

3.3 회귀분석

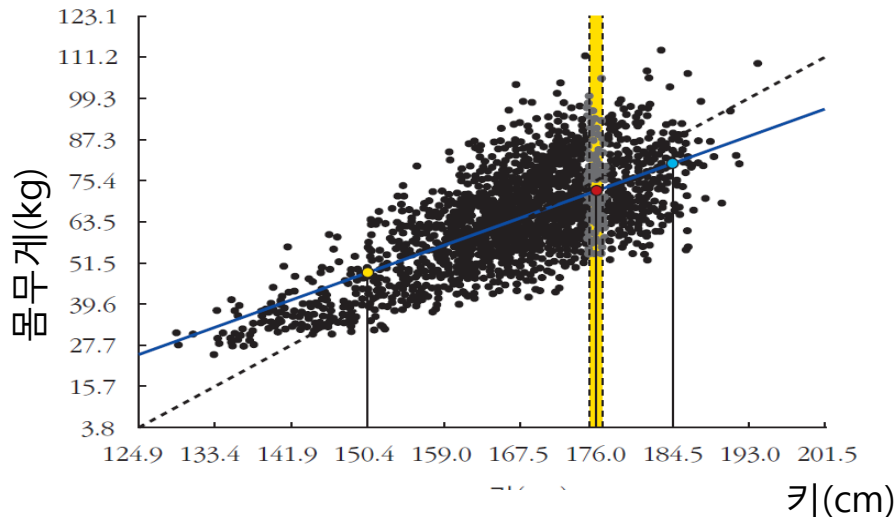
- 1) 변수간의 관계
- 2) 평균의 그래프
- 3) 회귀분석방법
- 4) 회귀효과

1. 변수간의 관계

회귀분석

- 회귀분석(regression analysis)은 집단별 평균을 분석하는 통계적 방법
- 집단을 구분하는 분류지표가 한 개인지, 둘 또는 그 이상인지에 따라 단순회귀분석과 중회귀분석으로 나누어짐

산포도



주: 실선은 회귀직선이고 점선은 표준편차선이

다.

류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 121

상관계수=0.67 :

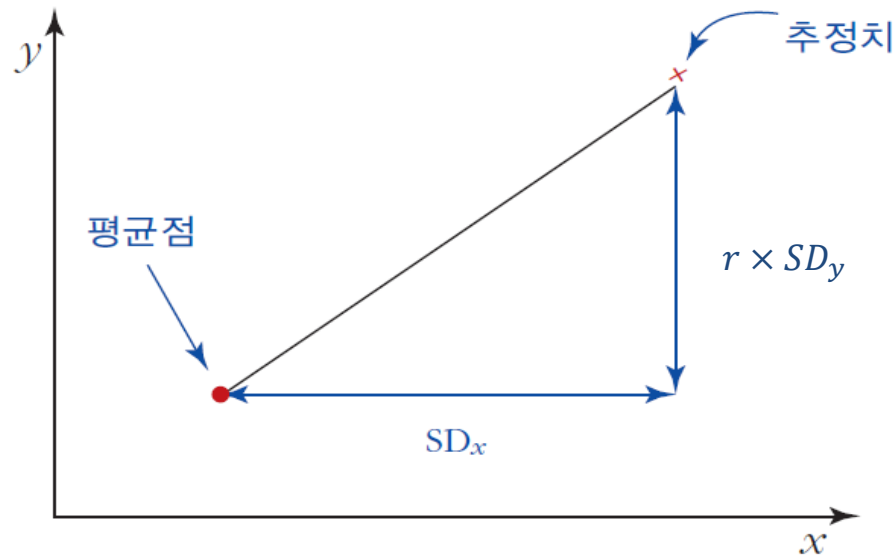
키가 평균보다 $1SD_x$ 만큼 큰 사람들
은 몸무게가 평균보다 대략

$0.67SD_y$ 만큼 더 나감

1. 변수간의 관계

회귀분석

- y 의 x 에 대한 회귀직선은 각각의 x 값에 대응하는 y 의 평균값을 추정



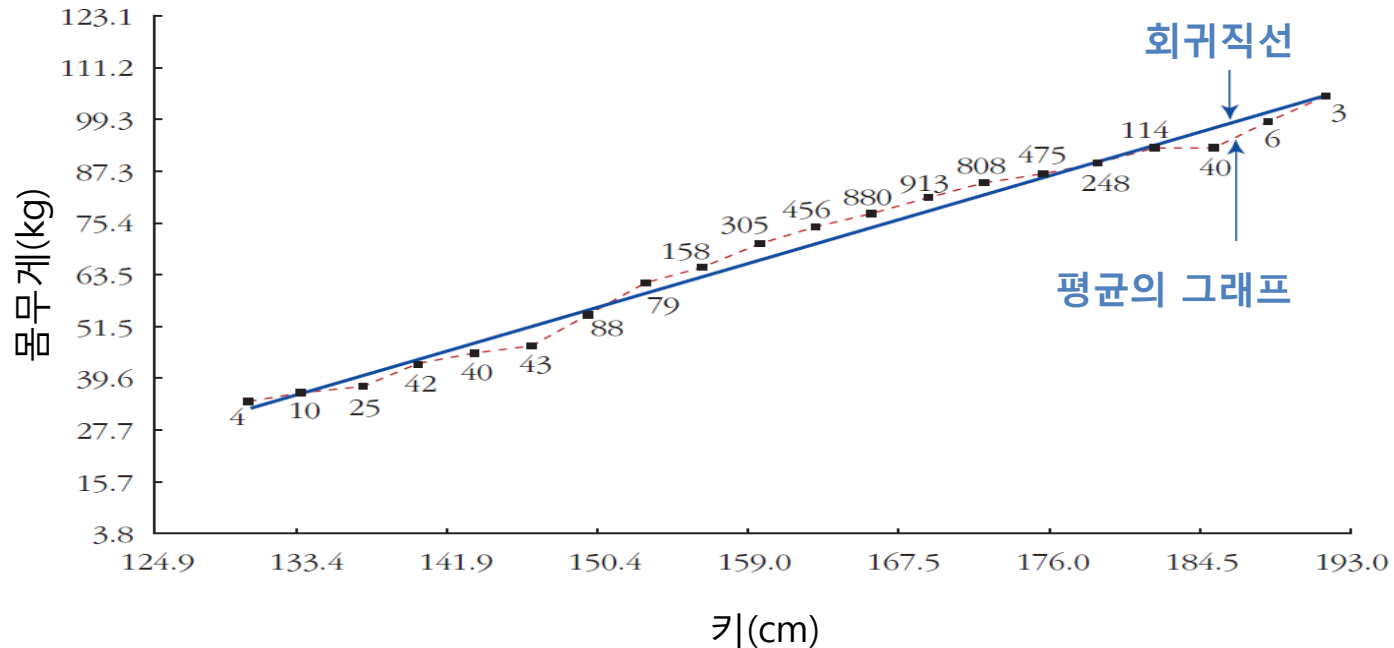
류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 123

- x 값이 x 평균값에서 $1SD_x$ 증가할 때 y 값은 y 평균값에서 $r \times SD_y$ 증가

2. 평균의 그래프

평균의 그래프와 회귀직선

그림 6-3 평균의 그래프와 회귀직선

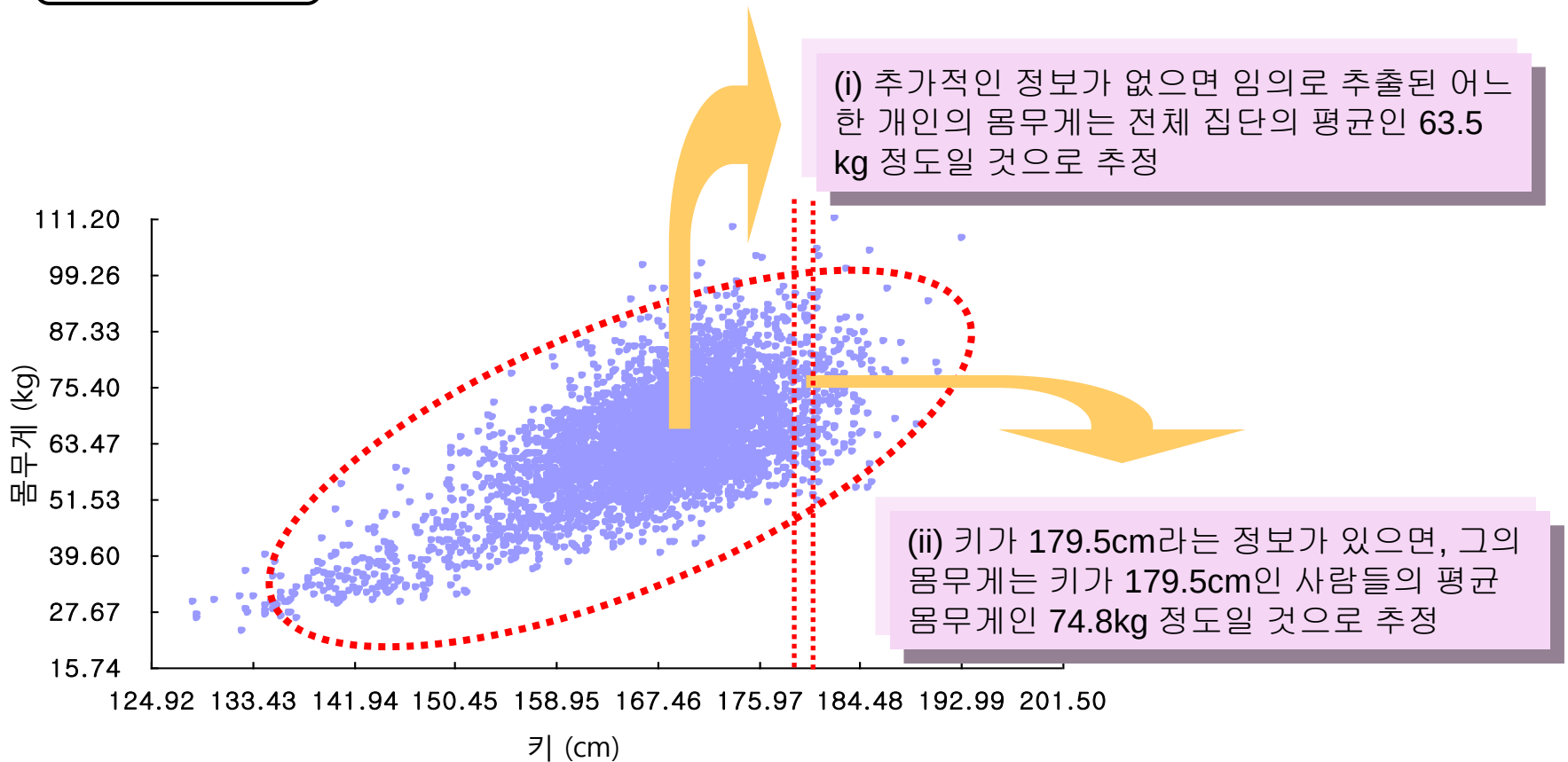


류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 124

- 회귀직선은 평균의 그래프를 하나의 직선으로 근사 시킨 것
- 평균의 그래프가 비선형이면 회귀직선으로의 선형 근사는 부적절

3. 회귀분석방법

몸무게 추정



3. 회귀분석방법

예1) 한 여학생의 학점 예측

예: 경제학부 10학번 학생 100명을 대상으로 조사한 결과

경제원론 평균 = 3.0 표준편차(SD_x) = 0.70

경제통계학 평균 = 3.0 표준편차(SD_y) = 0.60

상관계수 = 0.5

문제) 경제원론 학점이 3.70인 지희의 경제통계학 학점은?

답) 1. 경제원론 점수가 평균보다 $(3.7-3)/0.7=1SD_x$ 높으므로

2. 경제통계학 학점은 평균보다 rSD_y , 즉, $0.5*0.60=0.30$ 만큼 높을 것으로 예측

3. 따라서 경제통계학 학점의 예측치는 $3.0+0.3=3.3$ (B+) 이다.

3. 회귀분석방법

예2) 한 남학생의 백분위 예측

예: 경제학부 10학번 학생 100명을 대상으로 조사한 결과

경제원론 평균 = 3.0 표준편차(SD_x) = 0.70

경제통계학 평균 = 3.0 표준편차(SD_y) = 0.60

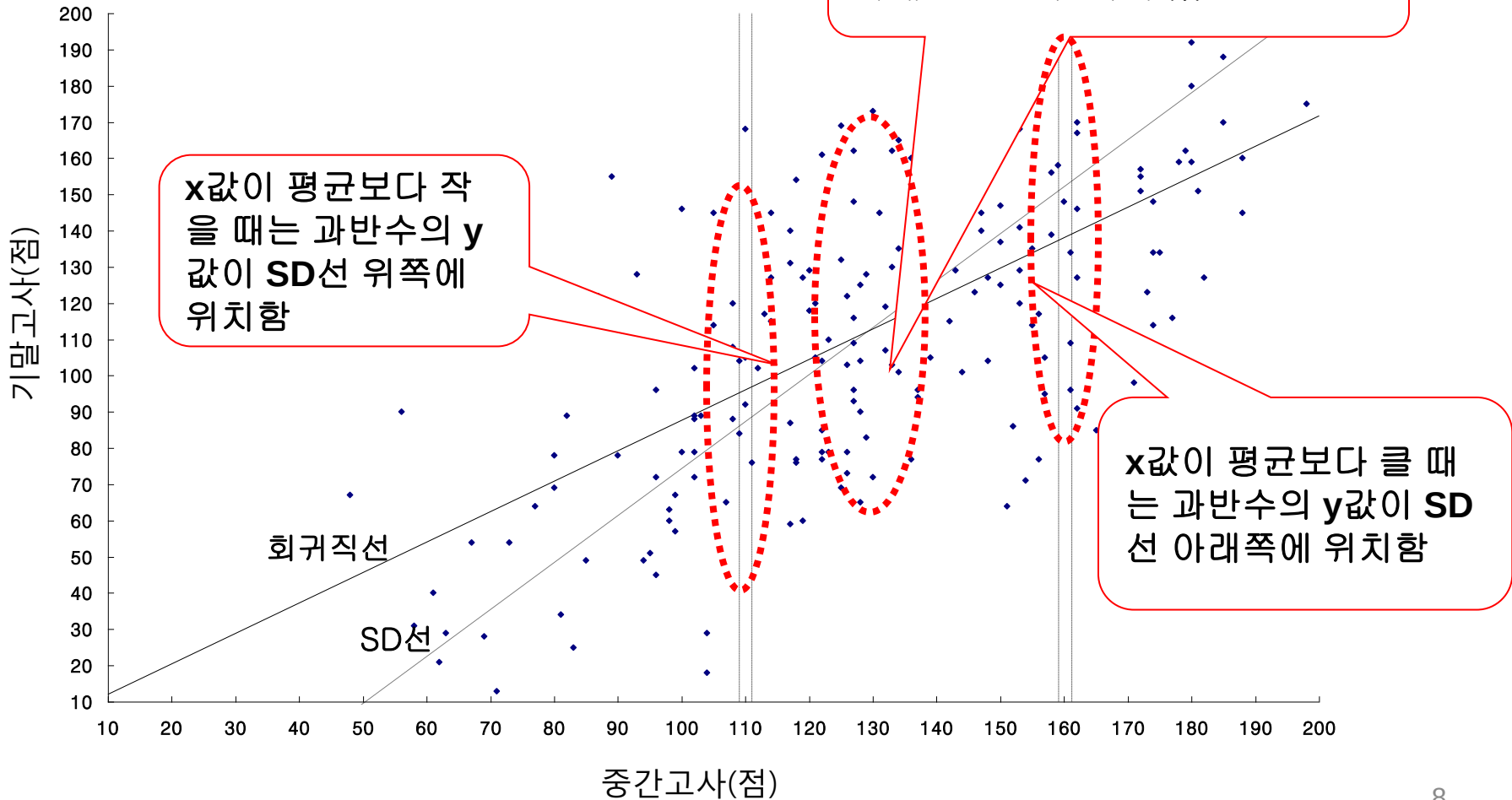
상관계수 = 0.5

문제) 경제원론의 백분위가 90%인 건희의 경제통계학 백분위 예측치는?

- 답) 1. 경제원론 분포가 정규분포를 따를 경우, 건희의 학점은 평균보다 $1.3SD_x$ 높음
2. 경제통계학 학점은 평균보다 $0.5 \cdot 1.3SD_y = 0.65SD_y$ 만큼 높을 것으로 추정됨
3. 이는 백분위로 환산하면 74% 에 해당

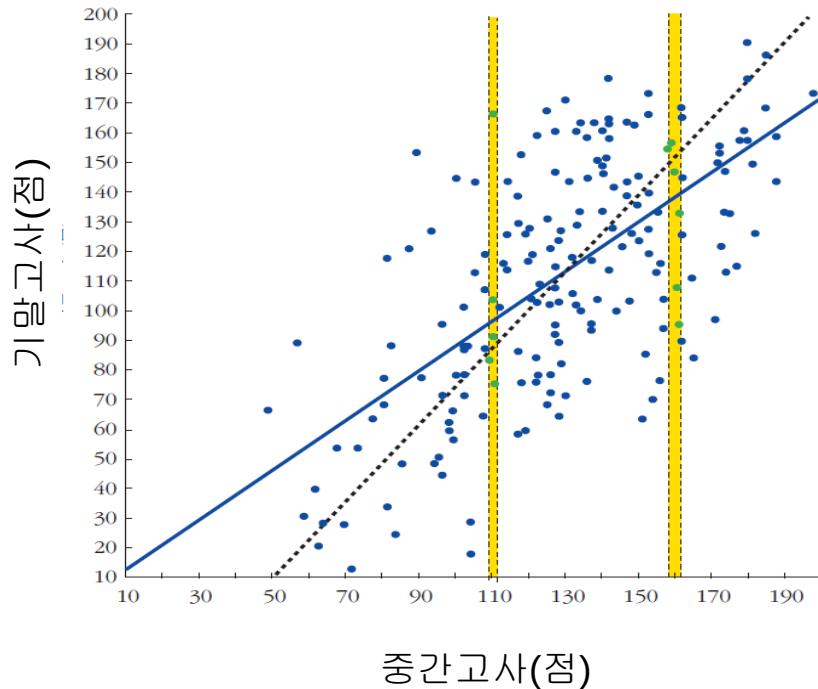
4. 회귀효과

평범으로의 회귀



4. 회귀효과

평범으로의 회귀



중간고사 성적의 평균 = 129.3

표준편차(SD_x) = 30.8

기말고사 성적의 평균 = 112.4

표준편차(SD_y) = 40.0

상관계수 = 0.64

평범으로의 회귀(regression to mediocrity)

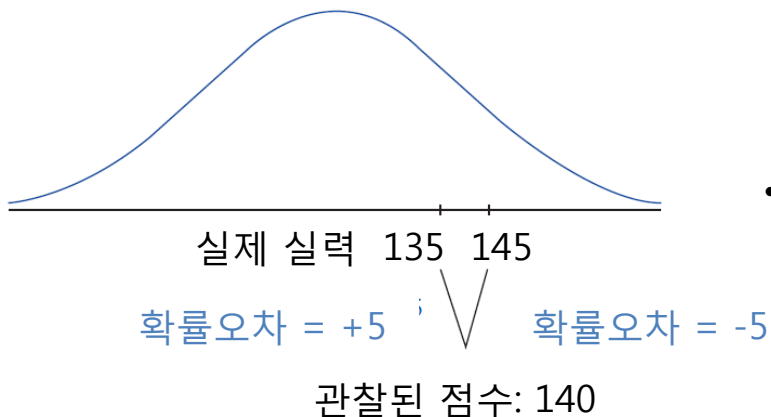
- 중간고사에서 평균보다 높은 160점 받은 학생들의 기말고사 평균 점수는 138점으로 표준단위로 볼 때 하락했음
- 반면 중간고사에서 평균보다 낮은 110점 받은 학생들의 기말고사 평균 점수는 96.1점으로 표준단위로 볼 때 상승했음

주: 실선은 회귀직선이고 점선은 표준편차선이다.

4. 회귀효과

회귀효과

- 회귀오류(regression fallacy)는 회귀효과를 무언가 중요한 효과로 착각하는 것
 - 회귀효과의 배경
 - (관찰된 점수) = (실제 실력) + (확률오차)
 - 예: 실제 실력은 평균 120, 표준편차 15의 정규분포를 따라 분포한다고 가정하자.
 - 또 관측치에 든 확률오차는 각각 0.5의 확률로 ± 5 라고 가정하자.
 - 실제 실력이 135인 사람의 관찰된 점수는 각각 1/2의 확률로 130 또는 140
 - 실제 실력이 145인 사람의 관찰된 점수는 각각 1/2의 확률로 140 또는 150

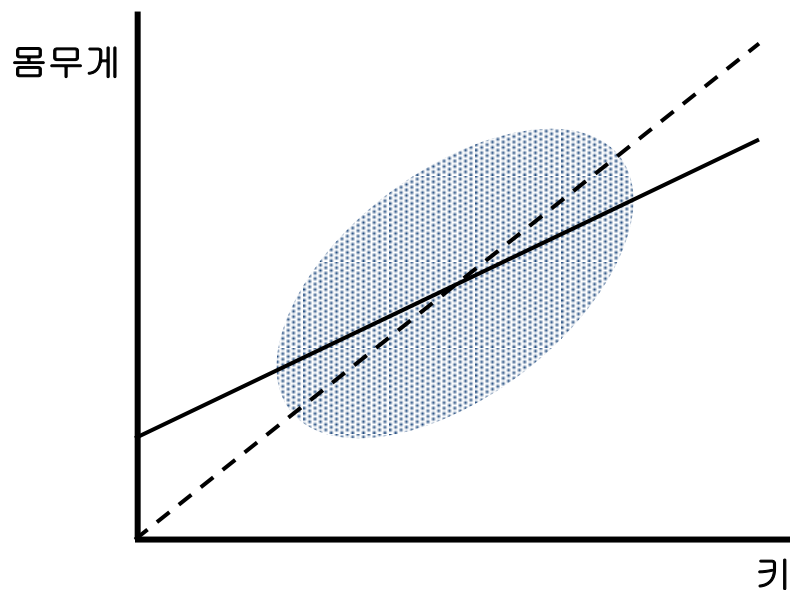


- 관찰된 점수로 140점 받은 사람의 경우 실제 실력은 135인데 확률오차가 +5인 경우가 실제 실력은 145인데 확률오차가 -5인 경우보다 더 가능성 높음
- 따라서 첫 번째 시험점수가 평균보다 높으면, 아마도 실제 실력은 관찰된 점수보다 낮을 가능성이 큼

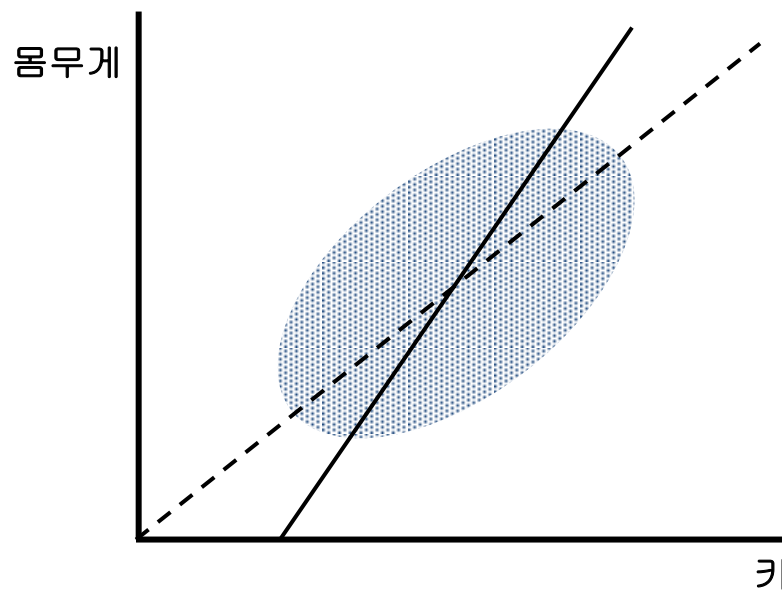
4. 회귀효과

두 회귀직선

몸무게의 키에 대한 회귀직선



키의 몸무게에 대한 회귀직선



단, 실선은 회귀직선, 점선은 표준편차선

류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 133