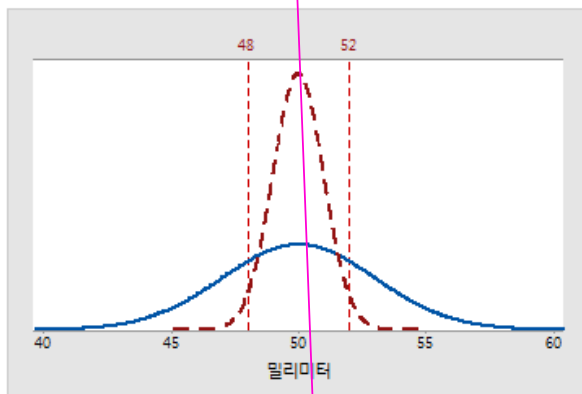


분산의 정의

Minitab 18에 대해 자세히 알아보기

분산은 데이터가 평균 주위에 흩어져 있는 정도를 측정합니다. 분산은 표준 편차의 제곱과 같습니다.

공정 분산을 줄이면 정밀도가 증가하고 결점 수가 감소하기 때문에 분산을 모니터링하는 것은 제조 및 품질 관리 분야에서 필수적입니다. 예를 들어 한 공장에서 50mm 길이의 못을 제조하는데, 못의 길이가 목표값 50mm로부터 2mm 내에 있으면 규격을 충족한다고 가정합니다. 공장에서 두 가지 종류의 기계를 사용하여 못을 제조합니다. 두 기계 모두 길이가 정규 분포를 따르고 평균 길이가 50mm인 못을 제조합니다. 그러나 각 기계에서 생산된 못의 분산이 다릅니다. 아래 그림에서 ~~표준 편차~~로 표시된 A 기계는 분산이 9mm²인 못을 제조하고, ~~표준 편차~~로 표시된 B 기계는 분산이 1mm²인 못을 제조합니다. 각 기계에서 생산된 못의 길이 분포가 세로 선으로 표시된 규격 상한과 하한을 따라 포개져 있습니다.



못 길이의 분포

A 기계에서 생산된 못 길이의 분산이 B 기계에서 생산된 못 길이의 분산보다 큼니다. 따라서 A 기계에서 생산된 못이 규격 한계를 벗어날 가능성이 B 기계에서 생산된 못보다 큼니다.

분산이 0이다. → 데이터 간 차이가 없다.



분산이 크다. → 데이터 간 차이가 크다.



세 집단의 평균 키가 비슷하다 → 분산이 작다

세 집단의 평균 키가 다르다 → 분산이 커진다

↑ 집단 간 분산





집단내 분산=0

→

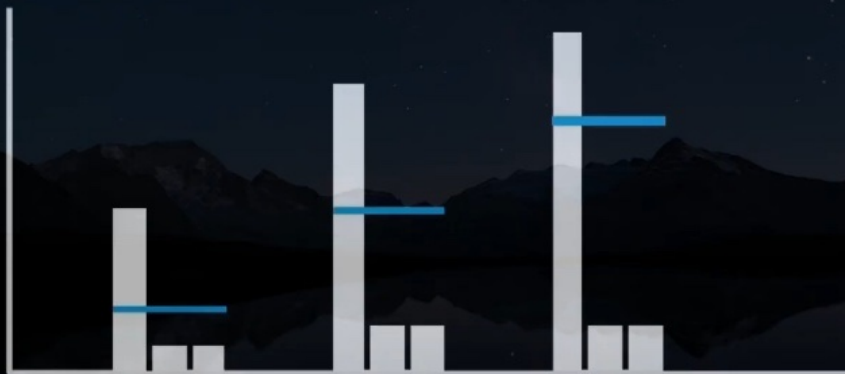
완벽한 집단들이다.



집단내 분산이 크다?

→

집단이라고 볼 수 없다.



$$\text{집단 간 차이} = \frac{\text{집단 간 분산}}{\text{집단 내 분산}}$$

→ 집단 간 평균 차이가 큰가요?
~~※~~ 애초에 집단이라고 할 수 있나요?

~~※~~

$$\text{집단 간 차이} = \frac{\text{집단 간 분산}}{\text{집단 내 분산}} = F \text{ 값}$$

등분산성(Homoscedasticity) 가정



“집단 간 집단내 분산이 서로 비슷한가?”

비교할 수 있는 상태인가?

세 집단의 집단 내 분산들이 다르다 → 서로 비교할 수 “없”다
세 집단의 집단 내 분산들이 비슷하다 → 서로 비교할 수 “있”다



집단내 분산이 서로 다르다? → 동등하게 비교하기 어렵다



등분산성의 가설

H_0 : 모든 집단의 집단내 분산이 같다.



H_a : 모든 집단의 집단 내 분산이 같지는 않다.



○ 가정 검정 (등분산성) ↖ ANOVA 전, '가정검정'으로 실시.

※ 분산분석은 반드시 두 번의 절차를 거쳐야 한다.

1. Omnibus test ↖ 첫 번째 실시 (ANOVA)

2. post-hoc test ↖ 두 번째 실시 (사후검정)