

10 장

목 차

01

연관성 분석

02

상관분석

03

공분산과 상관계수

01 연관성 분석

:: **Keywords** 연관성 분석의 정의 | 연관성 분석의 구분



연관성 분석

■ 연관성 분석(association analysis)

조사 대상에서 수집한 자료의 척도를 기준으로

변수들 간에 어느 정도 밀접한 관계가 있는지를 판단하기 위한 분석 방법

→ 자료의 척도를 기준으로 변수 간의 연관성을 파악하기 때문에
척도에 따라 연관성 분석 방법도 달라진다.

연관성 분석

■ 연관성 분석이 필요한 이유

실제의 연구조사에서 단 한가지의 목적으로만 이루어지는 조사는 없음.

→ 사회현상을 하나의 관점이 아닌 다차원적인 시각으로 고찰

∴ 조사 과정 자체에 많은 시간과 비용이 들어가기 때문에

효율성을 높이기 위해서

→ 조사를 여러 변수에 대해 진행한다는 의미

연관성 분석

■ 연관성 분석의 구분

사용된 척도의 종류에 따라 분석방법이 달라짐

- 연속형 척도

다른 변수들의 개입 여부에 따라 편상관분석과 피어슨 상관분석으로 구분

- 범주형 척도

서열척도로 구성된 변수들은 스피어만 서열 상관분석

명목척도로 구성된 변수들은 교차분석

구분	척도	분석 방법	기타 변수의 개입 여부
상관분석	등간척도, 비율척도	편상관분석	○
		피어슨 상관분석	×
	서열척도	스피어만 서열 상관분석	—
교차분석	명목척도	교차분석	—

02 상관분석

:: **Keywords** 상관분석의 정의 | 산포도



광고와 매출은 어떤 연관성이 있을까?

A씨는 은퇴 이후 창업을 생각하고 있다. 창업 자금은 충분히 확보되었지만, 가장 큰 문제는 매출을 어떻게 올릴 것인가이다. 광고를 하면 매출이 오를까? 혹시 비용만 더 많이 발생하는 것은 아닐까?

일반적으로 광고를 하면 매출이 오를 거라 예상하지만, 광고 방법에 따라 긍정적인 효과가 발생할 수도 있고, 반대로 부정적인 효과가 발생할 수도 있다. 따라서 어떤 광고를 해야 효과적인지 미리 알아야 한다. 즉 매출액과 광고비의 상관관계를 파악하는 것이 중요하다.



[그림 10-1] 광고와 매출

상관분석의 정의

■ 상관분석(correlation analysis)

조사 목적에 맞게 구성된 변수들 간의 연관성을 분석하는 방법

→ 상관관계는 두 개의 변수를 기준으로 양의 방향과 음의 방향으로 일정한 규칙이 나타나는 선형관계의 형태와 연관 정도를 수치로 표현한 것.

Ex. 키와 몸무게의 관계, 광고비와 매출액의 관계, 흡연과 수명의 관계 등

산포도

■ 산포도(scatter diagram)

두 개의 변수를 x 와 y 의 그래프로 나타내어, 이들의 분포 정도를 확인

→ 표본이 가지는 중심경향을 파악한 후, 표본이 분포하고 있는 정도를 나타낸 것

→ 산포도를 나타내는 지표
: 분산, 표준편차, 범위, 사분위수, 백분위수

산포도

Note

산포도의 형태



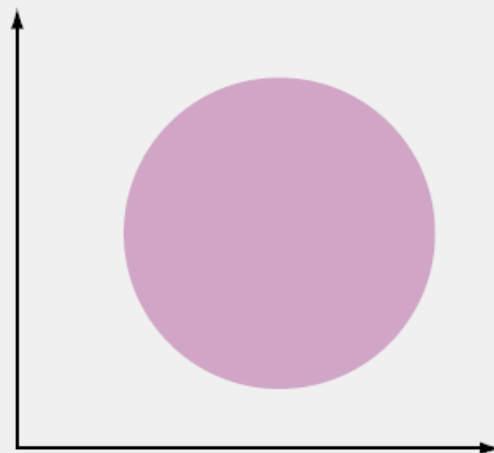
산포도는 [그림 10-2]와 같이 양의 상관관계, 음의 상관관계, 상관관계 없음의 3가지 형태로 구분된다.



(a) 양의 상관관계



(b) 음의 상관관계



(c) 상관관계 없음

[그림 10-2] 산포도의 형태

예제 10-1 산포도 출력

준비파일 | 10장_산포도.xlsx

A주식회사에서는 광고비 지출이 매출에 어느 정도의 영향을 주는지를 파악하기 위해 광고비와 매출액 간의 관계를 조사하였다. 2001년부터 2015년도까지 회사에서 지출한 광고비와 매출액을 조사한 결과가 다음과 같을 때, 광고비와 매출액 간에 어떤 연관성이 있는지 Excel을 이용하여 산포도를 구하라.

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

산포도 (예제 Excel 풀이)

10장_산포도 - Excel

파일 삽입 페이지 레이아웃 수식 데이터 검토 보기 ACROBAT POWERPIVOT

피벗 테이블 피벗 테이블 표 그림 온라인 그림 일러스트레이션 스토어 내 앱 추천 차트 피벗 차트 맵 Power View 꺾은선형 열 승패 (W) 슬라이더 시간 표시 막대 하이퍼링크 텍스트 기호

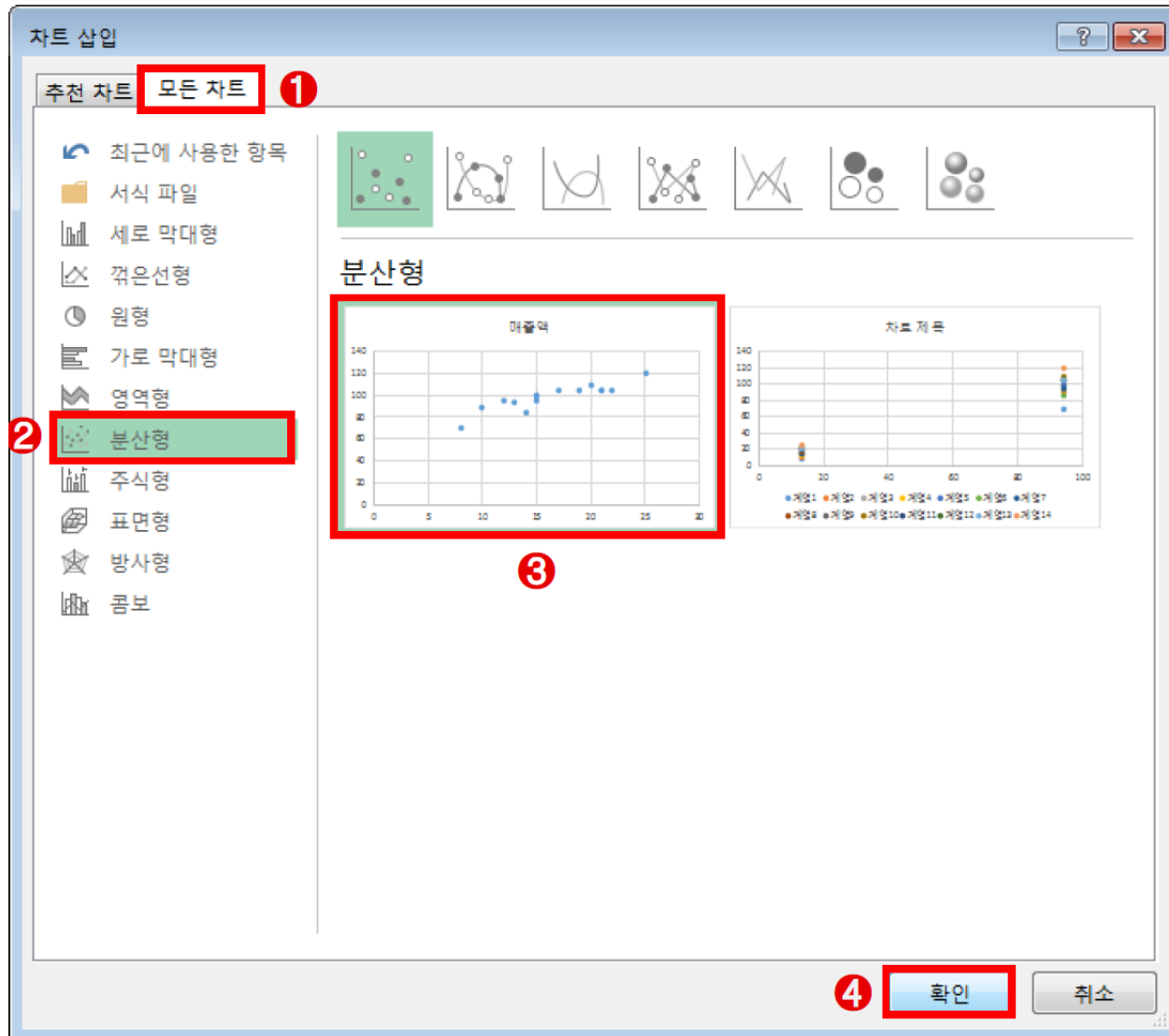
B3 : 광고비

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																
2	연도	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
3	광고비	13	8	10	15	12	15	14	15	17	19	20	21	22	21	25
4	매출액	94	70	90	100	95	100	85	95	105	105	110	105	104	105	121
5																
6																
7																
8																
9																

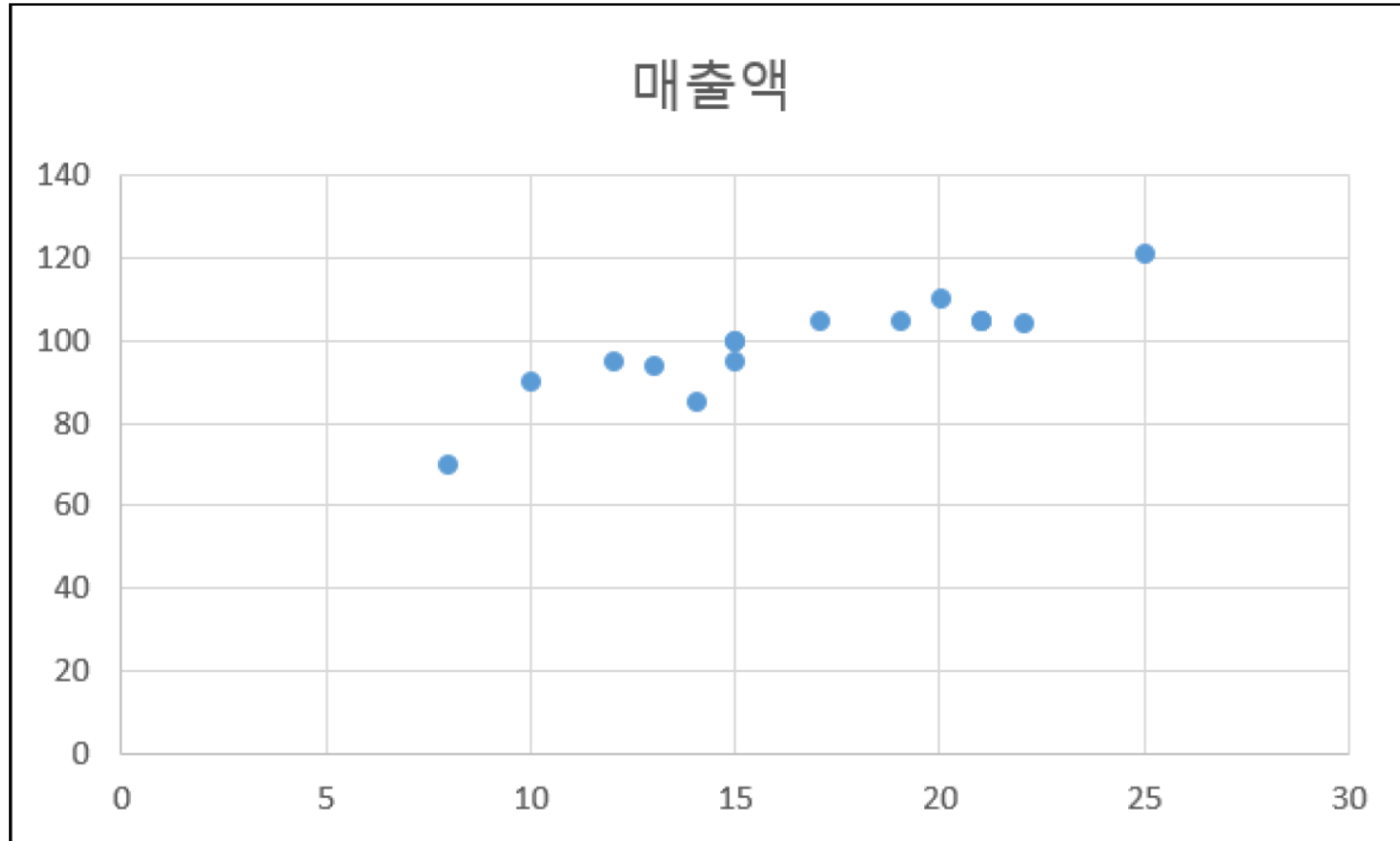
산포도

준비 평균: 57.7 개수: 32 합계: 1731 130 %

산포도 (예제 Excel 풀이)



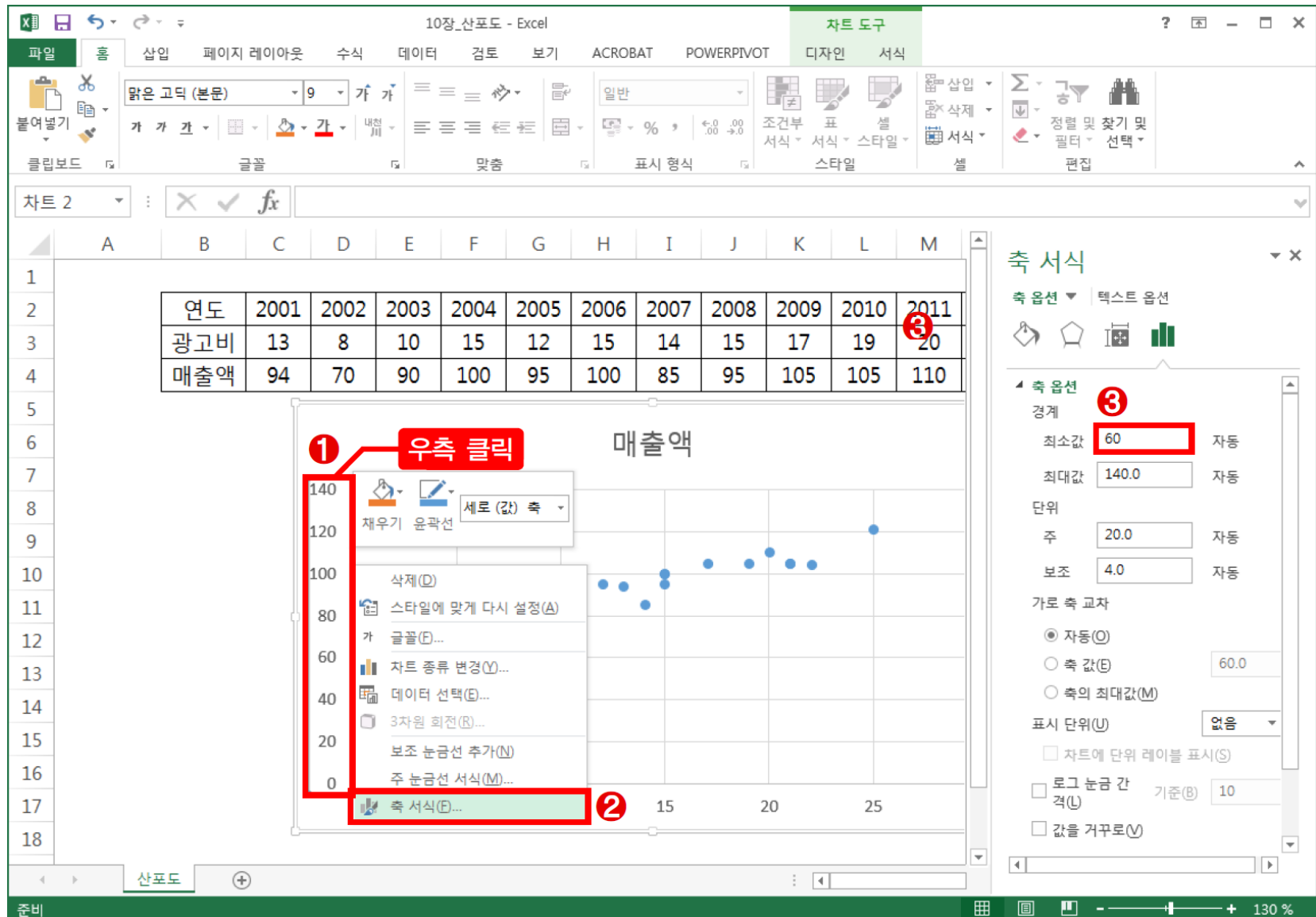
산포도 (예제 Excel 풀이)



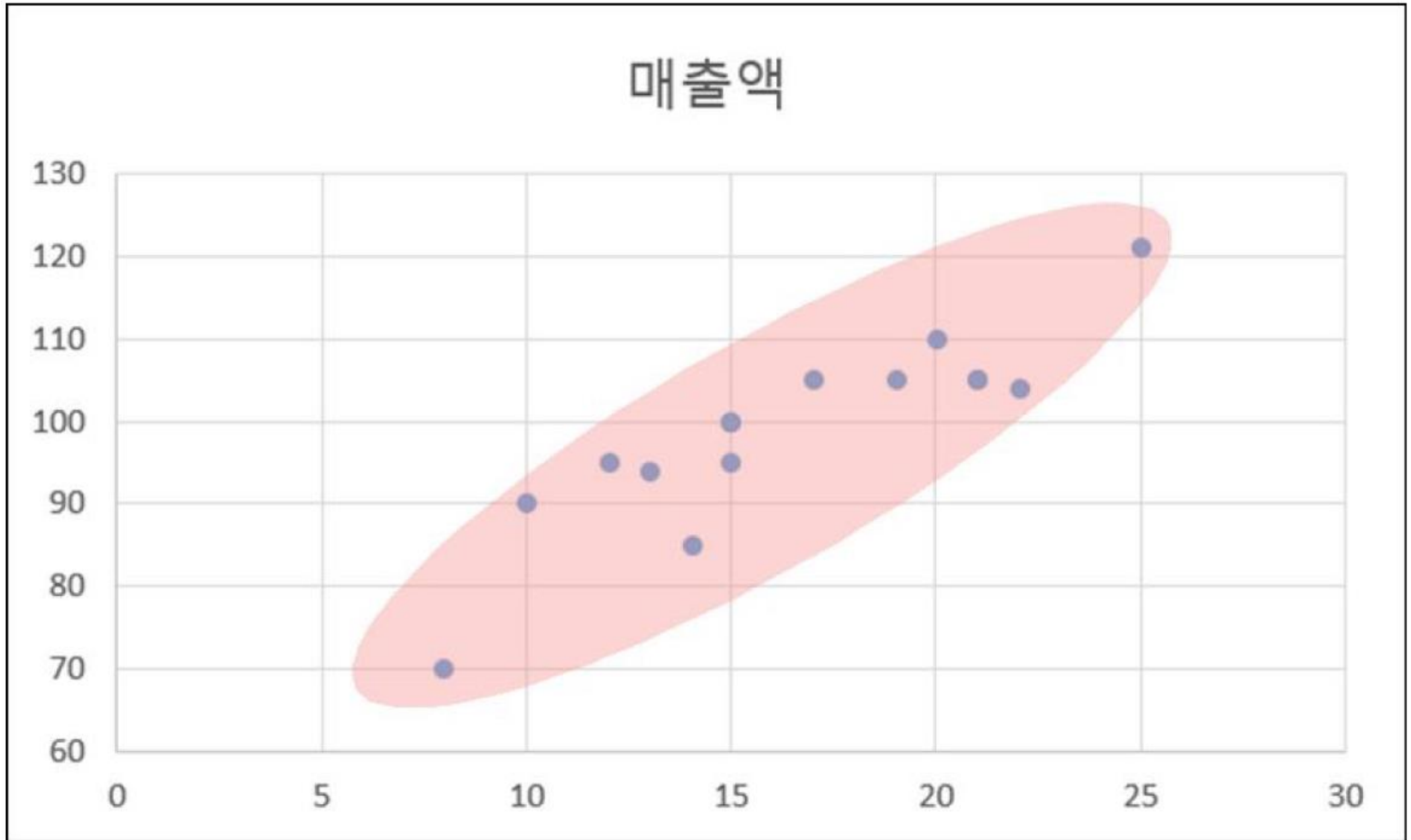
[그림 10-5] 차트 출력

산포도에서 세로축의 하한값이 70이므로, 그 이하의 공간에는 점이 표시되지 않아, 이 결과로는 산포의 방향을 정확하게 파악하기 어려움.

산포도 (예제 Excel 풀이)



산포도 (예제 Excel 풀이_완성)



[그림 10-7] 차트 완성

03 공분산과 상관계수

:: Keywords 공분산 | 상관계수



■ 공분산(covariance)

두 개의 확률변수에 대한 흠어짐의 정도가 동일한 방향인 양(+)의 방향인지
혹은 반대 방향인 음(-)의 방향인지를 나타내는 수치

→ 두 변수가 서로 변하는 정도를 수치로 나타낸 것

Ex. x 가 변하면 y 는 어떻게 변하는지를 표현

공분산

확률변수 X 에 대한 흠어짐의 정도를 분산으로 표시가능.

또 다른 확률변수 Y 에 대한 흠어짐의 정도도 분산으로 표시할 수 있다.

→ 이 각각의 분산에 대한 공통점이 공분산

이에 대한 분석 : 공분산 분석

공분산은 $Cov(X, Y)$ 로 표현하며,

$$Cov(X, Y) = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{N}$$

$$\begin{aligned} (\text{공분산}) &= \frac{[((\text{개별 } X \text{ 측정치}) - (X \text{의 평균})) \times ((\text{개별 } Y \text{ 측정치}) - (Y \text{의 평균})) \text{의 총합}]}{(\text{조합을 이루는 개수})} \\ &= \frac{[(X \text{의 평균편차}) \times (Y \text{의 평균편차}) \text{의 총합}]}{(\text{조합을 이루는 개수})} \end{aligned}$$

예제 10-2 공분산 계산

준비파일 | 10장_공분산.xlsx

[예제 10-1]의 광고비와 매출액의 관계를 확인하기 위해 공분산을 구하라.

공분산 (예제 풀이)

공분산을 구하기 위해 광고비와 매출액에 대한 평균을 구하면

$$\overline{x_{\text{광고비}}} = \frac{13 + 8 + \cdots + 21 + 25}{15} = 16.467$$
$$\overline{x_{\text{매출액}}} = \frac{94 + 70 + \cdots + 105 + 121}{15} = 98.933$$

광고비의 편차와 매출액의 편차를 계산하면

연도별 광고비-광고비 평균	연도별 매출액-매출액 평균	연도별 광고비편차×연도별 매출액 편차
-3.467	-4.933	17.103
-8.467	-28.933	244.976
-6.467	-8.933	57.77
-1.467	1.067	-1.565
-4.467	-3.933	17.569
-1.467	1.067	-1.565
-2.467	-13.933	34.373
-1.467	-3.933	5.77

⋮

⋮

⋮

공분산 (예제 풀이)

각 연도별 광고비 편차와 연도별 매출액 편차를 곱한 값을 모두 더하면

$$17.103 + 244.976 + \cdots + 27.502 + 188.298 = 703.471$$

따라서 공분산은 연도의 개수 15로 나누면

$$\frac{703.471}{15} = 46.898$$

상관계수

■ 상관계수(correlation coefficient)

공분산을 표준화한 값

→ 공분산을 계산하여 결과값을 얻더라도 크기만 짐작할 수 있으며,
계산된 결과로 어떠한 관계인지 정확히 파악하기 힘들.

∴ 공분산의 한계를 극복하기 위해 표준화된 공분산 계수를 상관계수라고 한다.

상관계수

공분산이 $Cov(X, Y) = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{N}$ 일 때,

상관계수는 $Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{\frac{\sum (X - \mu)^2}{N} \cdot \frac{\sum (Y - \mu)^2}{N}}}$

이를 정리하면 $Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}$

$$(\text{상관계수}) = \frac{(\text{공분산})}{(X\text{의 표준편차}) \times (Y\text{의 표준편차})}$$

상관계수

참고 상관계수의 개념

X 와 Y 가 서로 양(+)의 상관관계를 가지면 상관계수는 $0 < \text{Corr}(X, Y) < 1$ 의 값을 가지며, X 와 Y 가 서로 음(-)의 상관관계를 가지면 상관계수는 $-1 < \text{Corr}(X, Y) < 0$ 의 값을 가진다. X 와 Y 가 일정한 규칙 없이 (+)값과 (-)값이 동시에 대응하면 상관계수는 0이다.

상관계수

예제 10-3 상관계수 계산

준비파일 | 11장_상관계수.xlsx

[예제 10-1]의 광고비와 매출액의 관계를 확인하기 위해 상관계수를 구하라.

상관계수 (예제 풀이)

광고비와 매출액이 표준편차를 구하면

$$\sigma_{\text{광고비}} = \sqrt{\frac{(13 - 16.467)^2 + \dots + (25 - 16.467)^2}{15}} = 4.631$$

$$\sigma_{\text{매출액}} = \sqrt{\frac{(94 - 98.933)^2 + \dots + (121 - 98.933)^2}{15}} = 11.404$$

$$\begin{aligned} \text{Corr}(X, Y) &= \frac{\text{광고비와 매출액의 공분산}}{\text{광고비의 표준편차} \times \text{매출액의 표준편차}} = \frac{\text{Cov}(\text{광고비}, \text{매출액})}{\sigma_{\text{광고비}} \times \sigma_{\text{매출액}}} \\ &= \frac{46.898}{4.631 \times 11.404} = 0.888 \end{aligned}$$