

실험계획법 / 확률과 통계 기초

강의노트 20

회귀분석 (Linear Regression)

137. 회귀분석

관찰된 연속형 변수들에 대해 독립변수와 종속변수 사이의 상관관계에 따른 수학적 모델인 선형적 관계식을 구하여 어떤 독립변수가 주어졌을 때 이에 따른 종속변수를 예측한다. 또한 이 수학적 모델이 얼마나 잘 설명하고 있는지를 판별하기 위한 적합도를 측정하는 분석 방법이다. (위키)

138. 단순회귀분석과 다중회귀분석

1개의 종속변수와 1개의 독립변수 사이의 관계를 분석할 경우를 단순회귀분석(Simple Regression Analysis), 1개의 종속변수와 2개 이상의 독립변수 사이의 관계를 규명하고자 할 경우를 다중회귀분석(Multiple Regression Analysis)이라고 한다.

139. 회귀

회귀(Regress)의 원래 의미는 옛날 상태로 돌아가는 것을 의미한다. 영국의 유전학자 프란시스 갈튼(Francis Galton)은 부모의 키와 아이들의 키 사이의 연관 관계를 연구하면서 부모와 자녀의 키사이에는 선형적인 관계가 있고 키가 커지거나 작아지는 것보다는 전체 키 평균으로 돌아가려는 경향이 있다는 가설을 세웠으며 이를 분석하는 방법을 "회귀분석"이라고 하였다. 이러한 경험적 연구 후에 칼 피어슨(Karl Pearson)은 아버지와 아들의 키를 조사한 결과를 바탕으로 함수 관계를 도출하여 수학적 전개를 정립하였다.

140. 단순선형회귀

$$x = \beta_0 + \beta_1 x$$

$$i = \beta_0 + \beta_1 x_i + E_i$$

141. 최소제곱법

$$SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

$$\frac{\partial SSE}{\partial b_0} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i)$$

$$\frac{\partial SSE}{\partial b_1} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) x_i$$

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

142. 예제

관찰	X	Y
1	1	1.1
2	2	2.0
3	3	2.9
4	4	4.3
5	5	5.1

$$b_1 = \{5(1*1.1+2*2.0+3*2.9+4*4.3+5*5.1) -$$

-

$$(1+2+3+4+5)(1.1+2.0+2.9+4.3+5.1)\} / \{5*(1^2+2^2+3^2+4^2+5^2) - (1+2+3+4+5)^2\}$$

$$b_0 = (1.1+2.0+2.9+4.3+5.1) / 5 -$$

$$b_1 * (1+2+3+4+5) / 5$$

$$b_1 = 1.03 \dots \text{기울기}$$

$$b_0 = -0.01 \dots \text{절편}$$

143. Example

Observation	(x) Relative humidity, (%)	(y) Solvent evaporation, (%) wt
1	35.3	11.0
2	29.7	11.1
3	30.8	12.5
4	58.8	8.4
5	61.4	9.3
6	71.3	8.7
7	74.4	6.4
8	76.7	8.5
9	70.7	7.8
10	57.5	9.1
11	46.4	8.2
12	28.9	12.2
13	28.1	11.9
14	39.1	9.6
15	46.8	10.9
16	48.5	9.6
17	59.3	10.1
18	70.0	8.1
19	70.0	6.8
20	74.4	8.9
21	72.1	7.7
22	58.1	8.5
23	44.6	8.9
24	33.4	10.4
25	28.6	11.1