

# 회귀분석의 특징

## • 만약 데이터가

- 곡선 형태로 되어 있다면 기울기는 0 이다
- 원(구) 형태로 되어 있다면 기울기는 0 이다
- 직선의 형태로 되어 있지 않다면 분석할 수 없음

각 독립변수들과  
종속변수간의  
상관 관계를 먼저  
확인해야 함.

## ★ 회귀분석 전에 "산포도(scatter plot)을 찍어 보아야 함"

- 직선형태의 데이터 분포가 나타나지 않는 경우 다른 방법을 찾아야 함
- 왜냐하면 결국 회귀분석이란  $y = a + bx$ 의 직선을 구하는 것이므로

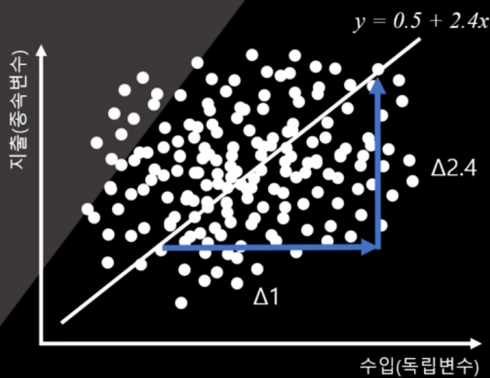
## • 회귀계수(기울기)는 결국 t-test의 평균값 차이와 동일한 개념

## ★ 따라서 회귀계수는 t-test로 그 유의성을 테스트 함

※ 독립변수와 종속변수가 서로 '선형 상관관계'여야지만, 해당 독립변수로  
종속변수를 예측하는 '선형 회귀식'을 구할 수 있음.

## 이 경우 회귀계수는 얼마라고 해야할까?

## • 회귀계수는?



	회귀계수	표준오차	t-value	p-value
절편	0.5	0.34	1.47	0.15
변수 X	2.4	1.9	1.26	0.21

- 회귀계수는 2.4로 나오나
- 유의하지 않으므로
- "0"이라고 본다
- 즉 아무 의미 없음
- 유의하지 않음

즉, '변수X'가  
종속변수에  
미치는 영향력은  
'0'이다.

독립변수가 5개인 경우 종속변수와 상관관계 분석에서는 모두 종속변수와 유의적인 상관관계가 나와도 회귀분석을 하면 대부분 독립변수 5개 중 2-3개 정도만 유의적인 영향력이 있다고 나옵니다. 5개 전부다 유의적으로 나오면 논문 조작 가능성이 99%입니다. 이렇게 나오려면 독립변수들간에 상관계수가 거의 0에 가까운 경우만 생깁니다(탐색적 요인분석을 해서 요인점수를 독립변수로 하였을 경우 이런 상황이 벌어집니다).

· 독립변수들 간의 상관관계까지 존재하게 되면, '다중공선성' 때문에 예측력이 떨어지게 됨.

· 결국 '다중공선성' 때문에, '2-3'개의 독립변수만 유의적인 영향력이 있다고 나옴.