



제1장 확률과 확률분포

제2장 통계적 검정 및 추정

제3장 상관분석 및 단순회귀분석

행운은 100% 노력한 뒤에
남는 것이다!
- 랭스턴 콜만 -

제3장 상관분석 및 단순회귀분석

1. 상관·회귀분석 개요

1.1 상관·회귀분석의 개념

- * 금속의 성분과 항장력 또는 항장력과 연신률, 소둔온도와 금속의 경도 등과 같이 만약 한쪽의 측정치가 변하면 다른 측정치가 이것에 따라 변화하는 관계가 있을 때 이들 사이에는 “상관이 있다.”고 말하며, 상관관계를 해석하는 방법을 상관분석이라 한다.
- * 상관분석은 X, Y 다 같이 랜덤하게 변하는 확률변수로서 2변량 정규분포를 하는 것이 전제로 되어 있다. 따라서 X 가 확률변수가 아닐 경우, 예를 들면 온도를 100°C , 110°C , 120°C , ...와 같이 지정하여 변화시키고, 이때의 반응생성물의 수확량을 측정하여, 이 온도와 수확량과의 관계를 문제로 삼을 경우에는, X 가 랜덤하게 변하는 확률변수가 아니므로 상관분석을 적용하는 것은 적절하지 못하다. 이와 같이 X 의 값을 알고 있는 확정수일 때에는 회귀분석으로 취급한다.

1.2 상관·회귀분석의 목적

- * 품질관리에서는 다음의 목적 때문에 상관·회귀가 사용된다.
 - ① 특성과 특성간의 상관관계, ② 대응 관리특성의 발견
 - ③ 특성과 요인간의 관계, ④ 층별하여 상관을 취함으로써 교호작용 발견 등
- * 상관과 회귀의 종국 목적은 변수 x 로부터 변수 y 를 추정하는데 있다.
- * 상관·회귀의 공통점은 변수들간의 관계를 규명하는데 쓰이는 것이며, 차이점은 회귀는 통계적 모형을 구축하고 그 모형으로부터 미래값을 예측하는데 활용된다는 데 있다.

2. 상관분석

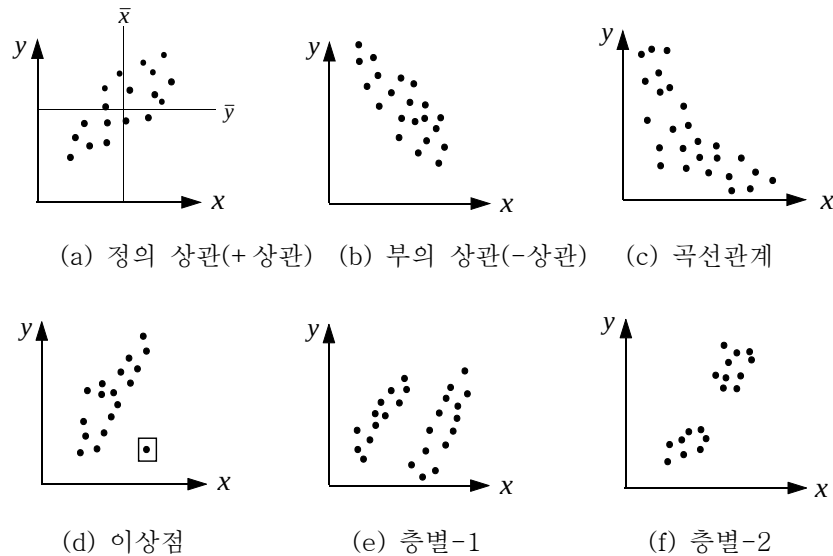
2.1 상관분석의 순서

- * 상관분석은 다음의 순서에 의해 이루어진다.
 - ① 산점도(산포도라고도 함)를 그린다.
 - ② 상관계수를 계산한다. (상관계수= r)
 - ③ 상관계수의 유의성에 대한 검정 및 추정을 한다.
 - ④ 유의이면 x 에 대한 y 의 회귀선을 구한다.
 - ⑤ 결정계수(기여율)을 구한다. [결정계수(기여율)= r^2]

2.2 상관분석을 위한 산점도의 작성

2.2.1 상관·회귀의 기초로서의 산점도

* 2변량 사이의 상관관계를 알기 위하여 2개의 변량 x, y 를 그래프에 타점하여 2개의 변량을 각자 세로축과 가로축에 잡아 측정치를 타점하여 [그림 3.1]과 같이 작성한 것을 산점도 (scatter diagram)라고 한다.



[그림 3.1] 산점도

* 산점도를 작성하는 방법은 제1장의 1항 데이터정리법에서 7가지 QC기초수법 중 산점도 항목을 참조바랍니다.

2.2.2 산점도를 그린 후 검토사항

* 산점도를 그린 후에 검토하여야 할 사항은 다음과 같은 것들이다.

- ① 점들이 산재해 있는 모양으로부터 x 와 y 사이에 관계가 있을 듯한가에 대해 검토한다.
또한 양(positive) 상관인가 음(negative) 상관인가를 알아 본다. [그림 3.1의 (a)와 (b)]
- ② x 와 y 가 직선관계인가 곡선관계인가를 살펴본다.
[그림 3.1]의 (c)처럼 곡선관계인 경우에는 상관계수를 구하는 것은 의미가 없고, 이와 같은 경우는 곡선회귀분석을 하여야 한다. 회귀선이 2차식이면 $r = 0$ 이다.
- ③ 이상한 데이터가 없나를 확인한다.
산점도에서 이상점의 발생은 데이터의 수집시 다른 모집단의 시료가 혼입되었거나, 측정이 나 계산이 잘못되었거나, 데이터의 기입 등에 착오가 있는 것이다. 이와 같은 이상점이 발견되면 원인을 규명하여 수정하여 주도록 한다.
- ④ 점들이 뚜렷하게 두 개 또는 그 이상으로 층별되는 경우가 있는가 검토한다.
[그림 3.1]의 (e)처럼 두 개의 층으로 나누어지는 경우에는 두 개의 모집단으로부터 얻어진 두 개의 시료가 섞여 있는 것이므로 두 개의 서로 다른 시료로 취급하여야 한다.

2.3 표본상관계수 r 을 구함

* 상관계수는 두 변수간의 선형(직선) 관계가 존재하는지 알아보는 계수이다. 회귀분석과 유사하지만 인과관계에 대한 분석은 아니다.

* 두 확률변수 X, Y 에 대하여 n 개의 데이터 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 이 얻어졌을 때 그 표본(또는 시료)상관계수(sample correlation coefficient)는 다음과 같이 정의된다.

$$r = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} \quad (3.1)$$

$$\text{여기서, } S_{(xx)} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$S_{(yy)} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$S_{(xy)} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

* 점들이 직선에 모여 있으면 상관계수는 커진다. 상관계수의 부호가 양이면 한 변수의 값이 커질수록(작아질수록) 다른 변수의 값도 커짐(작아짐)을 의미한다.

* 표본의 크기가 커지면 상관계수의 값이 커지며, 상관계수의 값이 얼마 이상이어야 한다는 기준은 없으므로, 가설검정에 의한 유의확률(P value)을 계산하도록 한다.

* 상관계수에 대해 다음 사항을 특히 주의하도록 한다.

- ① 상관계수는 두 변수간의 선형관계를 알아보는 것이다. 2차관계의 상관계수는 0이다.
- ② 상관계수는 점들이 직선에 모여 있는 정도를 나타내는 지표이지 직선의 기울기 크기를 나타내는 것이 아니다.

* 표본상관계수가 갖는 몇 가지 성질을 살펴보면 다음과 같다.

- ① r 의 범위는 $-1 \leq r \leq 1$ 이다.
- ② r 의 값은 두 확률변수 X, Y 간의 선형관계(linear relationship)를 나타내는 척도로서 [그림 3.1]의 (a)의 경우는 $0 < r < 1$ 이고 (b)의 경우는 $-1 < r < 0$ 이 된다.
 $r = \pm 1$ 인 경우는 모든 점이 기울기가 영이 아닌 일직선상에 놓이게 된다.
- ③ 만약 x 가 $ax + b$ 로 바뀌고, y 가 $cy + d$ 로 바뀌어도 a 와 c 의 부호가 같으면 r 의 값에는 변함이 없다. 따라서 계산의 편의상 수치변환(단, $h > 0, g > 0$ 임)

$$X = (x - x_0) \times h, Y = (y - y_0) \times g$$

을 하여도 x, y 간의 상관계수는 X, Y 의 상관계수와 동일하다. 즉,

$$r = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} = \frac{\frac{1}{hg}S_{(XY)}}{\sqrt{\frac{1}{h^2}S_{(XX)} \cdot \frac{1}{g^2}S_{(YY)}}} = \frac{S_{(XY)}}{\sqrt{S_{(XX)}S_{(YY)}}}$$

여기서 X, Y 는 확률변수를 나타내는 대문자의 의미로 사용한 것이 아니라 x, y 의 수치 변환된 값을 나타내고 있다.

2.4 상관계수 유의성 검정 및 추정

2.4.1 상관계수의 유의성 검정

(1) 상관계수의 유의성 검정 ($\rho = 0$ 일 때)

적용	H_0	H_1	검정통계량	기각역
t 표	$\rho = 0$	$\rho \neq 0$	$t_0 = r\sqrt{n-2}/\sqrt{1-r^2}$	$ t_0 > t_{1-\alpha/2}(n-2)$
r 표	$\rho = 0$	$\rho \neq 0$	$r_0 = S_{(xy)}/\sqrt{S_{(xx)}S_{(yy)}}$	$ r_0 > r_{1-\alpha/2}(n-2)$
비고	$r = S_{(xy)}/\sqrt{S_{(xx)}S_{(yy)}}$, r 분포표의 값은 부록의 <부표 12> r 분포표 참조.			

* 이러한 상관계수의 유의성 검정으로 H_0 가 기각되면, 회귀선을 $\hat{y} - \bar{y} = b(x - \bar{x})$ 로 구할 수 있게 된다.

참조 수치변환에 의한 상관계수의 계산 방법

상관계수(r)을 구하기 위한 수치변환의 경우는 다음과 같은 방법으로 행한다.

$X = (x - x_0) \times h$, $Y = (y - y_0) \times g$ 로 변환하면, 변동(편차제곱합)은

$$S_{(XX)} = \sum X^2 - (\sum X)^2/n, \quad S_{(YY)} = \sum Y^2 - (\sum Y)^2/n$$

$$S_{(XY)} = \sum XY - (\sum X)(\sum Y)/n$$

이고, 이들을 원래의 값으로 환원시키기 위해 다음과 같이 계산한다.

$$S_{(xx)} = S_{(XX)} \times \frac{1}{h^2}, \quad S_{(yy)} = S_{(YY)} \times \frac{1}{g^2}, \quad S_{(xy)} = S_{(XY)} \times \frac{1}{hg}$$

(2) 상관계수의 유의성 검정 ($\rho = \rho_0$ 즉 $\rho \neq 0$ 일 때)

H_0	H_1	검정통계량	기각역
$\rho = \rho_0$	$\rho \neq \rho_0$	$U_0 = \sqrt{n-3} \left(\frac{1}{2} \ln \left\{ \frac{1+r}{1-r} \right\} - \frac{1}{2} \ln \left\{ \frac{1+\rho_0}{1-\rho_0} \right\} \right)$	$ U_0 > u_{1-\alpha/2}$

2.4.2 상관계수에 관한 추정

(1) 두 확률변수 X 와 Y 간의 모상관계수 ρ 의 추정

* 두 확률변수 X , Y 간의 모상관계수 ρ 는 다음과 같다.

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

여기서 $\text{Cov}(X, Y) = \sigma_{xy}$, $V(X) = \sigma_x^2$, $V(Y) = \sigma_y^2$ 으로 바꿔 놓은 것이다.

* 이들의 시료에 의한 추정치는 각각

$$\hat{\sigma}_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{S_{(xy)}}{n-1}$$

$$\hat{\sigma}_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{S_{(xx)}}{n-1}, \quad \hat{\sigma}_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{S_{(yy)}}{n-1}$$

를 사용할 수 있으므로, ρ 의 점추정치는 다음과 같이 된다.

$$\hat{\rho} = \frac{\frac{S_{(xy)}}{n-1}}{\sqrt{\frac{S_{(xx)}}{n-1} \cdot \frac{S_{(yy)}}{n-1}}} = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} = r \quad (3.2)$$

(2) 모상관계수에 관한 신뢰구간의 추정 (대략 $n \geq 25$ 일 때)

* 상관계수의 분포에서 $Z = \frac{1}{2} \ln \frac{1+r}{1-r} = \tanh^{-1}(r)$ 로 정의되며, $E(Z)$ 값은 다음 식으로 구해진다.

$$E(Z) = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)$$

* Z 의 $100(1-\alpha)\%$ 신뢰구간은 다음 식으로 된다.

$$\left. \begin{matrix} Z_U \\ Z_L \end{matrix} \right\} = Z \pm u_{1-\alpha/2} \sqrt{\frac{1}{n-3}} \quad (3.3)$$

구간상한 Z_U 와 구간하한 Z_L 에 대응하는 ρ_U, ρ_L 의 값이 모상관계수의 구간추정값이다.

* Z 값으로부터 ρ 를 추정하는 방법은 다음의 방법들이 있다.

(가) 방법 1 [$r = \tanh(Z)$ 이용]

* $Z = \tanh^{-1}(r)$ 이므로 $r = \tanh(Z)$ 이다. 따라서 $\hat{\rho} \approx r = \tanh(Z)$ 이 된다.

$$Z_U \rightarrow \rho_U \text{ 변환 : } \hat{\rho}_U \approx r_U = \tanh(Z_U)$$

$$Z_L \rightarrow \rho_L \text{ 변환 : } \hat{\rho}_L \approx r_L = \tanh(Z_L)$$

* 따라서 ρ 의 구간추정은 $\hat{\rho}_L \leq \hat{\rho} \leq \hat{\rho}_U$ 로 구해진다.

(나) 방법 2 [$r = (e^{2Z} - 1)/(e^{2Z} + 1)$ 의 공식 사용]

* ρ 의 $100(1-\alpha)\%$ 신뢰구간은 다음의 식으로 계산되는 것이 증명되어져 있다.

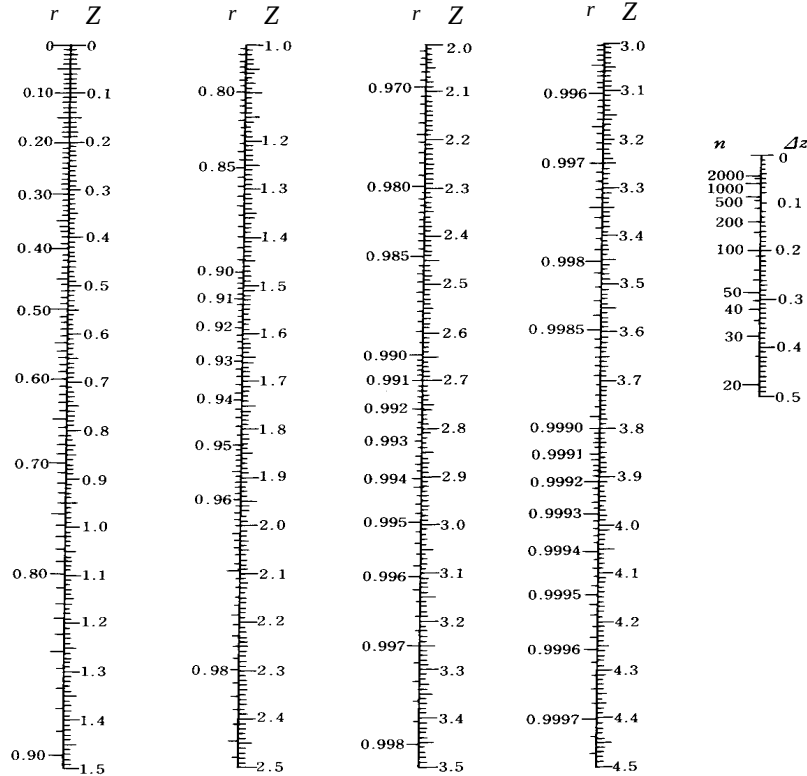
$$\hat{\rho}_L \approx r_L = \frac{e^{2Z_L} - 1}{e^{2Z_L} + 1} \quad (3.4)$$

$$\hat{\rho}_U \approx r_U = \frac{e^{2Z_U} - 1}{e^{2Z_U} + 1} \quad (3.5)$$

* 따라서 ρ 의 구간추정은 $\hat{\rho}_L \leq \hat{\rho} \leq \hat{\rho}_U$ 로 구해진다.

(다) 방법 3 (Z 변환표에 의한 경우)

- ▶ 순서 1 : 시료상관계수 r 의 값을 Z 변환도표 상에서 찾는다.
- ▶ 순서 2 : 시료의 크기 n 에 대응하는 ΔZ 의 폭을 구한다.
- ▶ 순서 3 : 앞에서 Z 변환도표 위에서 잡은 r 의 값에서 상하에 ΔZ 의 폭을 주었을 때 얻어지는 상하의 두 점의 r 의 값이 신뢰구간의 하한 ρ_L 과 상한 ρ_U 의 추정값이 된다.



$$Z = \frac{1}{2} \ln \frac{1+r}{1-r} = \tan h^{-1}(r) \quad r = \tan h(Z) \quad \Delta Z = \frac{1.96}{\sqrt{n-3}}$$

[그림 3.2] Z 변환도표

2.5 x에 대한 y의 회귀선

- * x에 대한 y의 회귀선은 다음의 식으로 구한다.

$$\hat{y} - \bar{y} = b(x - \bar{x}) \quad (3.6)$$

- * 상기 식에서 회귀계수 b 는 다음 식으로 구한다.

$$b = \frac{S_{(xy)}}{S_{(xx)}} \quad (3.7)$$

- * 상기 b 의 식에서 $S_{(xx)}$ 와 $S_{(xy)}$ 는 다음의 식에서 구해진 값을 구하여 대입시킨다.

$$S_{(xx)} = \sum x^2 - \frac{(\sum x)^2}{n}, \quad S_{(xy)} = \sum xy - \frac{\sum x \sum y}{n}$$

〔참조〕 Y 에 대한 x 의 회귀선을 구할 경우

$$x - \bar{x} = b(y - \bar{y}) \quad [\text{여기서, 회귀계수 } b = S_{(xy)} / S_{(yy)}]$$

〔참조〕 수치변환에 의한 회귀식의 계산의 경우

위의 회귀선에서의 x 의 평균 $\bar{x}$ 와 Y 의 평균 $\bar{y}$, 회귀계수 b 의 값을 구할 때 계산의 단순화를 위해 $X = (x - x_0) \times h$, $Y = (y - y_0) \times g$ 에 의해 수치변환한 값에 의한 경우에는 다음의 식으로 계산한 값을 식 (3.6)에 대입시킨다.

$$\bar{x} = x_0 + \frac{\sum X}{n} \times \frac{1}{h}, \quad \bar{y} = y_0 + \frac{\sum Y}{n} \times \frac{1}{g}$$

2.6 결정계수(기여율) R^2

* 상관계수(r)에 의거하여 결정계수(기여율) R^2 은 다음 식에 의하여 구한다.

$$R^2 = (r)^2 = \left(\frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} \right)^2 \quad (3.8)$$

* 결정계수(기여율)의 의미는 총변동(SST : Sum of Squared Total) 중에서 기여율에 해당하는 수치가 회귀에 의하여 설명되는 변동(SSR : Sum of Squared Regression)을 의미하고, 나머지 값은 잔차변동(SSE : Sum of Squared Error)을 나타낸다.

$R^2 > 0.65$ 이면 회귀식이 자료를 설명하기에 좋다고 판정한다.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (3.9)$$

결정계수는 두 변수간의 선형관계 정도가 높으면(관측치들이 가까이 모여 있다는 것을 의미함) 결정계수는 1에 가까워 진다. 특히 단순회귀 모형에서 상관계수는 $r = \pm\sqrt{R^2}$ 이 성립한다.

3. 단순회귀분석

3.1 회귀분석의 개념

- * 상관분석에 의하여 2개의 특성간의 관계를 조사할 경우, 측정치 x 와 y 는 어느 것이나 랜덤하게 변하는 확률변수이어야 한다는 것이 전제로 되어 있음을 보았다.
- * 특성치간의 관계를 다룰 때는, 이런 경우에 한하는 것은 아니다. 예를 들면 반응생성물의 수율 Y 에 영향을 미친다고 생각하는 반응온도 x 를 80°C , 90°C , 100°C , ...와 같이 지정하여 실험을 행하여 x 와 Y 의 관계를 알고 싶을 경우와 같이 x 가 확정수, Y 가 변수일 경우에는 상관분석을 적용할 수는 없고 회귀분석을 이용하여야 한다.
- * 일반적으로 2개의 특성치 x 와 Y 에 함수관계가 있을 때, x 와 Y 간에 회귀관계가 있다고 말하고, 이 함수를 회귀함수, 또는 줄여서 회귀라고 한다.

- * 이 회귀를 취급하는 해석법이 회귀분석으로서, 이것도 2개의 특성간의 관계를 취급하는 單회귀분석과, 3개 이상의 특성간의 관계를 취급하는 重회귀분석이 있다. 또한 회귀분석에서는 곡선회귀에 관한 분석도 있다.

이와 같이 회귀분석에서는 1개의 특성 y 를 다른 특성 x 로써 설명하는 것이 목적이며, 전자의 y 를 종속변수(또는 목적변수, 결과변수), 후자의 x 를 독립변수(또는 설명변수, 원인변수)라고 한다.

3.2 직선회귀 (단회귀)

3.2.1 직선회귀의 기초 개념

- * 변수 x 를 독립변수, y 를 종속변수라고 하면 그들의 관계는 다음과 같이 주어질 수 있다.

$$y = f(x)$$

- * x 에 대한 y 의 변화가 규칙적인 경우의 함수는 다음 수식으로 표시된다. 즉,

$$y = a + bx$$

- * 이러한 변량 사이의 관계를 표시하는 방정식을 회귀방정식이라 하고, 이 방정식이 그리는 선을 회귀선(regression line)이라 한다.

- * 그것이 직선인 경우에는 직선회귀라 하고, 곡선인 경우에는 곡선회귀라 한다.

- * 변수 x 와 변량 y 에 다음의 관계가 가정될 때 x, y 를 직선회귀의 문제로 다룰 수 있다.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

여기서, α : 상수($x_i = 0$ 일 때의 y 의 값, 즉 y 축의 절편)

β : 모회귀계수(기울기)

ε_i : 오차로서 평균치 0, 분산 σ^2 을 가지는 x_i 와는 독립된 변량

- * 지금 α 의 추정치를 a , β 의 추정치를 b 라고 하면 $y = a + bx_i$ 로 되나, 실제로는 오차가 있으므로 다음 식과 같이 된다.

$$y_i = a + bx_i + e_i$$

3.2.2 회귀직선의 추정

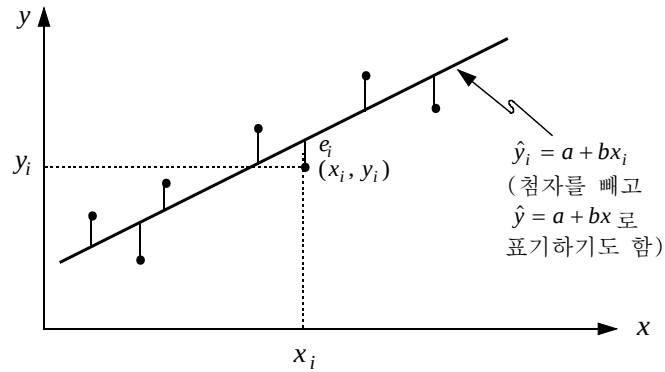
(1) 최소제곱법(MLS)에 의한 직선회귀 추정 ($y_i = a + bx_i + e_i$ 형)

- * 측정치 y_i 에서 일정수를 뺀 값인 편차의 제곱합

$$Q = \sum_{i=1}^n (y_i - a)^2 \quad (3.10)$$

을 생각하여 이 Q 를 최소로 하는 a 의 값은 $a = \bar{y}$ 가 될 때이다. 즉, y 의 편차제곱의 합이

$$S_{yy} = \sum (y_i - \bar{y})^2 \text{ 이 된다.}$$



[그림 3.3] 직선회귀의 오차

- * 지금 a 대신에 y 와 x 사이의 관계식 $y = a + bx$ 를 생각할 때 실제로 여러 요인에 의해 오차(잔차) e_i 가 발생하므로 y 의 추정치인 $(a + bx_i)$ 를 사용하여 [그림 3.3]에서와 같이 측정치 y_i 와 y 의 추정치 $\hat{y}_i$ 의 차이인 오차 e_i 의 제곱의 합

$$S_{y \cdot x} = Q = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - (a + bx_i))^2$$

을 생각하여 이 Q 가 최소가 되게 하면 x 로부터 y_i 를 추정할 때 가장 산포가 적은 y_i 의 추정치를 얻는 직선을 구할 수 있다. 즉, Q 가 최소가 되게 하는 a, b 를 구하면 된다.

이 방법이 최소제곱법(MLS : Method of Least Squares)으로 회귀를 구하는 방법이다.

- * 이것을 구하자면 $\partial Q / \partial a$ 와 $\partial Q / \partial b$ 에 의한 편미분으로 얻어지는 식을 0이라 두고 a, b 에 대해 풀면 된다.

$$\left. \begin{aligned} \partial Q / \partial a &= -2 \sum \{y_i - (a + bx_i)\} = 0 \\ \partial Q / \partial b &= -2 \sum x_i \{y_i - (a + bx_i)\} = 0 \end{aligned} \right\} \quad (3.11)$$

이므로, 이 식을 정리하면

$$\left. \begin{aligned} \sum y_i &= na + b \sum x_i \\ \sum x_i y_i &= a \sum x_i + b \sum x_i^2 \end{aligned} \right\} \quad (3.12)$$

- * 이 두 개의 방정식(정규방정식이라고 함)으로부터 정수항 a 와 회귀계수 b 를 구하면

$$a = \frac{\sum y_i - b \sum x_i}{n} = \bar{y} - b\bar{x} \quad (3.13)$$

$$\begin{aligned} b &= \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{\sum x_i y_i - (\sum x_i \sum y_i) / n}{\sum x_i^2 - (\sum x_i)^2 / n} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\ &= \frac{S_{(xy)}}{S_{(xx)}} \end{aligned} \quad (3.14)$$

가 되므로

$$\hat{y}_i = a + bx_i \quad (3.15)$$

에 대입하여 추정 회귀방정식을 구한다. 여기서, a 는 $x=0$ 일 때의 y 의 값으로서 절편, 또 b 는 회귀선의 기울기를 나타내며 회귀계수라 한다.

* 이 추정 회귀방정식에서 a 를 제거하고 달리 표현하면 다음 식으로 변형될 수 있다.

$$\begin{aligned} \hat{y}_i &= a + bx_i \\ &= a + b\bar{x} + b(x_i - \bar{x}) \\ &= \bar{y} - b\bar{x} + b\bar{x} + b(x_i - \bar{x}) \\ &= \bar{y} + b(x_i - \bar{x}) \end{aligned}$$

여기서 첨자 i 를 빼고 다음과 같은 변형한 식으로 나타낼 수도 있다.

$$\hat{y} - \bar{y} = b(x - \bar{x}) \quad (3.16)$$

여기서, $b, \bar{y}, \bar{x}$ 는 다음을 대입시킨다.

$$b = \frac{S_{(xy)}}{S_{(xx)}}, \quad \bar{y} = \frac{\sum y_i}{n}, \quad \bar{x} = \frac{\sum x_i}{n}$$

* 한편 추정 회귀식을 구하는 방법은 다음의 2가지가 있다.

① x 에 대한 y 의 추정 회귀선

$$\hat{y} - \bar{y} = b(x - \bar{x}) \quad (3.17)$$

윗 식을 정리하면 $\hat{y} = \bar{y} + b(x - \bar{x}) = a + bx$ 의 형태가 된다.

여기서 회귀계수 b 는 $b = S_{(xy)} / S_{(xx)}$ 이다.

② y 에 대한 x 의 추정 회귀선

$$\hat{x} - \bar{x} = b(y - \bar{y}) \quad (3.18)$$

윗 식을 정리하면 $\hat{x} = \bar{x} + b(y - \bar{y}) = a + by$ 의 형태가 된다.

여기서, $b = S_{(xy)} / S_{(yy)}$ 이다.

[참조] 상기 ①, ②에서 일반적으로 독립변수 x 를 가로축에, 종속변수 y 를 세로축에 취해 x 에 대한 y 의 회귀선에 대해 주로 생각한다.

[참조] 추정된(적합된) 회귀식의 성질

- ① $\sum e_i = 0$, 즉 오차의 합은 0이다.
- ② $\sum x_i e_i = 0$, 관측치 x_i 를 가중치로 한 오차의 가중합은 0이다.
- ③ $\sum \hat{y}_i e_i = 0$, 즉 예측치 $\hat{y}_i$ 를 가중치로 한 오차의 가중합은 0이다.
- ④ 적합된 회귀직선은 $(\bar{x}, \bar{y})$ 를 지난다.

(2) 최소제곱법(MLS)에 의한 회귀직선 추정 ($y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ 형)

* 단순회귀식의 적합의 경우로서 단순회귀의 경우에는 회귀선이 직선이 되므로 직선회귀라 부르고, 여기서 얻어지는 직선을 회귀직선이라고 부른다.

* 두 변수 x 와 y 간의 n 개의 데이터 (x_i, y_i) , $i = 1, 2, \dots, n$ 에 관한 직선회귀 모형은 다음과 같은 가정아래 표현된다.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (3.19)$$

여기서, $\varepsilon_i \sim N(0, \sigma^2)$ 이고 서로 독립

β_0, β_1 은 미지의 모수

x_i : i 번째 주어진 고정된 x 의 값, $i = 1, 2, \dots, n$

* 미지의 모수인 β_0 와 β_1 의 추정은 최소제곱법(MLS : Method of Least Squares)에 의하여 구한다. 이 방법은 식 (3.19)의 오차 ε_i 의 제곱, 즉 편차제곱의 합

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2 \quad (3.20)$$

을 최소로 하는 β_0 와 β_1 을 구하는 방법이다.

* 식 (3.20)의 오차제곱합을 최소로 하는 β_0 와 β_1 을 구하기 위하여 S 를 β_0 와 β_1 에 대하여 각각 편미분하여 다음의 결과를 얻는다.

$$\begin{aligned} \frac{\partial S}{\partial \beta_0} &= -2 \sum (y_i - \beta_0 - \beta_1 x_i) \\ \frac{\partial S}{\partial \beta_1} &= -2 \sum x_i (y_i - \beta_0 - \beta_1 x_i) \end{aligned}$$

* 위의 편미분 값을 영으로 만드는 β_0 와 β_1 의 추정치를 $\hat{\beta}_0, \hat{\beta}_1$ 으로 놓고 위의 식을 정리하면

$$\left. \begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i &= \sum y_i \\ \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 &= \sum x_i y_i \end{aligned} \right\} \quad (3.21)$$

을 얻게 되는데, 이 식을 정규방정식(normal equations)이라고 부른다.

* 이 식을 풀어서 얻어진 $\hat{\beta}_0, \hat{\beta}_1$ 을 최소제곱추정치(LSE : Least Squares Estimates)라고 부르고 이것을 풀면 다음과 같이 된다.

$$\left. \begin{aligned} \hat{\beta}_1 &= S_{(xy)} / S_{(xx)} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \end{aligned} \right\} \quad (3.22)$$

$$\text{여기서, } S_{(xx)} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum x_i^2 - (\sum x_i)^2 / n$$

$$S_{(xy)} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - (\sum x_i)(\sum y_i) / n$$

* 따라서 회귀직선의 추정식은 다음 식과 같다.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (3.23)$$

* 이를 다르게 표현하면 $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ 의 관계를 이용하여 다음과 같이 나타내기도 한다.

$$\begin{aligned} \hat{y}_i &= (\hat{\beta}_0 + \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 (x_i - \bar{x}) \\ &= (\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 (x_i - \bar{x}) \\ &= \bar{y} + \hat{\beta}_1 (x_i - \bar{x}) \end{aligned} \quad (3.24)$$

* 첨자 i 를 빼고 회귀직선의 추정식은 다음과 같이 쓰기도 한다.

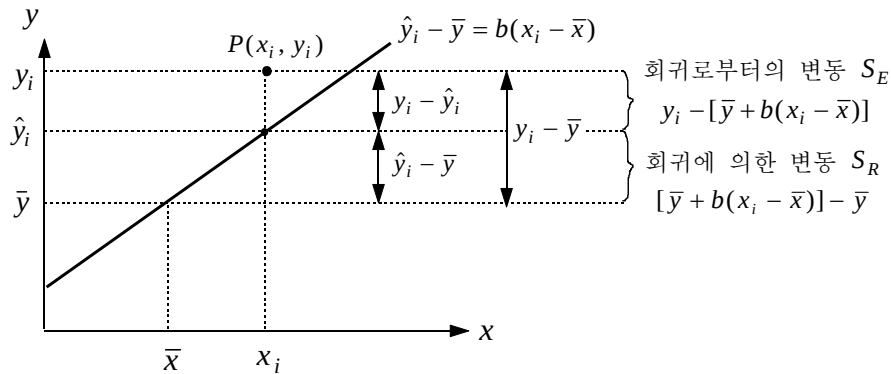
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad \text{또는} \quad \hat{y} = \bar{y} + \hat{\beta}_1 (x - \bar{x}) \quad (3.25)$$

이는 $\hat{y} - \bar{y} = b(x - \bar{x})$ 와 동일한 방법이 된다.

3.2.3 분산분석법 (회귀계수 $\beta = 0$ 의 검정)

(1) 회귀에 의한 변동, 회귀로부터의 변동에 대한 개념

- * 회귀식에서의 회귀계수 β 가 0인지 아닌지를 분산분석으로 검정할 수 있다.
- * 분산분석이란 실험계획법의 개념의 하나로서 「데이터에 관한 편차제곱합(변동)을 몇 개의 의미를 갖는 편차제곱합으로 분산하고 분산된 각 편차제곱합으로부터 산출되는 불편분산의 차의 유의성 검정을 행하는 것」을 말한다.



[그림 3.4] 총편차($y_i - \bar{y}$)의 분해

* [그림 3.4]에서 임의의 점 $P(x_i, y_i)$ 와 $y = \bar{y}$ 와의 거리 $y_i - \bar{y}$ 는

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \quad (3.26)$$

과 같이 $\hat{y}_i$ 를 가감시켜 표시할 수 있다. 이 식에 회귀식인

$$\hat{y}_i - \bar{y} = b(x_i - \bar{x})$$

을 대입하면 다음 식으로 고쳐 쓸 수 있다.

$$y_i - \bar{y} = b(x_i - \bar{x}) + (y_i - \hat{y}_i) \quad (3.27)$$

* 위의 식의 제 1항은 ‘회귀관계로부터 설명되는 편차’이다. 제2항은 ‘회귀관계로 설명되지 않는 편차’ 혹은 잔차(residual, 오차라고도 함), 즉 회귀관계로부터 벗어난 것을 나타내며, 잔차는 e_i 로도 나타낸다.

* 위의 식의 양변을 제곱하여 i 에 대해 총합을 구하면 식 (3.28)과 같이 된다.

$$\begin{aligned}\sum_i (y_i - \bar{y})^2 &= \sum_i \{b(x_i - \bar{x}) + (y_i - \hat{y}_i)\}^2 \\ &= b^2 \sum_i (x_i - \bar{x})^2 + \sum_i (y_i - \hat{y}_i)^2\end{aligned}\quad (3.28)$$

상기 식의 정리 과정에서 제외된 항목인 다음 식은 회귀의 성질을 이용한 것이다.

$$2b \sum (x_i - \bar{x})e_i = 2b[\sum x_i e_i - \bar{x} \sum e_i] = 2b(0 - 0) = 0$$

식 (3.28)의 우변은 y_i 의 $\bar{y}$ 로부터의 총변동이며, 우변의 제1항은 회귀에 의한 변동 S_R 을, 제2항은 회귀로부터의 변동(잔차변동) S_E 를 나타낸다.

* 따라서 식 (3.28)은 다음과 같이 변형시켜 볼 수 있다.

$$S_{(yy)} = b^2 S_{(xx)} + S_E = S_R + S_E \quad (3.29)$$

여기서 S_R 은 회귀에 의한 변동, S_E 는 회귀로부터의 변동을 나타내는데, S_E 는 S_{yx} 로 나타내기도 한다.

(2) 분산분석법

* 총변동의 분해에 따라 총변동(**SST** : Sum of Squared Total)은 회귀에 의한 변동(**SSR** : Sum of Squared Regression)과 회귀로부터 변동(오차변동, 잔차변동, **SSE** : Sum of Squared Error)으로 분해된다.

$$SST(S_{(yy)}) = SSR(S_R) + SSE(S_E) \quad (3.30)$$

* 이를 이용하여 단순회귀분석에서의 분산분석표에서와 같이 하여 분산분석을 실시한다.

3.3 단순회귀분석

3.3.1 단순회귀분석의 순서

* 단순회귀분석은 다음의 순서로 행한다.

- ① 산점도를 작성한다. ② 회귀의 유의성 검정을 한다.
- ③ 회귀선 $\hat{y} - \bar{y} = b(x - \bar{x})$ 을 구한다.
- ④ 모회귀계수 β 의 검정 및 추정을 한다. ⑤ 모회귀직선 η 를 추정한다.
- ⑥ 절편 α 의 검정 및 추정을 한다.

* 단순회귀(직선회귀) 분석의 기초분야에서는 ④, ⑤ 및 ⑥항은 생략되고 분석되기도 한다.

3.3.2 단순회귀분석

(1) 회귀분석을 위한 산점도의 작성

* 산점도를 작성 후 개략적으로 상관관계를 먼저 파악한다.

(2) 회귀의 유의성 검정

① 귀무가설 및 대립가설을 세운다. $H_0 : \beta = 0$, $H_1 : \beta \neq 0$

② 유의수준 α 를 정한다.

③ 분산분석표 작성에 의한 검정통계량의 값(F_0) 계산

요인	SS	DF	MS	F_0	$F_{1-\alpha}(v_R, v_E)$
R	$S_R = b^2 S_{(xx)} = \{S_{(xy)}\}^2 / S_{(xx)}$	1	$V_R = S_R / 1$	V_R / V_E	$F_{1-\alpha}(1, n-2)$
E	$S_E = S_{(yy)} - S_R$	$n-2$	$V_E = S_E / (n-2)$		
T	$S_{(yy)}$	$n-1$			

$$\text{단, } b = \frac{S_{(xy)}}{S_{(xx)}}, S_{(xx)} = \sum x^2 - \frac{(\sum x)^2}{n}, S_{(yy)} = \sum y^2 - \frac{(\sum y)^2}{n}, S_{(xy)} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

④ 기각역 설정. $F_0 > F_{1-\alpha}(1, n-2)$ 이면 H_0 기각

⑤ 판정. H_0 가 기각되면, 즉 모회귀계수 β 가 0이 아니므로 회귀선을 구한다.

(3) 회귀선을 구함

* 회귀선은 다음과 같이 구한다.

$$\hat{y}_i - \bar{y} = b(x_i - \bar{x}) \quad (3.31)$$

참자 i 를 빼고 $\hat{y} - \bar{y} = b(x - \bar{x})$ 로 표기하기도 한다.

* 여기서, 수치변환에 의해 계산할 때에는

$$X = (x - x_0) \times h, Y = (y - y_0) \times g$$

로 수치변환한 경우 다음 식에 의해 구한 값을 대입하여 구한다.

$$\bar{x} = x_0 + \frac{\sum X}{n} \times \frac{1}{h}, \bar{y} = y_0 + \frac{\sum Y}{n} \times \frac{1}{g}, b = \frac{S_{(xy)}}{S_{(xx)}}$$

(4) 모회귀계수 β 의 검정 및 추정

(가) 검정의 기초 개념

* (3)항에서 b 의 값이 $b = S_{(xy)} / S_{(xx)}$ 에 의해 계산된 일정한 값으로 주어지나 b 는 y 의 선형 결합으로 된 통계량이기 때문에 분포를 가진다.

* 회귀계수 b 는 $E(b) = \beta$, $V(b) = \sigma_E^2 / S_{(xx)}$ 인 모평균과 모분산을 갖는 정규분포를 따르

로, $\hat{\sigma}_E^2 = V_E$ 의 관계로부터 σ_E^2 대신 V_E 를 대입한 다음의 식

$$t = \frac{b - E(b)}{\sqrt{V(b)}} = \frac{b - \beta}{\sqrt{V_E / S_{(xx)}}} \quad (3.32)$$

가 자유도 $\nu = n - 2$ 인 t 분포를 하는 것을 이용해 모회귀계수 β 의 검정 및 추정을 한다.

(나) 모회귀계수 β 의 검정

① 가설 설정. $H_0 : \beta = 0$, $H_1 : \beta \neq 0$

② 유의수준 α 를 정함

③ 검정통계량의 값(t_0) 계산

$$t_0 = \frac{b - \beta}{\sqrt{V_E / S_{(xx)}}} \quad (3.33)$$

④ 기각역 설정. $|t_0| > t_{1-\alpha/2}(\nu)$ 이면 H_0 기각

⑤ 판정. 기각역의 조건을 만족하면 H_0 가 기각되어 $\beta \neq 0$ 이므로, β 의 추정을 할 수 있다.

(다) 모회귀계수 β 의 추정

* $\beta \neq 0$ 의 결과가 나오면 다음 식으로 β 의 추정을 한다.

$$\hat{\beta} = b \pm t_{1-\alpha/2}(n-2) \sqrt{\frac{V_E}{S_{(xx)}}} \quad (3.34)$$

4. 품질경영산업기사 실기 [기출유사 엄선문제]

4.1 상관분석

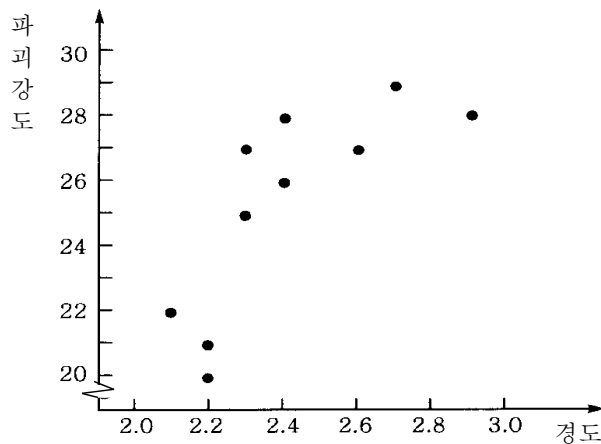
01 어떤 제품에 대한 경도(x)와 파괴강도(y)를 10회 측정해 보니 데이터가 다음과 같다.

번호	1	2	3	4	5	6	7	8	9	10
x	2.4	2.2	2.7	2.6	2.1	2.3	2.9	2.3	2.4	2.2
y	26	20	29	27	22	25	28	27	28	21

(1) 산점도를 그리시오. (2) 상관계수(r)를 구하시오.

해설

(1) 산점도 작성



$$(2) \text{상관계수} : r = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} = \frac{5.57}{\sqrt{0.569 \times 92.1}} = 0.769$$

$$\text{여기서, } S_{(xx)} = \sum x^2 - \frac{(\sum x)^2}{n} = 58.65 - \frac{(24.1)^2}{10} = 0.569$$

$$S_{(yy)} = \sum y^2 - \frac{(\sum y)^2}{n} = 6,493 - \frac{(253)^2}{10} = 92.1$$

$$S_{(xy)} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 615.3 - \frac{24.1 \times 253}{10} = 5.57$$

02 x 와 y 의 시료상관계수 r 을 구하기 위하여 $X = x - 30$, $Y = (y - 5) \times 10$ 으로 데이터 변환을 하여 X 와 Y 의 상관계수 $r' = 0.072$ 를 얻었다. x 와 y 의 시료상관계수 r 의 값은 얼마인가?

해설

☞ $X = (x - x_0) \times h$, $Y = (y - y_0) \times g$ 로 수치변환한 경우

$$r = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} = \frac{S_{(XY)} \times \frac{1}{hg}}{\sqrt{S_{(XX)} \cdot \frac{1}{h^2} \times S_{(YY)} \cdot \frac{1}{g^2}}} = \frac{S_{(XY)}}{\sqrt{S_{(XX)}S_{(YY)}}} = r' = 0.072$$

03 50개의 데이터로부터 $S_{(xx)}=7,064.1$, $S_{(yy)}=3,041.1$, $S_{(xy)}=4,017.3$ 을 얻었다.

(단, $S_{(xx)} = \sum (x_i - \bar{x})^2$, $S_{(yy)} = \sum (y_i - \bar{y})^2$, $S_{(xy)} = \sum (x_i - \bar{x})(y_i - \bar{y})$)

(1) 표본상관계수(r)를 구하시오.

(2) $H_0 : \rho=0$ 을 $H_1 : \rho \neq 0$ 에 대하여 검정하시오. ($\alpha = 0.05$)

해설

$$(1) r = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} = \frac{4,017.3}{\sqrt{7,064.1 \times 3,041.1}} = 0.8667$$

(2) 모상관계수의 상관관계 유무 검정

① 가설 설정 : $H_0 : \rho=0$, $H_1 : \rho \neq 0$

② 유의수준 : $\alpha = 0.05$

$$③ \text{검정통계량의 값}(t_0) \text{ 계산 : } t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.8667 \times \sqrt{50-2}}{\sqrt{1-(0.8667)^2}} = 12.038$$

④ 기각역 설정 : $|t_0| > t_{1-\alpha/2}(n-2) = t_{0.975}(50-2) = t_{0.975}(48) [< t_{0.975}(40) = 2.021]$ 이면 H_0 기각

⑤ 판정 : $|t_0| = 12.038 > t_{0.975}(40) = 2.021 [> t_{0.975}(48)]$ 이므로 유의수준 5%로 H_0 를 기각한다. 즉, 유의수준 5%로 상관관계가 있다고 할 수 있다.

04 다음 데이터의 공분산(V_{xy}), x 와 y 간의 상관계수를 구하시오.

x (cm)	6.8	7.1	6.5	7.8	7.5
y (g)	6.1	6.7	6.3	7.1	7.4

해설

$$(1) \text{공분산 : } V_{xy} = \frac{S_{(xy)}}{n-1} = \frac{1}{n-1} \left[\sum xy - \frac{(\sum x)(\sum y)}{n} \right] = \frac{1}{5-1} \left[240.88 - \frac{35.7 \times 33.6}{5} \right] = 0.244$$

$$(2) \text{시료상관 계수 : } r = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} = \frac{0.976}{\sqrt{1.092 \times 1.168}} = 0.864$$

$$\text{여기서, } S_{(xx)} = \sum x^2 - \frac{(\sum x)^2}{n} = 255.99 - \frac{(35.7)^2}{5} = 1.092$$

$$S_{(yy)} = \sum y^2 - \frac{(\sum y)^2}{n} = 226.96 - \frac{(33.6)^2}{5} = 1.168$$

$$S_{(xy)} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 240.88 - \frac{35.7 \times 33.6}{5} = 0.976$$

05 M공장에서는 포장재를 생산하는데 이 회사의 QC담당자는 제품의 특성인 신율이 품질을 좌우한다는 확신을 갖고 첨가제 x 와 y 의 상관관계를 32개의 시료에서 구한 결과 $r=0.674$ 을 얻었다. 다음에 대하여 유의수준 $\alpha=0.01$ 로 상관관계의 유무를 검정하시오.

해설

모상관계수의 상관관계 유무 검정

① 가설 설정 : $H_0 : \rho=0$, $H_1 : \rho \neq 0$ (양쪽검정) ② 유의수준 : $\alpha=0.01$

③ 검정통계량의 값(t_0) 계산 : $t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.674 \times \sqrt{32-2}}{\sqrt{1-0.674^2}} = 4.997$

④ 기각역 설정 : $|t_0| > t_{1-\alpha/2}(\nu) = t_{1-\alpha/2}(n-2) = t_{0.995}(30) = 2.750$ 이면 H_0 기각

⑤ 판정 : $|t_0|=4.997 > t_{0.995}(30)=2.750$ 이므로 유의수준 1%로 H_0 를 기각한다.

즉, 상관관계는 매우 유의하다고 할 수 있다.

06 백두전자에서는 신제품을 개발하여 공정의 온도(x)가 수율(y)에 미치는 영향을 조사하기 위하여 다음과 같은 데이터를 얻었다. 시료의 상관계수 r 과 공분산을 구하시오.

구분	1	2	3	4	5
온도(x)	20	30	40	50	60
수율(y)	3	4	6	8	9

해설

(1) 시료상관 계수 : $r = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} = \frac{160}{\sqrt{1,000 \times 26}} = 0.99$

$$\text{여기서, } S_{(xx)} = \sum x^2 - \frac{(\sum x)^2}{n} = 9,000 - \frac{(200)^2}{5} = 1,000$$

$$S_{(yy)} = \sum y^2 - \frac{(\sum y)^2}{n} = 206 - \frac{(30)^2}{5} = 26$$

$$S_{(xy)} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 1,360 - \frac{200 \times 30}{5} = 160$$

(2) 공분산 : $V_{xy} = \frac{S_{(xy)}}{n-1} = \frac{160}{5-1} = 40$

4.2 단순회귀분석

◆ 직선회귀 (단회귀) ◆

01 다음의 데이터는 원료의 양(x)과 생성물의 수량(y)의 관계를 나타낸 표이다.

원료(x)	1.5	2.0	3.5	4.3	5.0
수량(y)	30	35	66	66	87

- (1) x 와 y 에 대한 공분산을 구하시오. (2) x 에 대한 y 의 회귀식을 구하시오.
 (3) x 와 y 간의 상관계수를 구하시오.

해설

$$(1) V_{xy} = \frac{S_{(xy)}}{n-1} = \frac{1}{n-1} \left[\sum xy - \frac{(\sum x)(\sum y)}{n} \right] = \frac{1}{5-1} \left[1,064.8 - \frac{16.3 \times 284}{5} \right] = 34.74$$

- (2) x 에 대한 y 의 추정회귀직선식

$$\text{추정회귀직선식인 } \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \text{에 의거 } \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 5.6242 + 15.6981x$$

$$\text{여기서, } \hat{\beta}_1 = \frac{S_{(xy)}}{S_{(xx)}} = \frac{138.96}{8.852} = 15.6981, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 56.8 - 15.6981 \times 3.25 = 5.6242$$

$$\text{단, } S_{(xx)} = \sum x^2 - \frac{(\sum x)^2}{n} = 61.99 - \frac{(16.3)^2}{5} = 8.852$$

$$S_{(xy)} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 1,064.8 - \frac{16.3 \times 284}{5} = 138.96$$

$$\bar{x} = \frac{\sum x}{n} = \frac{16.3}{5} = 3.26, \quad \bar{y} = \frac{\sum y}{n} = \frac{284}{5} = 56.8$$

$$(3) r = \frac{S_{(xy)}}{\sqrt{S_{(xx)} S_{(yy)}}} = \frac{138.96}{\sqrt{8.852 \times 2,274.8}} = 0.9793$$

$$\text{여기서, } S_{(yy)} = \sum y^2 - \frac{(\sum y)^2}{n} = 18,406 - \frac{(284)^2}{5} = 2,274.8$$