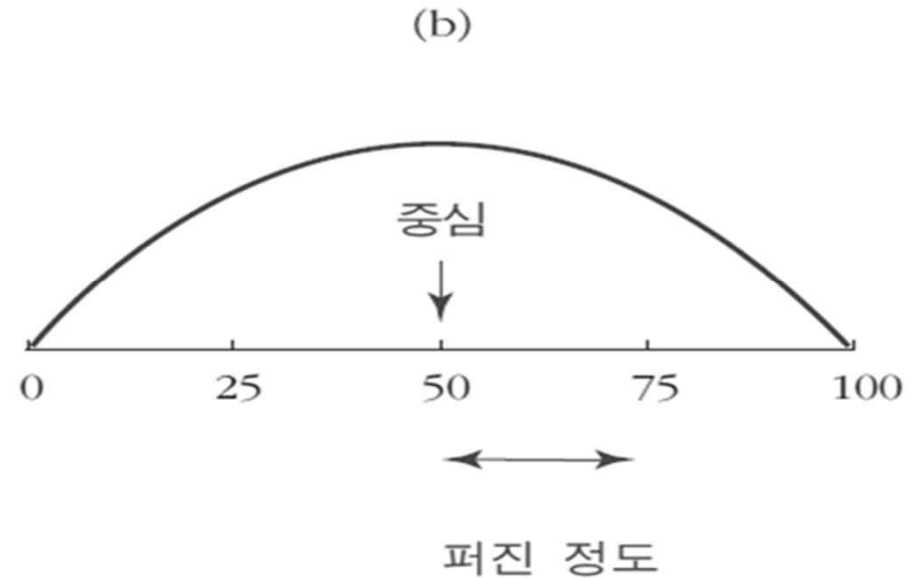
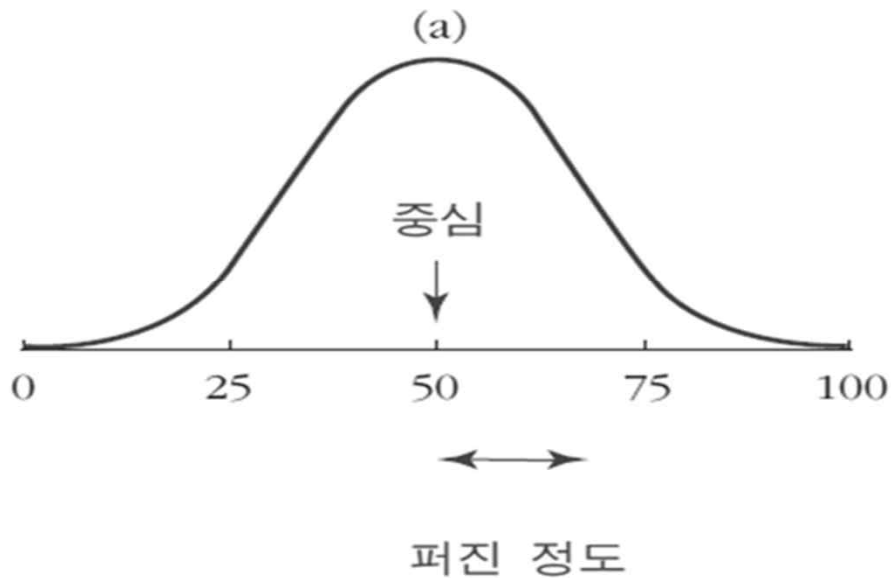


평균과 표준편차

히스토그램의 중심과 퍼진 정도

- 히스토그램에서 자료를 처리할 때 중심 (평균, 중앙값)과 중심 주위로 퍼진 정도 (표준편차, 사분위수 범위) 사용



집중 경향치

- 자료 분포의 중심으로 자료 전체를 대표할 수 있는 값
 - 산술평균: 일반적인 평균. 모든 자료의 합을 자료의 수로 나눈 값

모집단 :

표 본 :

: 개별 관측치, : 모집단 크기, : 표본크기

- 가중평균: 가중치를 둔 평균

: i 번째 관찰치

: i 번째 관찰치에 적용하는 가중치

대표값

- 평균의 특징
 - 모든 자료로부터 영향을 받는다
 - 수학적 연산이 가능하다.
 - 가중 평균 계산이 가능하다.
 - 분산의 계산과 모수의 추정 외에 가설 검정 등 통계 분석의 대표적인 값으로 사용된다.

예제 1

- 의료공학 프로젝트 1학년 수강생 최군은 6과목을 수강하였는데 2 과목 A, 2 과목 B, 1 과목 C, 1 과목 D를 획득하였다. A=4, B=3, C=2, D=1점이라고 하면 최군의 평점은 얼마인가?

중앙값

- 자료의 수를 크기 순서로 나열했을 때 중간에 해당하는 관찰값
- 중앙값을 산출하는 절차
 - 자료를 크기 순서로 나열
 - 자료의 수 n 이 홀수이면 $(n+1)/2$ 번째 자료 값이 중앙값
 - 자료의 수 n 이 짝수이면 $n/2$ 번째와 $n/2+1$ 자료 값을 평균한 값

예제 2

- 의료공학 프로젝트 수강생들의 나이는 22, 30, 23, 27, 24, 38, 21, 41, 34, 28이다. 이들 나이의 중앙값은 얼마인가?

최빈값 (Mode)

- 양적 또는 질적 자료 중에서 도수가 가장 많은 관찰치의 값을 의미한다.
- 도수가 모두 동일한 자료는 최빈값이 없으며, 동시에 두 개의 최빈값을 갖는 경우 쌍봉 (Bimodal), 세 개 이상의 최빈값을 갖는 경우 다봉 (Multimodal)이라 한다.

중앙값, 최빈값의 특성

- 자료내 극단적인 이상치가 있는 경우 극단적 관찰치에 덜 민감한 중앙값이 대표값으로 사용될 수 있음.
- 자료의 분포가 비대칭인 경우 평균의 보조 자료로 이용할 수 있음.
- 명목 자료와 서열 자료의 경우 평균과 중앙값을 계산할 수 없으므로 최빈값을 대표 자료로 사용
 - 남/녀, 학과, 상/중/하, 만족/보통/불만족 등

예제 3

- 다음 자료의 최빈값을 구하시오

– 0 0 2 3 3 4 4 5

– 1 3 5 7 9

– 22 27 30 30 30 30 34 38 60 60 60 60 65

산포도 (분산도)

- 자료 분포의 특성을 분석할 때 집중경향치 (평균, 중앙값, 최빈값)과 함께 산포도를 고려
 - 산포도 (분산도)는 자료가 흩어져 있는 정도를 측정
 - 두 분포에서 자료의 흩어짐을 비교하는데 이용
 - 측정변수: 범위, 중간범위, 평균절대편차, 분산, 표준편차, 변동계수 등

범위

- 범위: 자료에서 최대치와 최소치의 차이
- 중간범위: 최대치와 최소치의 합을 2로 나눈 값
- 평균절대편차: 편차 절대값을 모든 자료에 대해 구한 후 총 합을 자료의 총 수로 나눈 값
 - 편차: 자료에서 각 관찰 값이 자료의 평균으로부터 떨어진 정도

모집단 :

표 본 :

예제 4

- 다음 표본 자료에 대한 평균절대편차를 구하십시오
– 8 6 7 9 10

분산

- 모든 편차 (개별 자료와 평균의 차이)를 제공하고 이들을 합한 후 자료의 총 수로 나눈 값

$$\text{모집단: } \sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

$$\text{표본: } s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

분산의 특징

- 자료가 평균 주위로 얼마나 집중되어 있는가를 측정
- 분산 값이 작으면 자료의 변동이 심하지 않고 대체로 평균 가까이에 분포하고 있음을 의미
- 표본분산은 모분산을 구하고자 할 때 추정치로서 이용
- 표본분산 공식에서 $n-1$ 대신 n 을 사용하면 모분산을 과소추정. $(n-1)$ 을 사용하면 모분산의 불편추정치 (unbiased estimate)가 됨.
- 분산은 각 자료에 대한 편차 제곱을 하기때문에 원 자료의 단위보다 큰 단위로 표시됨.

자유도

- 전체 값 중 실질적으로 독립적인 값들의 개수
- 자유도 = 자료의 개수 - 1
- 자유도를 고려하지 않은 표준 편차는 실제 값을 과소 평가
- 극단적으로 자료의 개수가 하나인 경우 (3)
- 평균은 3이고 표준편차는 자유도를 고려하지 않을 경우 0, 자유도를 고려하면 계산 불능. 자료 하나로부터 퍼진 정도를 알 수 없기 때문에 퍼진 정도는 0이 아니라 불능이 정답

표준편차

- 분산의 양의 제곱근

$$\text{모집단: } \sigma = \sqrt{\sigma^2}$$

$$\text{표본: } s = \sqrt{s^2}$$

- 분산의 제곱근인 표준편차를 구하면 원래 자료의 단위로 환원되어 같은 단위로 측정된 평균이나 다른 통계량과 쉽게 비교할 수 있음.

분산, 표준편차 특성

- 자료가 흩어질수록 범위, 중간범위, 분산, 표준편차는 커짐
- 자료가 평균 주위로 집중하면 범위, 중간범위, 분산, 표준편차는 작아짐
- 자료가 모두 동일하면 범위, 중간범위, 분산, 표준편차 0
- 범위, 중간범위, 분산, 표준편차는 모두 양수 (+)임

예제 5

- 다음 자료의 분산과 표준편차를 구하십시오
 - 37 42 46 46 54 46 44
 - 자료가 모집단이라고 가정할 때
 - 자료가 표본이라고 가정할 때

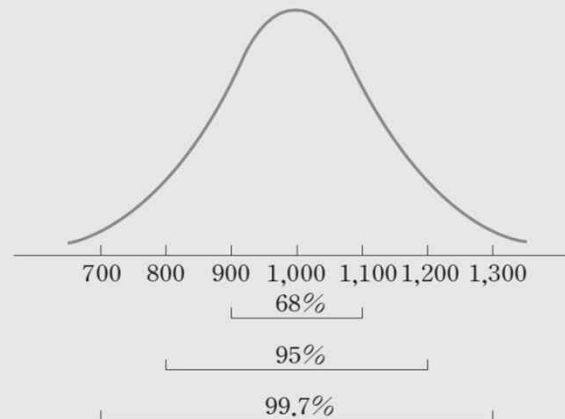
경험에 의한 규칙

- 자료의 분포가 평균을 중심으로 좌우대칭을 이루어 종 모양을 이루면 경험 법칙이 성립

경험법칙

- 모든 관찰치들의 약 68%는 평균±1표준편차 내에 포함된다.
- 모든 관찰치들의 약 95%는 평균±2표준편차 내에 포함된다.
- 모든 관찰치들의 약 99.7%는 평균±3표준편차 내에 포함된다.

그림 3-3 경험법칙의 예



체비셰프의 정리

- 자료의 분포가 정규분포가 아니거나 혹은 분포를 전혀 모르는 경우 체비셰프의 정리가 적용
- 체비셰프 정리
 - 표본 혹은 모집단의 어떤 자료에 있어 평균 $\pm k$ 표준편차 내에 존재하는 자료의 비율은 적어도 전체 자료의 $1-(1/k^2)$ 이다. 단 $k > 1$

예제 6

- 가톨릭관동대학교 1학년 학생 전체의 IQ는 평균이 120이고 표준편차가 15로 나타났다.
 - 체비셰프 정리를 기반으로 IQ가 90에서 150 사이에 있는 학생의 비율은?
 - 만일 IQ 분포가 종 모양의 좌우 대칭을 이룬다면 90과 150 사이의 비율은?

숙제

- 의료공학 프로젝트 수강생은 총 100명, 중간고사의 평균은 80점이고 표준편차는 5점이다.
 - 체비셰프 정리를 기반으로 70점과 90점 사이의 점수를 받은 학생은 최소 몇 명인가?
 - 만일 점수의 분포가 종 모양의 좌우 대칭을 이룬다면 70점과 90점 사이의 점수를 받은 학생은 몇 명인가?

변동계수

- 두 자료군의 상대적 변동을 측정하는데 이용되는 기법으로 상대적 표준편차라고도 함.
- 두 자료군 사이의 측정 단위가 서로 다르거나 평균에 있어 큰 차이가 있을 경우 표준편차로 비교하는데 무리가 있으므로 이 경우 변동 계수를 이용하여 비교함.

$$\begin{aligned} \text{모집단: } CV &= \frac{\text{표준편차}(\sigma)}{\text{평균}(\mu)} \times 100\% \\ \text{표본: } CV &= \frac{\text{표준편차}(s)}{\text{평균}(\bar{x})} \times 100\% \end{aligned}$$

숙제

- 최씨는 6주 동안 가격 변동을 나타낸 자산 1과 자산 2 중에서 하나를 골라 투자하려고 한다. 어떤 자산에 투자하는 것이 덜 위험할까?

	1주	2주	3주	4주	5주	6주
- 자산 1	55	68	63	72	65	69
- 자산 2	10	18	8	20	13	15

위치의 측정: 백분위수

- 자료를 크기 순으로 정리하여 백등분 하였을 때 각 등분 점에 위치하는 자료
- P 번째 백분위를 계산하는 절차
 - 자료를 작은 것부터 큰 순서로 정렬
 - 지수 i 를 다음과 같이 계산
 - 만일 i 가 정수이면 i 와 $(i+1)$ 의 위치에 있는 자료를 평균한 값이 P 번째 백분위수임. 만일 i 가 정수가 아니면 이를 올림한 것이 P 번째 백분위수임.

예제 7

- 다음은 의료공학 프로젝트 중간고사 시험 성적이다.

22 25 28 31 34 35 39 39 40 42
44 44 47 48 49 51 54 54 55 55
56 57 59 60 61 63 63 63 65 66
67 68 69 71 72 72 74 75 75 76
78 78 80 82 83 85 88 90 92 96

- 35번째 백분위 수를 구하십시오
- 60번째 백분위 수를 구하십시오
- 한 학생의 점수가 80점일 때 이는 몇 번째 백분위수에 해당하는가?

사분위수

- 백분위수 중 25번째 백분위수를 1사분위수 Q_1 , 50번째 백분위수를 2사분위수 혹은 중앙값 Q_2 , 75번째 백분위수를 3사분위수 Q_3 이라고 함.
- 사분위수의 범위: 1사분위수와 3사분위수의 차이

예제 8

- 다음은 의료공학 프로젝트 중간고사 시험 성적이다.

22 25 28 31 34 35 39 39 40 42
44 44 47 48 49 51 54 54 55 55
56 57 59 60 61 63 63 63 65 66
67 68 69 71 72 72 74 75 75 76
78 78 80 82 83 85 88 90 92 96

- 1사분위수?
- 2사분위수?
- 3사분위수?
- 사분위수 범위?