분산분석 예제: 새 기계의 생산량 (2)

기계1: 48 49 49 49 50 : 평균 = 49
기계2: 56 55 56 57 56 : 평균 = 56
기계3: 50 51 51 52 51 : 평균 = 51

기계1: 38 39 39 39 40 : 평균 = 39
기계2: 61 60 61 62 61 : 평균 = 61
기계3: 50 51 51 52 51 : 평균 = 51

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

Case 1,2는 그룹 내 분산이 동일하나 Case2가 그룹 간 분산이 큼.

귀무가설이 Case2에서 기각될 가능성이 높음

분산분석 예제: 음료수 색상과 판매량

그룹	무색	분홍	오렌지	초록
1	26.5	31.2	27.9	30.8
2	28.7	28.3	25.1	29.6
3	25.1	30.8	28.5	32.4
4	29.1	27.9	24.2	31.7
5	27.2	29.6	26.5	32.8
평균	27.32	29.56	26.44	31.46
총평균	28.70			

1. 총변동 $SST = \sum_i \sum_j (X_{ij} - \bar{\bar{X}})^2$

$SST = (26.5-28.7)^2 + (28.7-28.7)^2 \cdots + (32.8-28.7)^2 = 115.93$

2. 설명 안된 변동 $SSE = SSW = \sum_j \sum_i (X_{ij} - \bar{X}_j)^2$

$SSW = (26.5-27.32)^2+(28.7-27.32)^2+ \cdots+(32.8-31.46)^2 = 10.68+8.57+13.2+6.6=39.08$

3. 설명된 변동 $SSR = SSB = \sum_i N_i(\bar{X}_i - \bar{\bar{X}})^2$

$SSB = 5(27.32-28.7)^2 + 5(29.56-28.7)^2 +5(26.44-28.7)^2 + 5(31.46-28.7)^2 = 76.85$

분산분석 표 (ANOVA table)

변동의 원천	DF	Sum of Square	분산	F값
그룹 간	r-1	SSB	MSB	$\frac{MSB}{MSW}$
그룹 내	N-r	SSW	MSW	
합	N-1	SST		

1. SST = SSW+SSB

2. 자유도: N-1 = (r-1) + (N-r)
   DF of SST = DF of SSB + DF of SSW

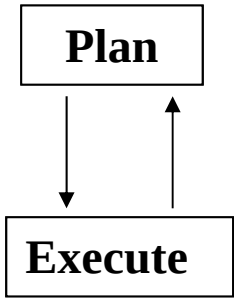
3. 그룹 간 분산(Mean Square Between): $MSB = \frac{SSB}{r-1}$

4. 그룹 내 분산(Mean Square Within): $MSW = \frac{SSW}{N-r}$

5. F-통계량: $F_{r-1,\,N-r} = \frac{MSB}{MSW}$

제곱합 분석 – 유의사항

Problem Solving Questions



1)당신은 무엇을 알고자 합니까?
2)당신이 알기 위해 필요한 것을 어떻게 알아보고자 합니까?
3)어떤 종류의 도구가 당신이 알아보기 위해 필요한 것을 알려 줄 것인가?
4) 어떤 종류의 자료가 선택된 도구에 필요한가?
5)어디에서 당신이 필요한 종류의 자료를 얻을 수 있습니까?

- 당신이 알고자 하는 것 : 어떤 변화 군들이 당신의 spec에 영향을 미칩니까?
- 우리가 알고자 하는 것에 대한 방법 : 백분율 기여도를 보여주는 Bar or Pie chart
- 우리는 사용하는 도구 : 분산분석의 제곱 합 (ANOVA).
- 우리가 필요로 하는 자료 : Y response, cycle time, 잠재 변화 군에 대한 X “설정들”.
- 우리는 어디에서 그 자료를 얻을 수 있나?

제곱합 분석을 이용한 연습 문제

PrtType	PrtArea	PrtShape	LinerTyp	Housing	MatlAge	DamperOC	Response
1	28	1	2	3	8	1	8.25
2	144	2	0	2	75	1	9.66
1	80	3	1	2	44	1	8.33
2	100	2	0	2	75	1	8.82
3	168	3	1	2	30	1	12.06
3	28	1	2	3	4	1	9.67
1	64	2	1	2	45	1	17.51
1	64	2	1	1	16	1	10.79

제곱합 분석을 이용한 연습 문제

위의 파일은 Funacbb.mtw에 들어 있다.

❑ 반응(response)은 용광로에서 나온 조립 완료품에 대한 측정이다.
컬럼에는 특정부품에 대한 데이터가 들어있다.

- PrtType: 조립유형
- prtarea: 부품의 면적
- Prtshape: 부품의 모양 요인
- LinerTyp: 용광로 안쪽의 라이너
- Housing: 조립품을 넣어 이동시키는 컨테이너
- MatlAge: 원자재 구매일로부터의 날짜 수
- DamperOC: 용광로의 환기구가 열림(1), 닫힘(2)

프로젝트에서 부품처리 분야의 특성을 검토하고 있다.
사전에 정해진 조사 또는 실험이 아니라 우연에 의해
얻어진 데이터로 검토하려고 한다.

제곱합 분석을 이용한 연습 문제

Analysis of Variance for Response

Source	DF	SS
PrtShape	2	66.10
Error	86	752.36
Total	88	

$$\frac{66.1}{818.46} = 8.1\%$$

Analysis of Variance for Response

Source	DF	SS
LinerTyp	2	77.36
Error	86	741.10
Total	88	818.46

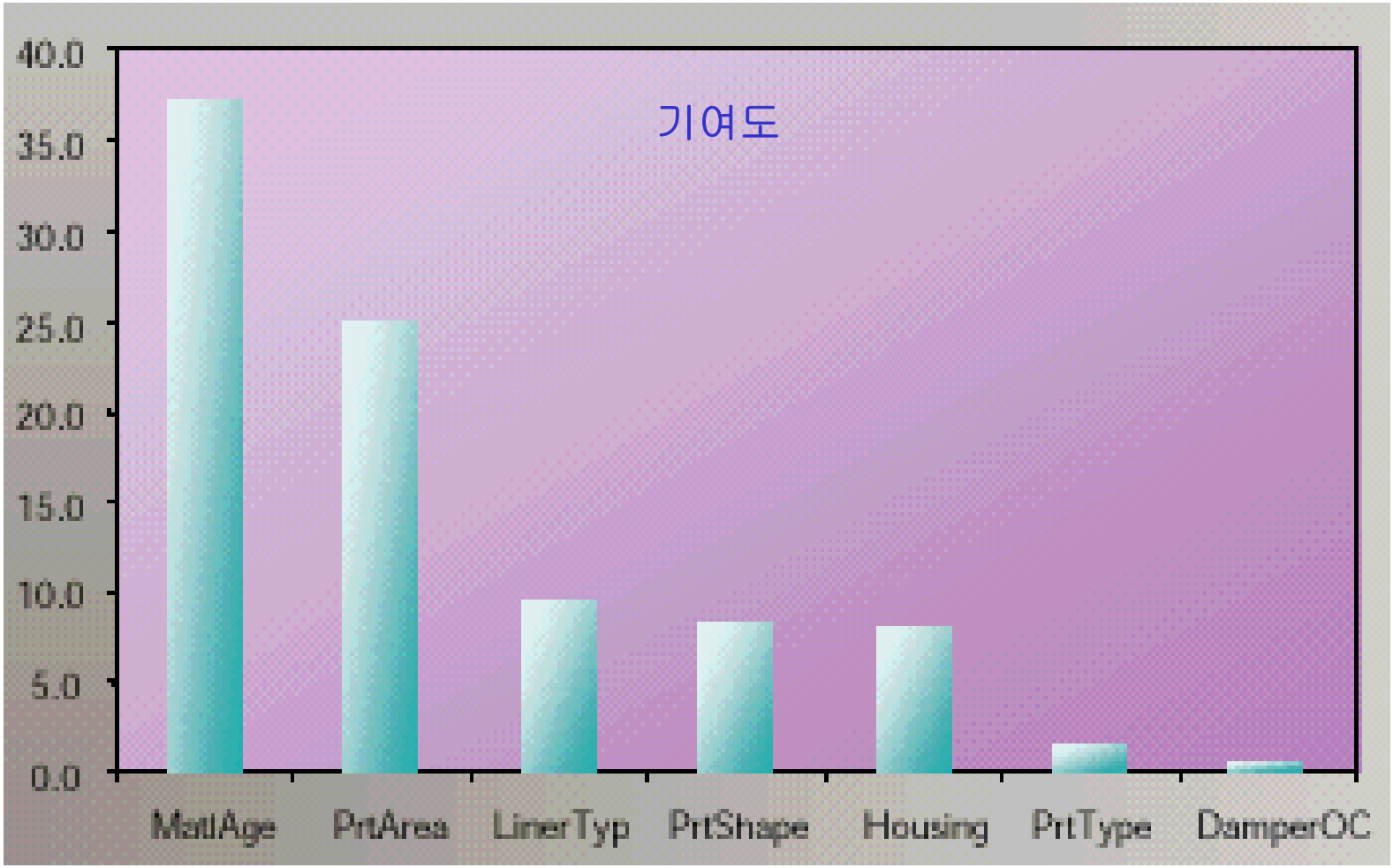
$$\frac{77.36}{818.46} = 9.5\%$$

Analysis of Variance for Response

Source	DF	SS
MatlAge	36	315.4
Error	53	533.5
Total	89	848.9

$$\frac{315.4}{848.9} = 37.2\%$$

제곱합 분석을 이용한 연습 문제



상관과 회귀

참조파일: Scatt39.mtw

Erosion.mtw

상관과 회귀의 목적

- 상관 관계는 두 변수들(Y와 X, 또는 두개의 X) 사이에 관련이 있음을 알려준다.

- 회귀 분석은 관계의 본질을 분명히 하기 위해 사용되며, Y에 대한 더 나은 이해와 관리 가능성을 향상 시킬 수 있도록 Y를 예측할 수 있게 해 준다.

정의

- ***상관:***
  - 변수 출력과 변수 입력 사이의 연관성의 강도를 상관계수=r를 통해 계수화할 수 있는 테크닉.

- ***회귀 등식:***
  - 입력의 값을 사용해서 상응하는 출력에 대한 예측을 할 수 있게끔 해 주는 예측 등식으로서 반드시 선형적일 필요는 없다.

- ***결정계수:***
  - r^2, 회귀 모형의 적합성 또는 회귀 등식에 의해 설명된 변동량을 나타낸다.

상관계수

- ***관련성의 정량화***

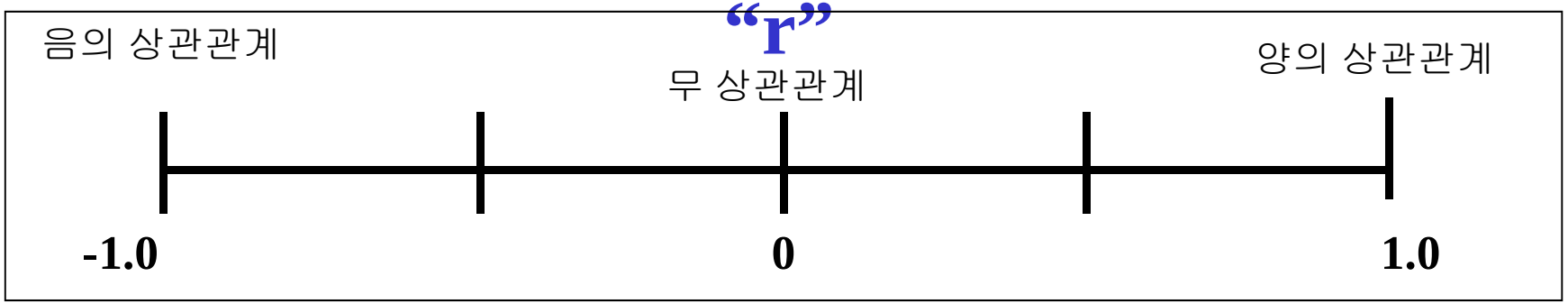
  관련 정도를 측정하는 가장 일반적인 방법: 상관계수 사용
  (Minitab은 Pearson product moment)상관계수를 사용한다).

- **상관 계수(r):**

  항상 –1과 +1 사이다
  한 변수의 값이 증가함에 따라 다른 변수도 증가하면 양(+)
  한 변수의 값이 증가함에 따라 다른 변수의 값이 감소하면 음(–)

- **보통 표본크기를 기준으로 하여**

  만약 $|r| > 0.80$ 이면 관계가 중요하다.
  만약 $|r| < 0.20$ 이면 관계가 유의적이지 않다.



상관 데이터가 갖추어야 할 조건

- 상관 조사를 하기 위한 필요사항:
  - **2변량 데이터**: 이 용어는 변수인 두 개의 데이터를 기술하는데 사용된다.
- 수학적으로 표현하면, 2변량 데이터는 순서쌍으로 이루어져 있다 – 이것들을 x 와 y (x, y)로 부르도록 하자.
- 관례적으로 X를 입력 변수(독립 변수)라고 부르고, Y 를 출력 변수(종속 변수)라고 부른다.

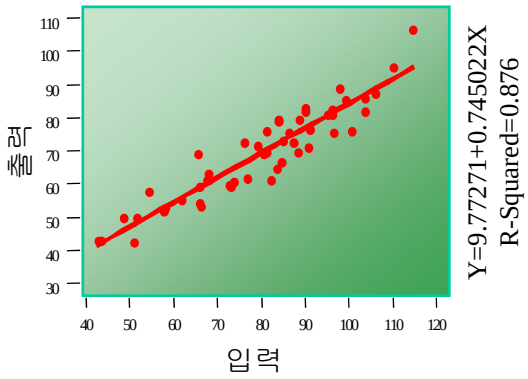
상관 공식

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

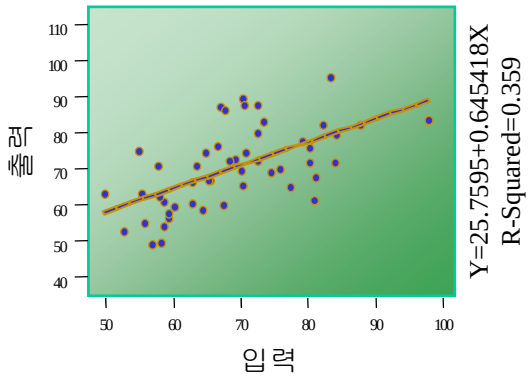
(반응)	(요인)
Y	X
y1	x1
y2	x2
y3	x3
y4	x4
y5	x5
:	:
y_i	x_i
:	:
y_n번째	x_n번째

상관의 유형과 크기

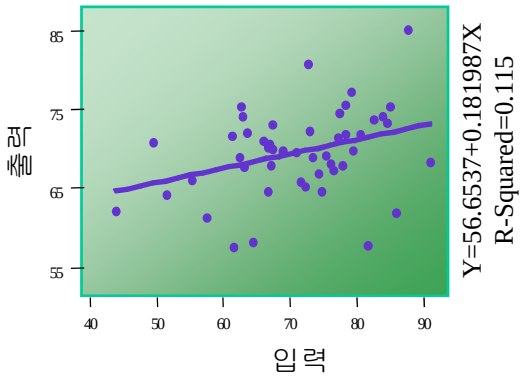
강한 양의 상관관계



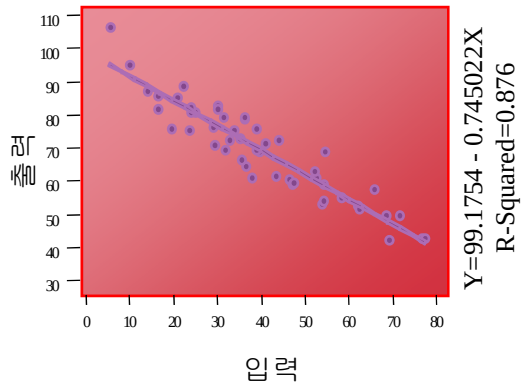
중도적인 양의 상관관계



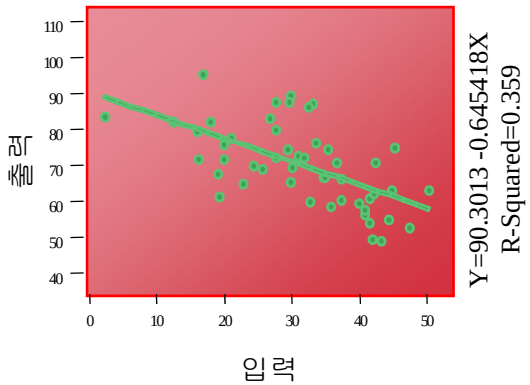
약한 양의 상관관계



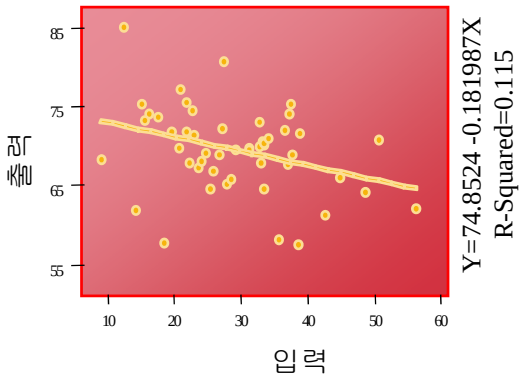
강한 음의 상관관계



중도적인 음의 상관관계



약한 음의 상관관계



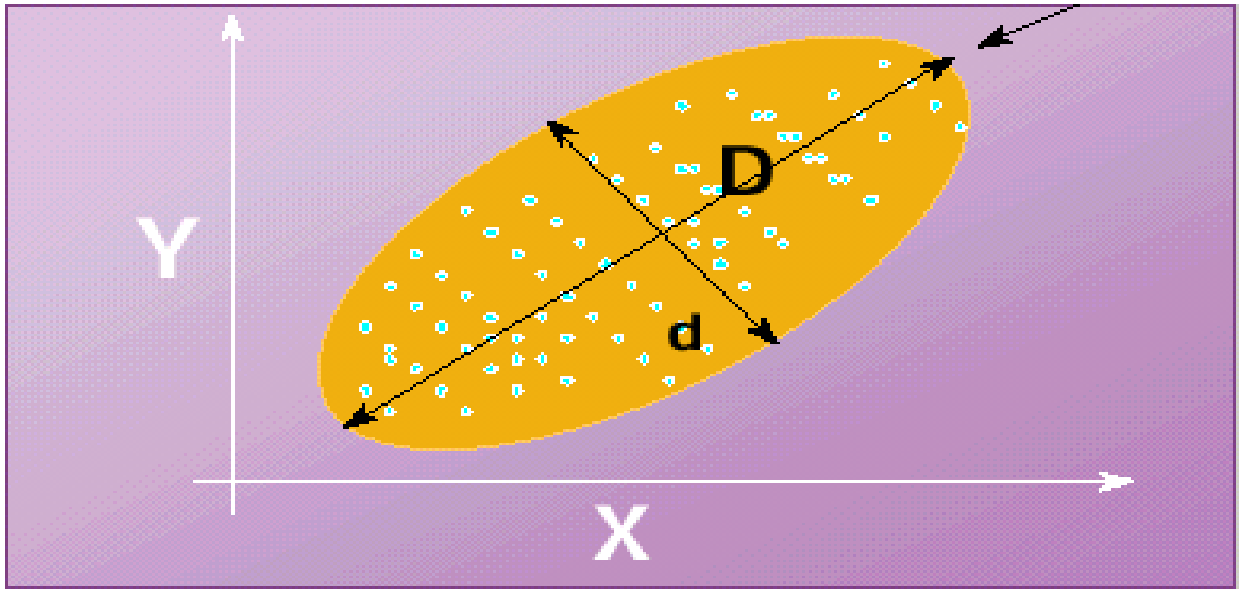
시각적 상관 추정

1)자로 타원형의 최대직경(D)를 측정한다.

2)자로 타원형의 최소직경(d)를 측정한다.

3) “r”값을 다음을 이용하여 계산한다. + / - (1 - d / D)

4)D의 경사방향에 따라 기호를 붙인다.

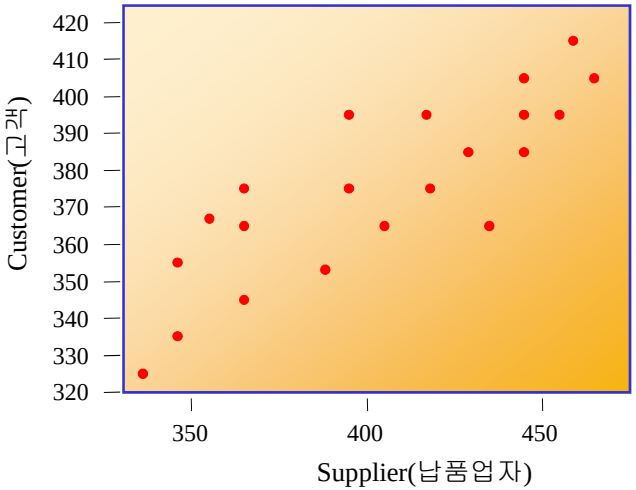


상관의 남용과 오용

- y와 x1 간에 상관이 있다는 것을 입증했다 하더라도, 이것이 반드시 x1의 변동이 y의 변동에 의해서 초래되었다는 것을 의미하지는 **않는다**.
  - x1과 y 둘 다에 변동을 초래하는 제3의 변수가 "숨어" 있을 수 있다.
  - 두 변수 간에 관계가 있다는 결론이 인과관계를 의미하는 **것**은 아니다.

상관은 인과관계를 파악하는 것이 *아니다*!

미니탭의 상관분석



a. **항상** 데이터를 그래프 상에 타점하는 것을 제일 먼저 실시한다.
   참조파일: scatt39.mtw
   그래프 > 산점도

b. 상관분석을 실시한다
   통계학 > 기초통계 > 상관계수

상관(피어슨)

납품업자와 고객의 상관 = 0.834, P-값 = 0.000

참조파일: Scatt39.mtw