제곱법

실제 데이터를 측정하다 보면 독립변수에 따라 종속변수의 변화하는 정도가 다르게 나타나는 경우가 있는데, 이러한 개별 측정치들 간에 차이

표본으로부터 도출된 회귀식을 $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

미지의 모회귀식을 $Y_i = \beta_0 + \beta_1 X_i$

표본에서의 $\hat{\beta}$ 를 모수 β에 가장 가깝게

추정한 회귀식을 도출하는 것이 가장 잘한 분석

최소자승법

$\hat{Y}_{\rm i} = \hat{\beta}_0 + \hat{\beta}_1 X_i$ 에서 회귀식과 측정치의 간의 차이인 잔차 $\widehat{\epsilon_i}$ 가 필연적으로 발생

그러므로 추정 회귀모형은 $\hat{Y}_{\rm i} = \hat{\beta}_0 + \hat{\beta}_1 X_i + \widehat{\epsilon_i}$

잔차 $\widehat{\epsilon_i}$ 의 모든 합이 최소가 되는 회귀식을 구하면 된다.

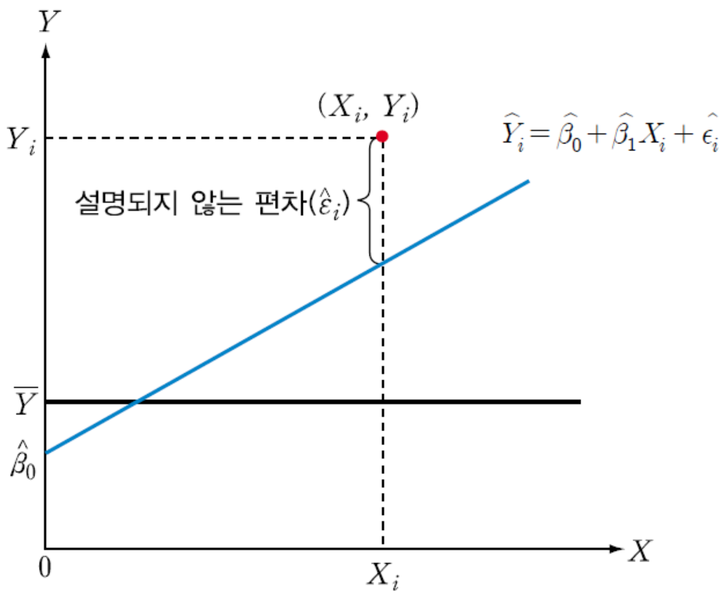
$\hat{Y}_{\rm i} = \hat{\beta}_0 + \hat{\beta}_1 X_i$의 위$(+)$와 아래$(-)$로 분포되어 합이 0이 되므로

모두 제곱하여 합을 구한다.

최소자승법

■ 최소자승법(method of least squares)

잔차의 제곱합 $\sum \hat{\epsilon}_i^2$ 을 최소로 하는 방법



$\hat{\beta}_0 = \overline{Y} - \beta_1 \overline{X}$ 와 $\hat{\beta}_1 = \dfrac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$ 를 통해 구해진 $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$

단순회귀분석

예제 11-1 단순회귀분석[4]

[표 11−1]에 대한 회귀분석을 실시하라.

단순회귀분석 (예제 풀이)

$\hat{\beta}_0$, $\hat{\beta}_1$ 을 구하면

$$\widehat{\beta}_1 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2}$$

$$= \frac{[(13-16.467)(94-98.933)+(8-16.467)(70-98.933)+\cdots]}{[(13-16.467)^2+(8-16.467)^2+\cdots]} = 2.186$$

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \cdot \overline{X} = 98.933 - 2.186 \cdot 16.467 = 62.929$$

따라서 회귀식은 $\hat{Y} = 62.929 + 2.186\, X_i$

단순회귀분석 (예제 풀이)

절편 $\hat{\beta}_0$ 는 광고비를 전혀 지출하지 않더라도

62.929만큼 매출이 발생하고 있다는 것을 의미

$\hat{\beta}_1$ 에서

A주식회사에서의 광고비와 매출액의 인과관계를 판단할 수 있는데,

광고비가 1단위 증가할 때 매출액은 2.186만큼 증가한다는 것을 의미

적합도 검정과 분산분석

■ 적합도 검정(goodness-of-fit test)

도출한 회귀식 $Y_i = \beta_0 + \beta_1 X_1$ 이
표본 측정치를 얼마나 잘 설명하는지를 확인하는 것

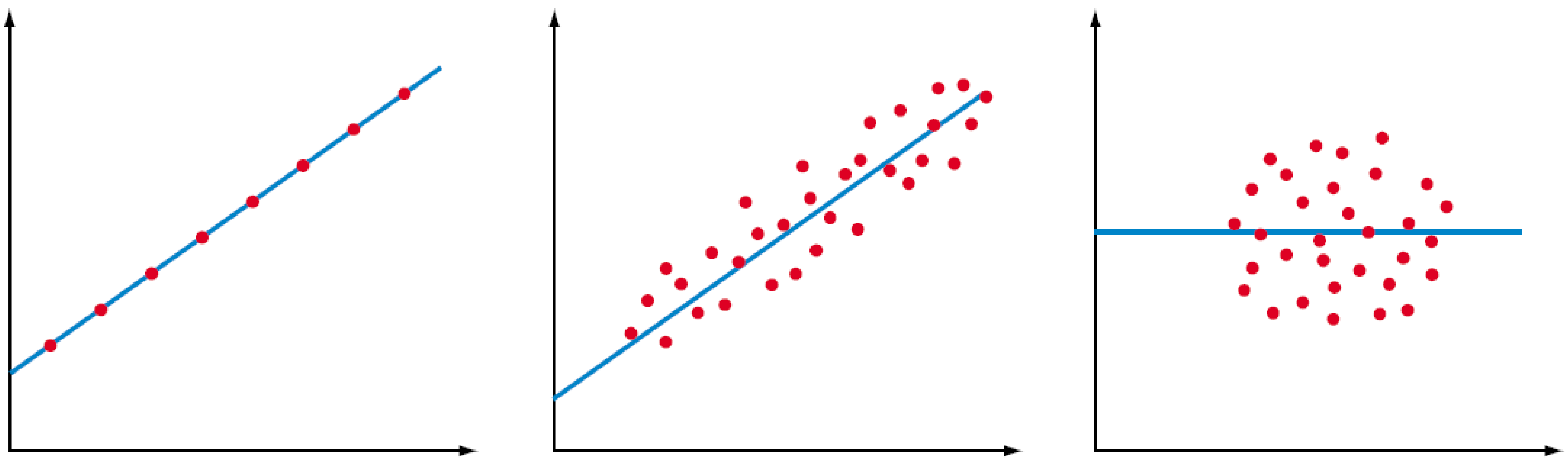
적합도 검정과 분산분석

표본에 대한 회귀선의 설명력 (R^2)

추정된 회귀식이 어느 정도 측정치들과 일치하는지의 정도

설명력은 0~1까지의 숫자로 나타내거나
몇 %의 설명력을 가지는지 확률로 표현

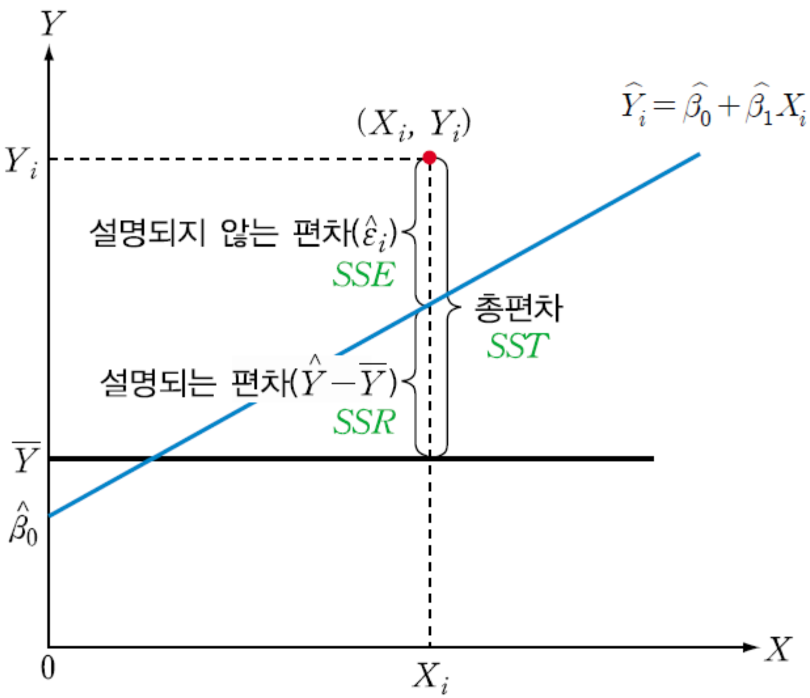
적합도 검정과 분산분석



(a) $R^2 = 1$ (b) $0 < R^2 < 1$ (c) $R^2 = 0$

R^2의 값은 1에 근접할수록
회귀선이 측정치를 잘 반영하고 있다고 할 수 있다.

적합도 검정과 분산분석



R^2은 결정계수라고도 하며

$$\text{결정계수 } R^2 = \frac{\text{회귀 제곱합}(SSR)}{\text{총제곱합}(SST)}$$

적합도 검정과 분산분석

■ 분산분석

회귀분석에서 분산분석이 등장하는 이유는
총편차를 분해하는 과정이 분산분석과 동일하기 때문

각 제곱합 SST, SSR, SSE를 구한 후에는
분산분석에서의 평균제곱을 구하기 위해 각각의 자유도를 알아야 한다.

SST 의 자유도 : $n-1$
SSR 의 자유도 : 1
SSE 의 자유도 : $n-2$

적합도 검정과 분산분석

참고 SST, SSR, SSE의 자유도

SST의 자유도는 전체에서 $\sum(Y_i-\overline{Y})=0$이 되어야 하는 제약 1개로 구성되므로 자유도가 $n-1$이다. SSR의 자유도는 $\widehat{Y}_i$가 독립변수 X_i에 따라 제약되므로 자유도는 1이다. SSE의 자유도는 모수 $\widehat{\beta}_0$, $\widehat{\beta}_1$의 2개를 추정하므로 자유도는 $n-2$이다.

적합도 검정과 분산분석

분산비율 F값을 알기 위해
총 제곱합의 구성 부분인 오차 제곱합과 회귀 제곱합을 각각의 자유도로 나누면

평균 오차제곱($Mean\ Sauare\ Error$: MSE) 과
평균 회기제곱($Mean\ Sauare\ Regerssion$: MSR)을 구할 수 있다.

$$MSE = \frac{SSE}{n-2}, \quad MSR = \frac{SSR}{1}$$

회귀와 잔차에 대한 분산비율 F는

$$F = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE}$$

예제 11-2 단순회귀분석의 결정계수와 분산비율 계산

[표 11–1]에 대해 결정계수와 분산비율을 구하라.

단순회귀분석 (예제 풀이)

총제곱합 SST, 회귀 제곱합 SSR, 오차 제곱합 SSE를 구하면

$$SST = \sum (Y_i - \overline{Y})^2$$

$$= (94 - 98.933)^2 + \cdots + (121 - 98.933)^2 = 1950.933$$

$$SSR = \sum (Y_i - \widehat{Y}_i)^2$$

$$= (91.354 - 98.933)^2 + \cdots + (117.591 - 98.933)^2 = 1538.123$$

$$SSE = \sum (\widehat{Y}_i - \overline{Y})^2$$

$$= (94 - 91.354)^2 + \cdots + (121 - 117.591)^2 = 412.811$$

단순회귀분석 (예제 풀이)

결정계수 R^2 을 구하면

$$R^2 = \frac{SSR}{SST} = \frac{1538.123}{1950.933} = 0.788$$

회귀식은 78.8%의 설명력을 가진다.

분산비율 F를 구하면

$$F = \frac{SSR/1}{SSE/(n-2)} = \frac{1538.123/1}{412.811/13} = 48.438$$

유의성 검정과 회귀계수의 신뢰구간

회귀식의 설명력에 대한 검정을 한 후,

표본을 설명하는 회귀식이 유의한지를 확인하는 과정이 필요

■ 회귀식의 유의성 검정

단순회기분석에서 회귀식 $\hat{Y} = 62.929 + 2.186X_i$에 대하여

F검정을 통해 회귀식의 유의성을 검정

회귀와 잔차에 대한 분산비율 F값을 계산해서

F분포표의 값보다 더 크면 귀무가설을 기각하고 대립가설을 채택

유의성 검정과 회귀계수의 신뢰구간

참고 분산비율 F와 F분포표의 값을 비교하는 의미

분산비율 F는 평균 회귀제곱(MSR)을 평균 오차제곱(MSE)으로 나눈 값으로, 표준오차보다 회귀식으로 설명되는 부분이 어느 정도 더 많은지를 나타내는 수치이다. 그러므로 F값이 분포표의 임계치보다 크다는 것은 회귀식으로 설명할 수 있는 부분이 더 많다는 의미가 되므로, 회귀식이 유의하다는 판단을 내릴 수 있는 근거가 된다.

유의성 검정 (예제 풀이_회귀식)

예제 11-3 단순회귀식의 유의성 검정

다음은 A주식회사의 연도별 광고비와 매출액을 기준으로 작성된 분산분석 결과이다. 이를 기반으로 회귀식이 유의한지 $\alpha = 0.05$에서 검정하라.

모형	제곱합	자유도	평균제곱	분산비율 F
회귀모형	1538.123	1	1538.123	48.438
잔차	412.811	13	31.755	–
합계	1950.933	14	–	–

유의성 검정 (예제 풀이)

회귀식에 대한 유의성을 검정하기 위해 회귀식에 대한 가설을 수립니다.

H_0 : 회귀식이 유의하지 않다.
H_1 : 회귀식이 유의하다.

분산비율이 $F = 48.438$이고, F분포표에서 $F_{(1.13)} = 4.67$
F값이 임계치보다 크므로 귀무가설을 기각하고 대립가설을 채택한다.
따라서 회귀식은 유의

유의성 검정과 회귀계수의 신뢰구간

■ 회귀계수 $\boldsymbol{\beta_0, \beta_1}$ 의 유의성

표본에서 도출한 회귀식 $\hat{Y}\mathrm{i} = \hat{\beta}_0 + \hat{\beta}_1 X_i$ 에서 $\hat{\beta}_0$ 는
회귀선의 절편, $\hat{\beta}_1$ 은 기울기를 의미

회귀계수 $\hat{\beta}_0$ 과 $\hat{\beta}_1$ 를 검정한다는 의미는
표본에서 도출한 회귀식이 의미를 가지는지 아닌지의 여부를 판단하는 것

아무런 의미가 없다면 $\hat{\beta}_0 = 0$ 또는 $\hat{\beta}_1 = 0$ 으로 가설을 수립
반면, 유의미한 역할을 한다면 영향력이 0이 아니라는 의미에서

$$\hat{\beta}_0 \neq 0 \text{ 또는 } \hat{\beta}_1 \neq 0 \text{ 으로 가설을 수립}$$

유의성 검정과 회귀계수의 신뢰구간

가설이 수립되면 $\hat{\beta}_0$ 와 $\hat{\beta}_1$ 에 대한 검정통계량을 계산한다.

$$t_{n-2} \text{의 검정통계량} = \frac{\widehat{\beta}_0 - \beta_0}{s_{\widehat{\beta}_0}} = \frac{\widehat{\beta}_0 - 0}{s_{\widehat{\beta}_0}} = \frac{\widehat{\beta}_0}{s_{\widehat{\beta}_0}}$$

$$t_{n-2} \text{의 검정통계량} = \frac{\widehat{\beta}_1 - \beta_1}{s_{\widehat{\beta}_1}} = \frac{\widehat{\beta}_1 - 0}{s_{\widehat{\beta}_1}} = \frac{\widehat{\beta}_1}{s_{\widehat{\beta}_1}}$$

참고 **회귀계수의 검정통계량을 t로 하는 이유**

최소자승법을 이용하여 표본에서 표본 회귀계수를 계산하게 되는데, 표본에서 도출된 회귀계수는 표본의 구성에 따라 변한다. 이때 표본의 회귀계수에 대한 분포를 확인해야 하는데, 표본의 잔차에서 구한 분산을 이용한다. 모수를 알 수 있다면 z분포를 이용할 수 있겠으나, 그렇지 않으므로 t분포를 이용한다. t분포는 자유도에 따라 민감하게 반응하므로 자유도가 중요한 개념이 된다.

유의성 검정 (예제 풀이_회귀계수)

예제 11-4 회귀계수의 유의성 검정

[예제 11-3]에서 회귀계수 $\widehat{\beta}_0$와 $\widehat{\beta}_1$에 대한 유의성을 판단하라.

유의성 검정 (예제 풀이_회귀계수)

β_0, β_1의 유의성을 검정하기 위해 $\hat{\beta}_0$ 와 $\hat{\beta}_1$ 에 대한 가설을 수립한다.

$$H_0: \beta_0 = 0, \quad H_1: \beta_0 \neq 0$$

$$H_0: \beta_1 = 0, \quad H_1: \beta_1 \neq 0$$

$\beta_0 = 62.929, \ \beta_1 = 2.186, \ MSE = 31.755, \ \sum \widetilde{X^2} = 321.733$ 이므로

$s_{\widehat{\beta}_0}$ 와 $s_{\widehat{\beta}_1}$ 를 구하면

$$s_{\widehat{\beta}_0} = \sqrt{\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum \widetilde{X^2}}\right) MSE} = \sqrt{\left(\frac{1}{15} + \frac{16.467^2}{321.733}\right) 31.755} = \sqrt{28.88} = 5.374$$

$$s_{\widehat{\beta}_1} = \sqrt{MSE / \sum \widetilde{X^2}} = \sqrt{\frac{31.755}{321.733}} = 0.314$$

유의성 검정 (예제 풀이_회귀계수)

$\hat{\beta}_0$ 와 $\hat{\beta}_1$ 에 대한 $t_n - 2$의 검정통계량을 구하면

$$\frac{\widehat{\beta_0}}{s_{\widehat{\beta_0}}} = \frac{62.929}{5.374} = 11.710, \ |\pm t_{(0.025,\ \alpha/2)}| = 2.1604$$

$$\frac{\widehat{\beta_1}}{s_{\widehat{\beta_1}}} = \frac{2.186}{0.314} = 6.960, \ |\pm t_{(0.025,\ \alpha/2)}| = 2.1604$$

그러므로 $\hat{\beta}_0$ 와 $\hat{\beta}_1$ 모두 귀무가설을 기각하고 대립가설을 채택

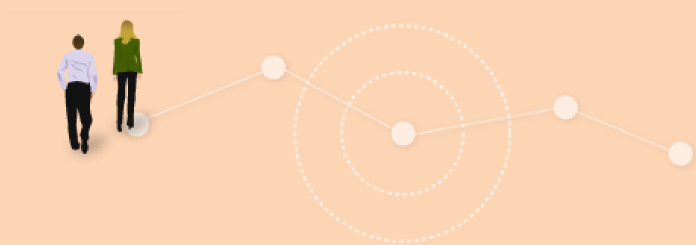
단순회귀분석 (예제 Excel 풀이)

예제 11-6 **단순회귀분석 출력** 준비파일 | 11장_단순회귀분석.xlsx

[예제 11-1]에 대해 Excel을 이용하여 단순회귀분석을 실시하라.

02 다중회귀분석

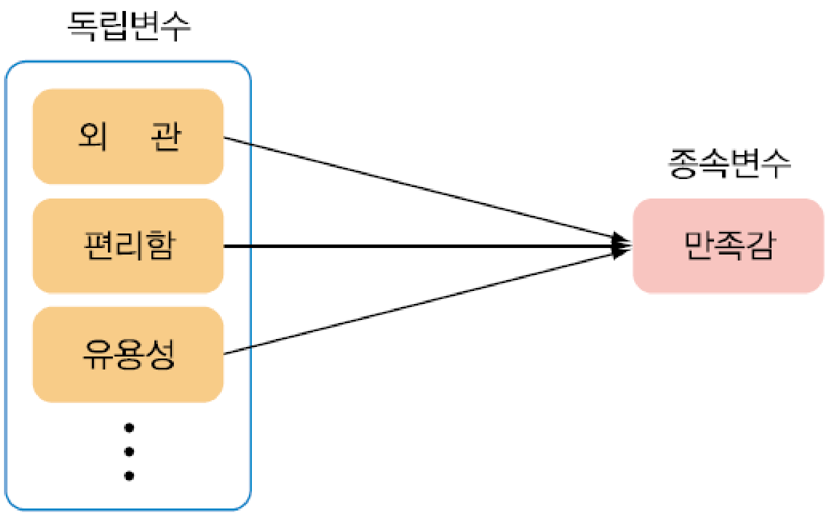
:: **Keywords** 다중회귀분석 | 단순회귀분석과의 차이점 | 수정된 R^2



다중회귀분석의 개념과 특징

■ 다중회귀분석(multiple regression analysis)

종속변수에 영향을 미치는 독립변수가 여러 개인 경우에 실시하는 회귀분석



다중회귀분석의 개념과 특징

다중회귀모형에서 독립변수의 개수가 i개라면
X_i 와 각 독립변수에서의 측정치 개수를 j로 구분하여 설명할 수 있다.

다중회귀모형의 회귀식은

$$E(Y \mid X_{ij}) = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_i X_{ij}$$

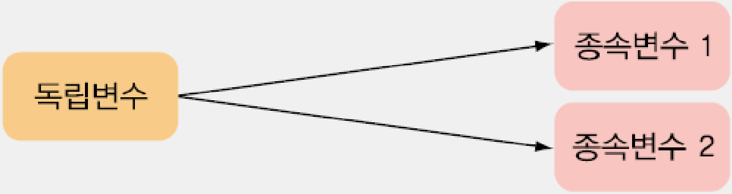
개별 측정치들 사이의 차이인 잔차를 회귀식에 반영한 확률적 회귀모형은

$$\widehat{Y} = \widehat{\beta_0} + \widehat{\beta_1} X_{1j} + \widehat{\beta_2} X_{2j} + \cdots + \widehat{\beta_i} X_{ij} + \epsilon_i$$

다중회귀분석의 개념과 특징

참고 종속변수가 여러 개인 경우 회귀분석

기본적으로 회귀분석은 독립변수는 1개 이상이며, 종속변수는 1개로 정한다. 다시 말해, 회귀분석은 종속변수가 여러 개인 경우에는 실시할 수 없다. 변수들 간의 인과관계를 파악하고자 연구모델을 설정하면 독립변수를 고려한 연구모델이 최종적으로 결정되겠으나, 더욱 정교한 모델을 구성하기 위해서는 [그림 11-12]와 같이 종속변수가 다수인 경우도 생각할 수 있다. 과연 이러한 경우는 어떻게 분석을 해야 할까?



[그림 11-12] **다수의 종속변수**

그 답은 종속변수의 개수만큼 회귀분석을 실시하는 것이다. 즉 독립변수가 종속변수1에 미치는 영향을 먼저 확인한 후, 다시 독립변수가 종속변수2에 미치는 영향을 확인해야 한다. 회귀분석은 독립변수와 종속변수 간의 인과관계를 표현하여 사회현상을 잘 설명할 수 있는 분석 방법이지만, 종속변수의 개수에 대해 분석 방법을 반복해야하므로 번거로울 수도 있다.[10]

다중회귀분석의 회귀계수 계산

다중회귀분석에서 회귀계수 $\hat{\beta}_i$를 계산하는 방법은
독립변수의 수가 늘었다는 것 외에는 단순회귀분석과 동일

$Y = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_i X_{ij}$에서 $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1j} + \widehat{\beta}_2 X_{2j} + \cdots$
$+ \widehat{\beta}_i X_{ij} + \widehat{\epsilon}_i$를 뺀

$Y - \widehat{Y}$의 나머지 $\sum \epsilon_i$를 최소로 하는 최소자승법을 사용

적합도 검정과 분산분석

적합도 검정

도출한 회귀식이 어느 정도 측정치들을 설명하는지를 나타냄

다중회귀분석에서 결정계수는

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

독립변수가 i개로 늘어나기 때문에
독립변수의 개수가 늘어나는 만큼 R^2이 변화하는 정도를 수정해야 한다.

적합도 검정과 분산분석

수정된 $R^2(adiusted\ R^2)$은
R^2이 변화하는 정도를 반영하면서 각 자유도로 나누어 불편추정량을 계산

$$수정된\ R^2 = \frac{\frac{SSR}{i}}{\frac{SST}{n-1}} = 1 - \frac{\frac{SSE}{n-i-1}}{\frac{SST}{n-1}}$$

적합도 검정과 분산분석

■ 분산분석

다중회귀분석도 총편차를 분해하는 과정이 분산분석과 동일

→ 각 제곱합 SST, SSR, SSE를 구한 후에는

분산분석에서의 평균제곱을 구하기 위해 각각의 자유도를 알아야 한다.

SST의 자유도 : $n-1$

SSR의 자유도 : i

SSE의 자유도 : $n-i-1$

참고 **SST, SSR, SSE의 자유도**

SST의 자유도는 전체에서 $\sum(Y_i-\overline{Y})=0$이 되어야 하는 제약 1개로 구성되므로 자유도는 $n-1$이고, SSR의 자유도는 $\widehat{Y}_i$가 독립변수의 개수 X_i에 따라 제약되므로 자유도는 i이다. SSE의 자유도는 모수 $\widehat{\beta}_0$, $\widehat{\beta}_i$가 $i+1$개이므로 $n-(i+1)=n-i-1$이다.

적합도 검정과 분산분석

평균오차제곱(MSE)과 평균 회귀제곱(MSR)은

$$MSE = \frac{SSE}{n-i-1}, \quad MSR = \frac{SSR}{i}$$

회귀와 잔차에 대한 분산비율 F는

$$F = \frac{SSR/i}{SSE/(n-i-1)} = \frac{MSR}{MSE}$$

다중회귀분석

예제 11-8 다중회귀분석의 결정계수와 분산비율 계산

[예제 11-7]를 이용하여 결정계수와 분산비율을 구하라.

다중회귀분석 (예제 풀이)

총 제곱합 SST, 회귀 제곱합 SSR, 오차 제곱합 SSE를 구하면

$$SST = (5 - 3.475)^2 + (3 - 3.475)^2 + \cdots + (1.33 - 3.475)^2 = 45.992$$

$$SSR = (4.644 - 3.475)^2 + (3.801 - 3.475)^2 + \cdots + (2.002 - 3.475)^2 = 28.827$$

$$SSE = (5 - 4.644)^2 + (3 - 3.801)^2 + \cdots + (1.33 - 2.002)^2 = 17.165$$

결정계수 R^2를 구하면

$$R^2 = \frac{SSR}{SST} = \frac{17.165}{45.992} = 0.373$$

다중회귀분석 (예제 풀이)

R^2을 불편추정량으로 변경하기 위하여 수정된 R^2을 계산해야 한다.

$$\text{수정된 } R^2 = 1 - \frac{28.827/(80-3-1)}{45.992/(80-1)} = 1 - \frac{0.379}{0.582} = 0.349$$

따라서 회귀식은 34.9%의 설명력을 가진다.

분산비율 F을 구하면

$$\frac{SSR/i}{SSE/(n-i-1)} = \frac{17.165/3}{28.827/76} = 15.084$$

유의성 검정과 회귀계수의 신뢰구간

■ 회귀식의 유의성 검정

다중회기분석에서 회기식 $\hat{Y} = 3.514 + 0.269X_1 + 0.211X_2 + 0.162X_3$에 대해 F검정을 통해 회귀식의 유의성을 검정해야 한다.

회귀식의 유의성을 검정하는 방법은
회귀와 잔차에 대한 분산비율 F값을 계산해서

F분포표의 값보다 더 크다면 귀무가설을 기각하고
대립가설을 채택하는 것

유의성 검정과 회귀계수의 신뢰구간

예제 11-9 다중회귀식의 유의성 검정

다음은 스마트폰의 만족도에 영향을 주는 3가지 요인으로 외관, 편의성, 유용성에 대한 회귀분석을 실시한 결과이다. 이를 기반으로 회귀식이 유의한지 $\alpha = 0.05$에서 검정하라.

모형	제곱합	자유도	평균제곱	분산비율 F
회귀모형	17.165	3	5.722	15.084
잔차	28.827	76	0.379	−
합계	45.992	79	−	−

유의성 검정과 회귀계수의 신뢰구간 (예제 풀이)

회기식에 대한 유의성을 검정하기 위해 회귀식에 대한 가설을 수립한다.

H_0 : 회귀식이 유의하지 않다.

H_1 : 회귀식이 유의하다.

분산비율이 F=15.084이고, F분포표에서 $F_{(3.76)} = 2.72$

F값이 임계치보다 크므로,

귀무가설을 기각하고 대립가설을 채택

따라서 회귀식은 유의하다.

유의성 검정과 회귀계수의 신뢰구간

■ 회귀계수의 신뢰구간

표본에서 도출한 회귀식 $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_{1j} + \cdots + \hat{\beta}_i X_{ij} + \hat{\epsilon}_i$에서

$\hat{\beta}_i$는 $\sum \hat{\epsilon}_i = 0$인 경우를 유의한 경우로 가설을 수립하면

$$H_0 : \widehat{\beta}_i = 0 \text{과 } H_1 : \widehat{\beta}_i \neq 0$$

가설이 수립되면 $\hat{\beta}_i$에 대한 검정통계량을 계산한다.

$$t_{n-i-1} \text{의 검정통계량} = \frac{\hat{\beta}_i - \beta_i}{s_{\widehat{\beta}_i}} = \frac{\hat{\beta}_i - 0}{s_{\widehat{\beta}_i}} = \frac{\hat{\beta}_i}{s_{\widehat{\beta}_i}}$$

다중회귀분석

예제 11-11 **다중회귀분석 출력**

준비파일 | 11장_다중회귀분석.xlsx

[예제 11-7]을 이용하여 다중회귀분석을 출력하라.