1장 정보의 시각화

* 일상에서의 통계

예제①: 자신과 남의 통신비 지출내역을 비교하며
경우에 따라 여름방학 알바를 하지 않아도
된다는 결론을 유도해 볼 수 있음.

예제②: 소득분배 관련 통계청의 가계동향조사
결과를 어떻게 받아들이는가에 따라
사회와 정치에 대한 이해가 달라짐.

* 통계란?: 주어진 데이터에서 정보를
추출하는 것.

- 데이터: 수집된 사실 또는 숫자
- 정보: 어떤 의미가 부여된 데이터

* 통계 학습 목적

① 통계 산출 방법 학습
② 통계 활용법 학습

* 통계를 배워야 하는 이유

① 세상에서 벌어지는 일들을 이해하고 설명하기 위해
② 사실 관계를 정확하게 판단하기 위해

44쪽

* 정보전달 방법 선택

• 동일한 데이터에서 다른 정보, 강조에 따라 상반되는 정보를 이끌어 낼 수 있다.

• 예제: 어떤 회사의 월별 실적

(단위: 억원)

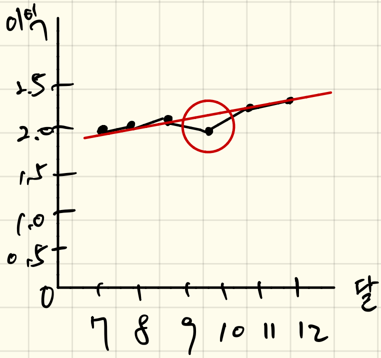
월	7	8	9	10	11	12
이익	2.0	2.1	2.2	2.1	2.3	2.4

어떤 정보를 확인할 수 있나? 아마도 아래와 같이 답할 가능성이 높다.

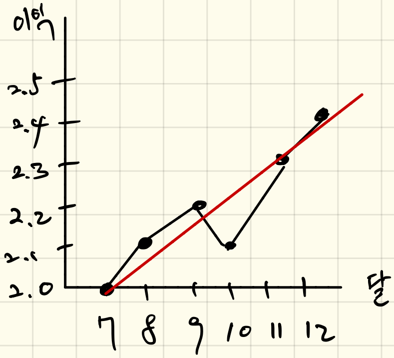
"월별 이익이 완만하지만 꾸준히 증가한다"

정말 이것 뿐일까? 아래 두 그래프를 비교해보자.

①



②



회사의 CEO라면 경우에 따라 두 그래프 중 하나를 선택해서 사용할 것이다.

경우 ①: 회사가 꾸준히 성장하고 있음을 강조하고 싶을 때 그래프 ① 사용

경우 ②: 회사의 실적을 과장하여 보여주고 자 할 때 그래프 ② 사용

* 정보를 전달하는 네 가지 방법

① 파이차트 (원그래프)
② 막대그래프
③ 히스토그램
④ (꺾은)선 그래프

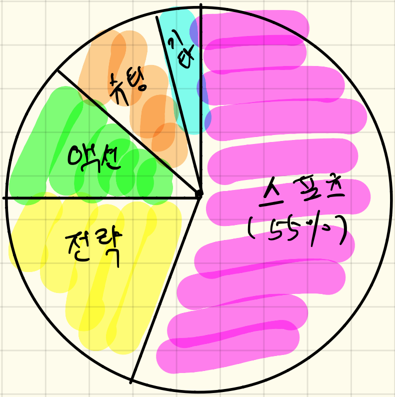
(4쪽) ① 파이차트

어떤 게임회사에서 판매한 게임의 장르별 판매량

도수

범주

장르	판매량
스포츠	27,500
전략	11,500
액션	6,000
슈팅	3,500
기타	1,500



장점: 각 제품이 차지하는 비율을 한눈에 볼 수 있게 도와줌.

단점: 항목별 숫자가 별 차이를 갖지 않을 때 의미 없음.

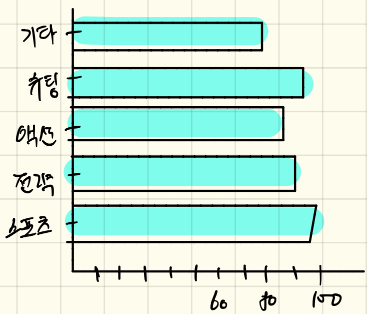
* 알아두어야 할 개념 두 가지

① 범주 : 동일한 성질을 갖는 부류 또는 범위

② 도수 : 특정 범주에 속하는 개체의 개수

② 막대그래프

장르	만족도(%)
스포츠	99
전략	90
액션	85
슈팅	95
기타	80



(수평) 막대그래프

* 막대그래프 장점

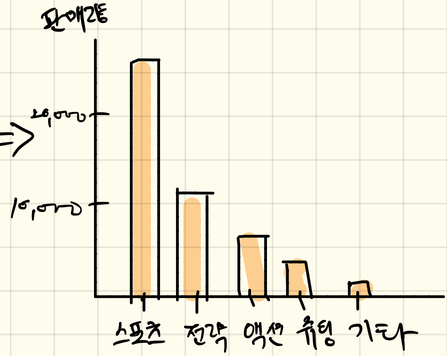
① 범주별 미세한 차이를 명확히 보여줌.

② 복수의 데이터를 하나의 막대그래프로 보여줄 수 있음.

① 의 예 : 파이차트를 막대그래프로 나타낼 수 있음.

장르	판매량
스포츠	27,500
전략	11,500
액션	6,000
슈팅	3,500
기타	1,500

⇒



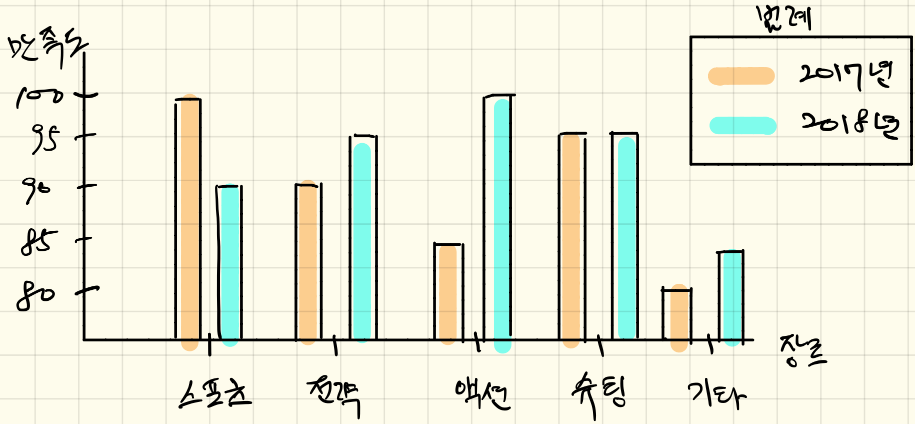
② 의 예

장르	만족도(%)
스포츠	99
전략	90
액션	85
슈팅	95
기타	80

(2017년)

장르	만족도(%)
스포츠	90
전략	95
액션	99
슈팅	95
기타	85

(2018)



58쪽 ③ 히스토그램

* 데이터 종류

① 범주형 데이터 (질적 데이터)
범주의 성질이나 특성에 따라 나뉘어진 데이터

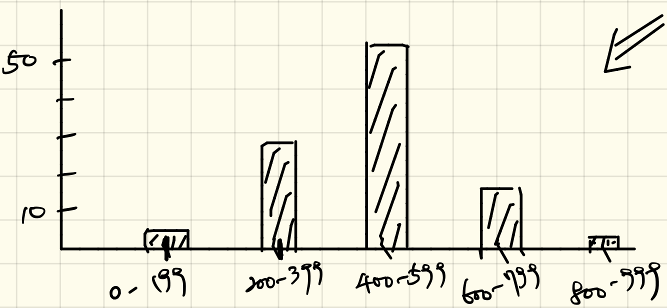
② 수치적 데이터 (양적 데이터)
무게, 길이, 점수, 시간 등 숫자의 범위에 따라 나뉘어진 데이터

61쪽

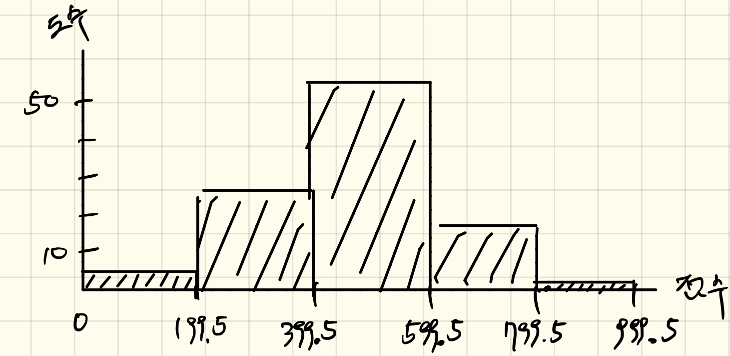
예) 시험

범주

점수	도수(명)
0-199	5
200-399	29
400-599	56
600-799	17
800-999	3



문제:
구간별의 연속성을 반영 못함



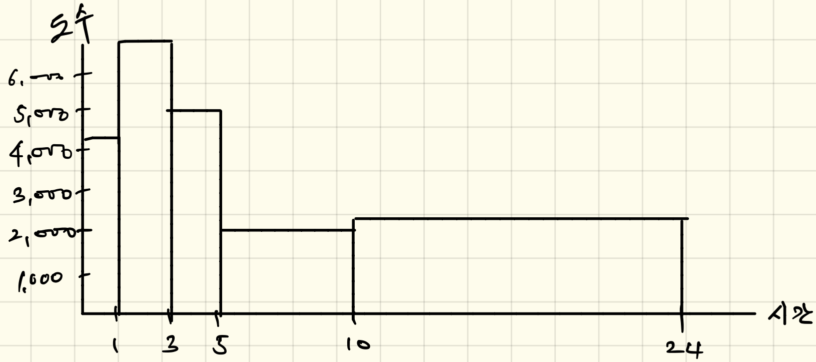
* 도수밀도

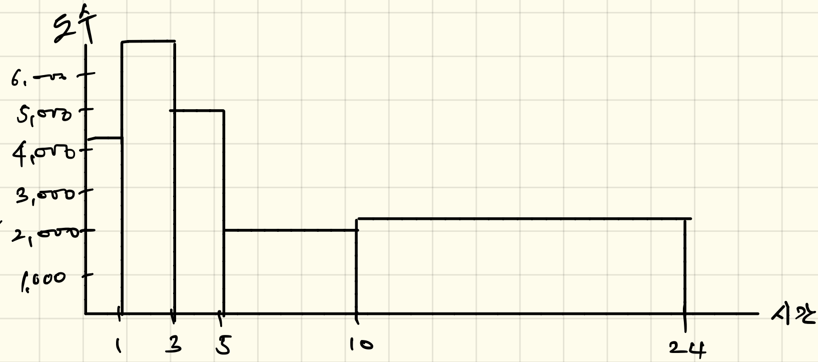
64쪽

예) 하루동안 게임하는 시간

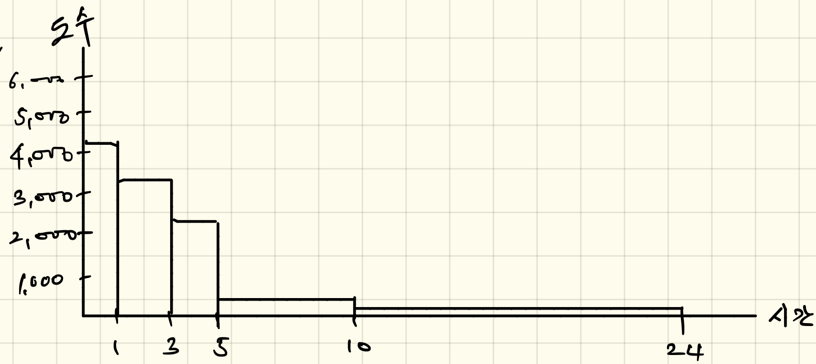
범주

시간	도수
0 - 1	4,300
1 - 3	6,900
3 - 5	4,900
5 - 10	2,000
10 - 24	2,100





$$도수밀도 = \frac{도수}{구간의\ 폭} = \frac{도수}{막대의\ 폭} = 막대의\ 높이$$



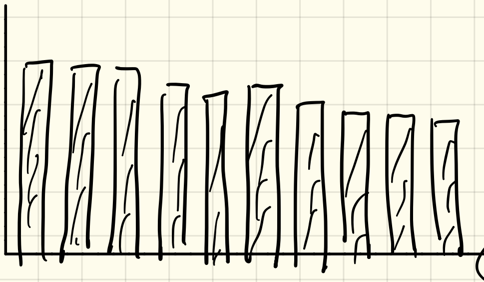
④ 선 그래프

• 활용예제 ①

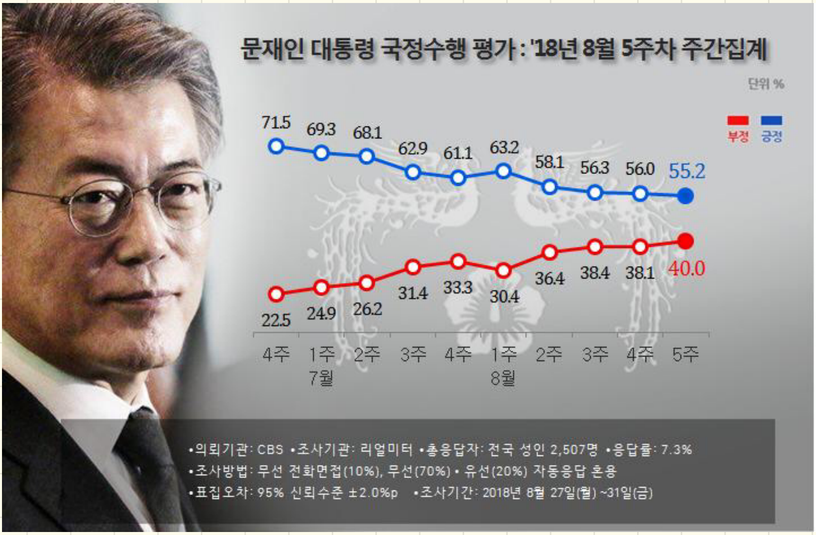
양적 데이터의 특수한 경우

↓

시간, 연도 등 일정한 흐름을 보여주는 데이터



시계열 데이터



P.74

② 활용예제 ②

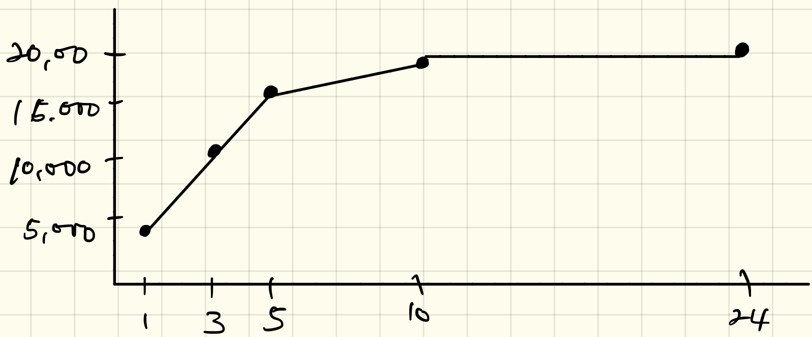
누적도수 그래프를 그릴 때 유용함.

?

시간	도수	상한시간	누적도수
0 − 1	4,300	1	4,300
1 − 3	6,900	3	11,200
3 − 5	4,900	5	16,100
5 − 10	2,000	10	18,100
10 − 24	2,100	24	20,200

질문 : 최대 3시간 게임하는 사람의 수?

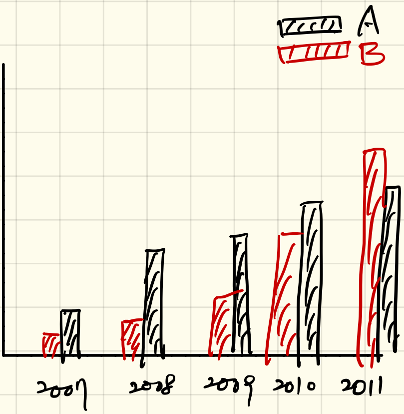
답 : 4,300 + 6,900 = 11,200 명

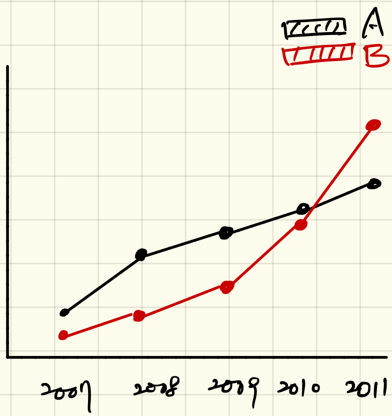


* 막대그래프 사용 불가!

• 활용예제 ③

P.79





2장. 평균구하기

1. 평균 (Average) 의 종류

① 평균값 (Mean)

② 중앙값 (중위값, Median)

③ 최빈값 (Mode)

2. 평균값 (Mean)

P-88 예제 : 헬스 교실 참가자들의 나이

19, 20, 20, 20, 21

질문 : 참가자들 나이의 평균값?

답 : $\frac{19+20+20+20+21}{5} = 20$

P. 90 일반화 : 평균값(μ) = $\frac{\text{모든 데이터의 합}}{\text{도수의 합}}$

$$= \frac{\Sigma X}{n}$$

* 기호설명

합

개별 데이터를 가리키는 기호

$$\mu = \frac{\sum X}{n}$$

어떤 주어진 데이터의 평균값을 나타내는 기호

주어진 데이터의 개수를 나타내는 기호

$\Rightarrow$ $\sum X$ = 모든 데이터들의 합

예) 유한개의 데이터가 주어졌다고 가정.

$x_1, \cdots, x_n$: n 개의 데이터

$$\Rightarrow \sum X = \sum_{i=1}^{n} x_i$$

$$= x_1 + \cdots + x_n$$

* 이상치 (Outlier)

이상한 값

p.26 예제 : 중푸교실 참가자들 나이

나이	19	20	21	145	147
도수	3	6	3	1	1

$$\mu = \frac{19\cdot 3 + 20\cdot 6 + 21\cdot 3 + 145 + 147}{14}$$

$$= 33$$

⇒ 평균값이 주어진 데이터를 제대로 대변하지 못함.

- 145, 147세를 제외한 참가자들 나이의 평균값 : 20세

- 이상치 : 평균값을 편향 시키는 값

  예 : 145, 147세

p.101

3. 중앙값 (Median) (중위값)

정의: 데이터를 크기순으로 나열했을 때 한가운데에 위치한 값.

예제:

데이터 개수 짝수일 때: 19, 19, 19, 20, 20, 20, 20, 20, 20, 21, 21, 21, 145, 147

평균값: 20

홀수: 19, 19, 20, 20, 20, 21, 21, 100, 102

중앙값

* 중앙값 구하기 3단계

전제: N개의 데이터가 크기순으로 나열되어 있음.

① N 짝수인 경우:

$$\text{중앙값} = \frac{N}{2}\text{번째와 } \left(\frac{N}{2}+1\right)\text{번째}$$

값들의 평균값

② N 홀수인 경우:

$$\text{중앙값} = \frac{N+1}{2}\text{번째 값.}$$

* 편향된 데이터의 평균값과 중앙값 비교

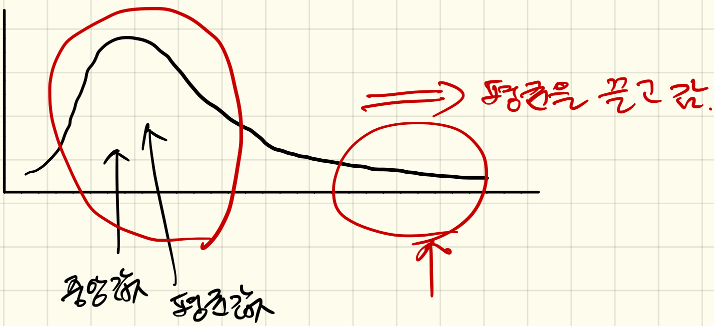
① 오른쪽으로 편향된 데이터

예제:

값	1	2	3	4	5	6	7	8
도수	4	6	4	4	3	2	1	1

평균값(μ) = 3.44

중앙값 = 3

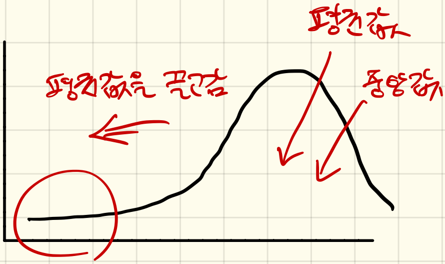


② 왼쪽으로 편향된 데이터

값	1	4	6	8	9	10	11	12
도수	1	1	2	3	4	4	5	5

평균값 = 9.28

중앙값 = 10



* 최빈값 (mode)

p.108~110

예제: 부모와 아이가 함께하는 수영교실

나이	1	2	3	31	32	33
도수	3	4	2	2	4	3

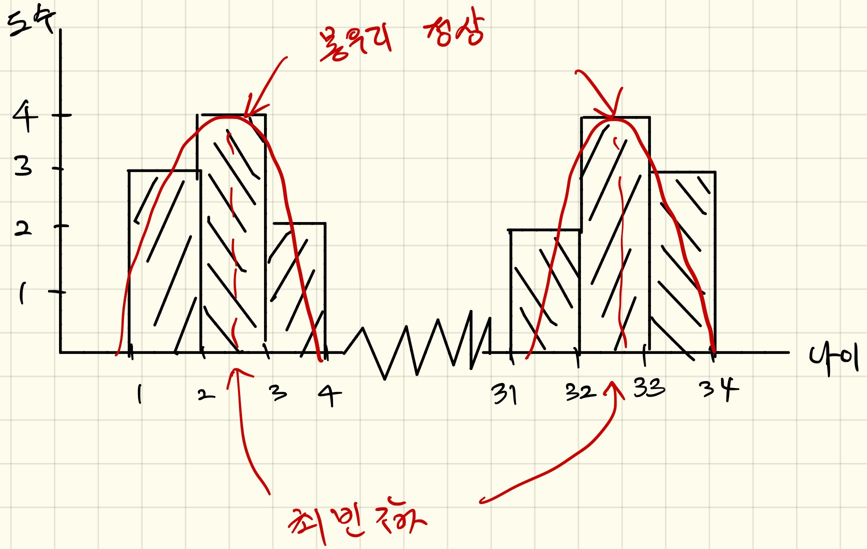
μ = 17(세)

중앙값 = 17(세)

} 주어진 데이터를 전혀 대표하지 못함

=> 아이들의 모습을 대표하는 값 : 2세

부모들의 모습을 대표하는 값 : 32세



p.119

연습문제

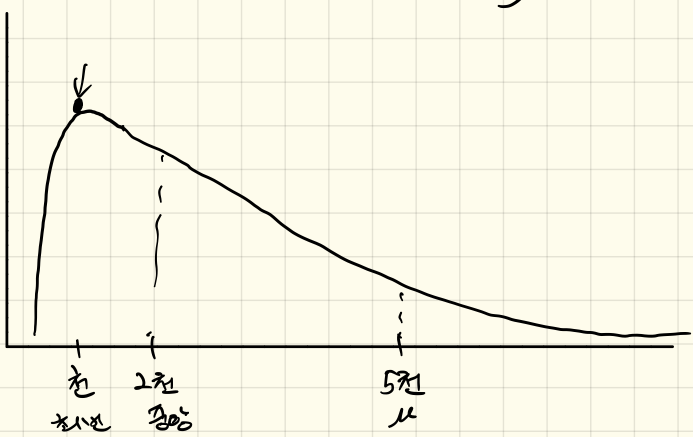
회사의 CEO가 모든 직원들의 연봉을 올려주고자 함.

방식 ① : 모두의 연봉을 동일하게 2백만원씩 인상하기

방식 ② : 각자의 연봉을 10% 인상하기

상황 설정 : 노조위원장 입장에서 어느 방식을 선택해야 할지 판단해야 함.

평균값 = 5천만원
중앙값 = 2천만원
최빈값 = 1천만원



3장. 변이와 분포 측정하기

p.125

예제 : 세 명의 농구 선수 기록 비교

선수 1

게임당 점수	7	9	10	11	13
도수	1	2	4	2	1

선수 2

게임당 점수	7	8	9	10	11	12	13
도수	1	1	2	2	2	1	1

선수 3

게임당 점수	3	6	7	10	11	13	30
도수	2	1	2	3	1	1	1

세 선수 기록 모두 평균값, 중앙값, 최빈값이 10으로 동일.

⇒ 평균으로는 세 선수의 특성을 구분 못함.

* 세 선수 기록의 특성을 구별하는 새로운 방법이 필요.

⇒ 일관성을 기준으로 사용 가능

⇓

일관성을 평가하는 방법 필요!

p.126

* 일관성 평가 방법 1 : 범위

예) 선수 2의 기록

7 8 9 9 10 10 11 11 12 13

범위

하한 상한

$$\text{범위} = \text{상한} - \text{하한}$$
$$= 13 - 7$$
$$= 6$$

예) 선수 3의 기록

$\text{범위} = 30 - 3 = 27$

⇒ 선수2의 일관성이 선수3의 일관성보다 높다라고 할 수 있음.

* 범위 사용의 한계

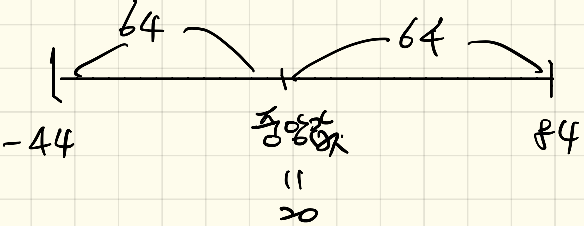
- 이상치가 있는 경우 문제 발생.

예) 2장에서 다룬 공포교실 참가자 나이 데이터

나이	19	20	21	145	147
도수	3	6	3	1	1

범위 = 147 − 19 = $\underline{\underline{128}}$

⇓



⇒ 이상치가 범위를 왜곡시킴.

평균값 뿐만 아니라 데이터의

⇒ 사분 범위

p.132~133

$20 \cdot \frac{1}{4}$ Q₂(중앙값) $20 \cdot \frac{3}{4}$

예) 1 1 1 2 2 | 2 2 3 3 3 | 3 3 4 4 4 | 4 5 5 5 10

제외 Q_1 사분 범위 Q_3 제외

사분 범위 = 4 − 2 = 2

예) 콩쥐교실 데이터

$\lceil 14 \cdot \frac{1}{4} \rceil$ 올림 $\lceil 14 \cdot \frac{3}{4} \rceil$ Q_2

19 19 19 20 20 20 20 20 20 21 21 21 145 147

제외 Q_1 사분 범위 Q_3 제외

사분 범위 = 21 − 20 = 1

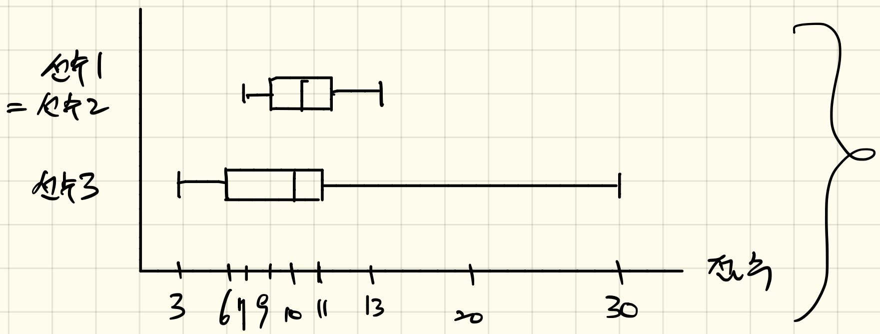
⇒ 사분 범위 구하기

① 사분위수 활용 : Q_1, Q_2, Q_3

P. 140

상자수염 다이어그램

예) 농구선수 3의 점수 데이터 범위를 한 눈에 알아볼 수 있도록 시각화하는 방법



P. 143

사분범위의 한계

- 사분범위 안에 위치한 데이터들이 얼마나 자주 발생하는지 또는 얼마나 자주 중앙값에 가까운 점수를 내는지 알려주지 않음.

⇒ 일관성을 측정하는 방식으로 사분범위 이외의 다른 기준이 요구됨.

⇒ 변이 (변동) 개념을 활용할 것임.

p.144

* 변이를 측정하기 위해 두 가지 방법 활용

① 분산

② 표준편차

* 변이(변동)란?

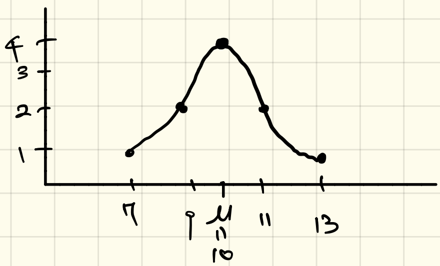
- 단순히 데이터들의 분포 양상이 아니라, 데이터의 산점성까지 포함하는 개념.

↳ 데이터 각각이 평균값(μ)으로부터 떨어져 있는 정도

⇙

분산과 표준편차를 활용하여 계산!

예) 농구 선수1과 선수2의 기록 그래프 비교



선수1

선수

⇒ 선수1의 기록이 선수 2의 기록보다 평균에 집중되어 있음.

p.145

• 평균거리 : 각 데이터와 평균값의 차이의 평균값

예) 농구 선수 3의 기록의 평균거리

게임당 점수	3	6	7	10	11	13	30
도수	2	1	2	3	1	1	1
μ와의 차이	−7	−4	−3	0	1	3	20

$$평균거리 = \frac{(3-10)*2 + (6-10) + (7-10)*2 + (10-10)*3 + (11-10) + (13-10) + (30-10)}{11}$$

$= 0$ (평균거리는 항상 0임)

• 평균 편차 :

각 데이터가 평균으로 떨어진 정도의 평균값

$$평균편차 = \frac{|3-10|*2 + |6-10| + |7-10|*2 + |10-10|*3 + |11-10| + |13-10| + |30-10|}{11}$$

$$= \frac{14+4+6+1+3+20}{11}$$

$$= \frac{48}{11} = 4.3xxx$$

하지만 평균편차는 사용하지 않음.
대신에 분산과 표준편차를 범위를 측정하는 데에 사용함.

$$\text{분산}\ (\sigma^2) = \frac{(3-10)^2 * 2 + (6-10)^2 + (7-10)^2 * 2 + (10-10)^2 * 3 + (11-10)^2 + (13-10)^2 + (30-10)^2}{11}$$

$$= \frac{98 + 16 + 18 + 1 + 9 + 400}{11}$$

$$= \frac{542}{11}$$

$$= 49.1xx$$

⇒ 분산은 너무 큼

대신에 표준편차를 사용.

$$\text{표준편차}(\sigma) = \sqrt{\text{분산}}$$

$$= \sqrt{49.1xx}$$

$$= 7.xx$$

- 선수1의 기록의 표준편차와 분산

게임당 점수	7	9	10	11	13
도수	1	2	4	2	1

$$\sigma^2 = \frac{9+2+2+9}{10} = \frac{22}{10} = 2.2$$

$$\sigma = \sqrt{2.2} = \underline{1.48xx}$$

- 선수2의 기록의 표준편차와 분산

게임당 점수	7	8	9	10	11	12	13
도수	1	1	2	2	2	1	1

$$\sigma^2 = \frac{9+4+2+2+4+9}{10}$$

$$= \frac{30}{10} = 3$$

$$\sigma = \sqrt{3} = \underline{1.71xx}$$

p. 155

연습문제

회사의 CEO가 모든 직원들의 연봉을 올려주고자 함.

방식 ① : 모두의 연봉을 동일하게 2백만원씩 인상하기

방식 ② : 각자의 연봉을 10% 인상하기

기존 연봉의 표준편차 = σ_{old} ,

⇒ 질문 : 연봉을 올릴 때 표준편차가 어떻게 변하는가?

방식 ① : $\sigma_{new}^2 = \dfrac{\sum((x+\cancel{200}) - (\mu_{old}+\cancel{200}))^2}{n}$

$= \dfrac{\sum(x-\mu_{old})^2}{n}$

$= \sigma_{old}^2$

소득격차가 10% 만큼 벌어짐.

방식 ② : $\sigma_{new}^2 = \dfrac{\sum(x*1.1 - \mu_{old}*1.1)^2}{n}$

$= (1.1)^2 \dfrac{\sum(x-\mu_{old})^2}{n}$

$= (1.1)^2 \cdot \sigma_{old}^2$

$$분산(\sigma^2) = \frac{\sum (x-\mu)^2}{n}$$

$$표준편차(\sigma) = \sqrt{\frac{\sum (x-\mu)^2}{n}}$$

$$분산(\sigma^2) = \frac{\sum (x-\mu)^2}{n}$$

$$= \frac{\sum (x^2 - 2\mu x + \mu^2)}{n}$$

$$= \cdots$$

$$= \frac{\sum x^2}{n} - \mu^2$$

보다 간단한 계산.

분산(표준편차)의 특성

- 개별 데이터에 대한 정보를 반영
- 계산 쉬움.
- 분산이 (표준편차) 클수록 데이터들이 평균값으로부터 멀리 흩어져 있음을 의미함.

p.159

- 互为逆元

5장. 이산 확률 분포

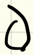
주제 : 어떤 사건의 발생할 확률에 대한 정보를
활용하여 각각의 사건이 미칠 영향을 알아본다.

이산 데이터 : - 예: ① 1, 2, 3, 4, --- 처럼 따로 따로
(discrete) 구분할 수 있는 데이터
② 연도, 월,

≠
연속 데이터
↳ 예: 키, 몸무게, ---

예제 : 슬롯머신

창 : 3개

그림 : $, , , 기타.

달러, 체리, 레몬

• 잭팟 : $\boxed{(\$, \$, \$) \text{ 가 나오는 사건}}$

↳ 정해진 순서대로 그림이 나오는 사건

p.238

예제

게임 규칙

- 게임당 1000원 지불
- 아래 4 사건이 발생하는 경우 상금 받음.

① ($, $, $) : 20,000원

② $ * 2회, Ø * 1회 : 15,000원

③ (Ø, Ø, Ø) : 10,000원

④ (0, 0, 0) : 5,000원

- 나머지 경우 : 꽝
- 각각의 창에서 특정 그림에서 멈출 확률
  - $: 0.1
  - Ø : 0.2
  - 0 : 0.2
  - 기타 : 0.5
- 전제 : 각각의 창은 서로 독립.

$	Ø	0	기타
0.1	0.2	0.2	0.5

p.239

- ①~④의 사건이 발생할 확률

- $P(①) = 0.1 * 0.1 * 0.1 = 0.001$

$P(②) = P(\text{Ø}, \$, \$)$

$+ P(\$, \text{Ø}, \$)$

$+ P(\$, \$, \text{Ø})$

$= (0.2 * 0.1 * 0.1) + (0.1 * 0.2 * 0.1) + (0.1 * 0.1 * 0.2)$

$= 0.002 * 3 = 0.006$

$P(③) = (0.2)^3 = 0.008$

$P(④) = (0.2)^3 = 0.008$

$P(\text{꽝}) = 1 - (0.001 + 0.006 + 0.008 + 0.008)$

$= 0.977$

⇒ 확률분포표

예) 슬롯게임의 경우

예) 슬롯게임의 경우

사건	①	②	③	④	꽝
확률	0.001	0.006	0.008	0.008	0.977

* 기대치

- 기대치 : 기대되는 값
- 예제 : 슬롯머신의 경우, 게임을 1회 실행할 때 볼 수 있는 액수의 평균값

슬롯머신

$P(X=19{,}000)=0.001$

$P(X=14000)=0.006$

$P(X=9000)=0.008$

$P(X=4000)=0.008$

$P(X=-1000)=0.977$

⇒

$$\begin{aligned}
&19{,}000 * 0.001 \\
+\ &14{,}000 * 0.006 \\
+\ &9{,}000 * 0.008 \\
+\ &4{,}000 * 0.008 \\
+\ &(-1{,}000) * 0.977 \\
\hline
=\ &-770
\end{aligned}$$

p.242

- 확률변수
  - 특정한 확률과 연관되어 있는 값들을 대상으로 하는 변수

    예) 슬롯머신에서 확률변수는 따는 액수 중 하나를 가리킴.

    즉, 19,000, 14,000, 9,000, 4,000, -1,000 중에 하나를 가리키는 변수.

  - 일반적으로 $X, Y, X_1, X_2, Y_1, Y_2, \dots$ 을 확률변수로 사용함.

- 확률변수 X가 x를 가질 확률 : $P(X=x)$

특정값

이전 : $P(A) = ?$

	①	②	③	④	꽝
x	19,000	14,000	9,000	4,000	-1,000
$P(X=x)$	0.001	0.006	0.008	0.008	0.977

⇒ 슬롯머신 게임을 1회 시행할 때 벌 수 있는 액수에 대한 확률분포표.

* 기대치 (일반화)

$$E(X) = \sum x \cdot P(X=x)$$

* 주의: 평균값 계산과 비슷.

$$\mu = \frac{\sum x}{n} = \sum \frac{x}{n} = \sum x \cdot \frac{1}{n}$$

→ 확률 변수 X에 대한 기대치.

* 확률변수 X에 대한 표준편차

$$Var(X) = \sum (x - E(X))^2 \cdot P(X=x)$$
$$= \sum (x-\mu)^2 \cdot P(X=x)$$

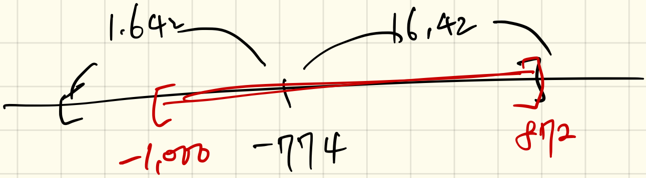
$$\sigma^2 = \frac{\sum (x-\mu)^2}{n} = \sum (x-\mu)^2 \cdot \frac{1}{n}$$

* 표준편차 $(\sigma) = \sqrt{Var(X)}$

예) 슬롯머신의 경우.

$$
\begin{aligned}
Var(X) = \ & (19000+774)^2 * 0.001 \\
& + (14000+774)^2 * 0.006 \\
& + (9000+774)^2 * 0.008 \\
& + (4000+774)^2 * 0.008 \\
& + (-1000+774)^2 * 0.977 \\
= \ & 2,697,100 \quad (2.6971 \text{ 달러})
\end{aligned}
$$

$$\sigma = 1,642$$



확률 분포와 선형변환

p.252
예제

게임 규칙

- 게임당 ~~1000~~ 2000원 지불
- 아래 4 사건이 발생하는 경우 상금 받음.

① ($, $, $) : ~~20,000원~~ 100,000원

② $ * 2회, Ø * 1회 : ~~15,000원~~ 75,000원

③ (Ø, Ø, Ø) : ~~10,000원~~ 50,000원

④ (O, O, O) : ~~5,000원~~ 25,000원

- 나머지 경우 : 꽝
- 각각의 칸에서 특정 그림에서 뽑을 확률
  - $: 0.1
  - Ø : 0.2
  - O : 0.2
  - 기타 : 0.5

	①	②	③	④	꽝
y	98,000	73,000	48,000	23,000	-2,000
$P(Y=y)$	0.001	0.006	0.008	0.008	0.977

$E(Y) = -850$

$Var(Y) = 67{,}427{,}500$

$\sigma = 8211.4$

-850

이렇게 매번 (수동으로)
계산해야 하나?

-2000 $\qquad$ -850 $\qquad$ 7361.4

$\Rightarrow$ 십중팔구 2000원 잃음.

그럴 필요 없음.

X와 Y 사이의 관계를 이용하면
$E(X)$와 $Var(X)$로 부터
$E(Y)$와 $Var(Y)$를 유도할 수 있음.

$$Y = X$$

$\underline{X}$는 상금액수 아님.

a가 어떤 사건 발생할 때의 상금.

$$\begin{cases} X = a - 1{,}000 \Rightarrow a = X + 1{,}000 \\ \underline{\underline{Y}} = 5a - 2{,}000 \end{cases}$$

$$= 5(X + 1{,}000) - 2{,}000$$

$$= 5 * X + 3{,}000$$

$$\Rightarrow \boxed{Y = 5 \cdot \underline{\underline{X}} + 3{,}000}$$

$$\Rightarrow E(Y) = E(5 \cdot X + 3000)$$

$$= 5 \cdot E(X) + 3{,}000$$

$$Var(Y) = Var(5X + 3000)$$

$$= 5^2 \cdot Var(X) +$$

$$\sigma_y = 5 \cdot \sigma_x$$

두 확률변수 X, Y 사이에 선형관계가
다음과 같이 있다고 가정:

$$Y = aX + b$$

$$\Rightarrow \begin{cases} E(Y) = a \cdot E(X) + b \\ Var(Y) = a^2 \cdot Var(X) \\ \sigma_Y = a \cdot \sigma_X \end{cases}$$

p.262

* 독립관측

x	-1,000	5,000
$P(X=x)$	0.9	0.1

$E(X)$
$Var(X)$

z	-2,000	4,000	10,000
$P(Z=z)$	0.81	0.18	0.01

$\|$

0.9 * 0.1 * 2

$\Rightarrow$ $E(Z) = ?$

$Var(Z) = ?$

Z 와 X 사이의 관계!

$$\boxed{Z = X + X}$$

$$\neq \boxed{2X}$$

$$\Rightarrow E(Z) = E(X+X)$$
$$= E(X) + E(X)$$
$$= \boxed{2 \cdot E(X)}$$

$$Var(Z) = Var(X+X)$$
$$= Var(X) + Var(X)$$
$$= 2 \cdot Var(X)$$
$$\neq Var(2X) = 2^2 \cdot Var(X)$$

X

$$Z = X_1 + X_2$$

X_1, X_2 : X의 복제

$$\begin{cases} E(X_1) = E(X_2) = E(X) \\ Var(X_1) = Var(X_2) = Var(X) \end{cases}$$

$$\begin{aligned} E(Z) &= E(X_1 + X_2) \\ &= E(X_1) + E(X_2) \\ &= 2 \cdot E(X) \end{aligned}$$

$$\begin{aligned} Var(Z) &= Var(X_1 + X_2) \\ &= Var(X_1) + Var(X_2) \\ &= 2 \cdot Var(X) \end{aligned}$$

일반화 $\Rightarrow$ $X_1, \cdots, X_n$

$$Z = X_1 + \cdots + X_n$$

$$E(Z) = E(X_1) + \cdots + E(X_n)$$

$$Var(Z) = \underline{Var(X_1) + \cdots + Var(X_n)}$$

* 확률변수의 뺄셈.

$Z = X - Y$: 두 확률변수 X와 Y의 차이를 가리킴.

p.279

예) A와 B 두 개의 레스토랑 있음.

메뉴가격 :

A식당

x	20,000	30,000	40,000	45,000
$P(X=x)$	0.3	0.4	0.2	0.1

B식당

y	10,000	15,000	18,000
$P(Y=y)$	0.2	0.6	0.2

$$E(Z) = E(X) - E(Y)$$

$$Var(Z) = Var(X) + Var(Y)$$

8장 정규분포 이야기

· 정규분포란?

- 연속(적인) 데이터를 다룰 때
  일반적으로 기대할 수 있는 값들의 분포.

보통 엄밀한 측정을 요하는 데이터.

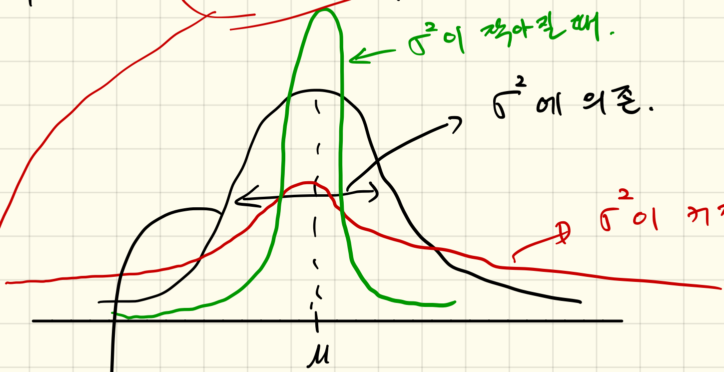
정규분포를 특징짓는 표현.

· 정규분포의 확률밀도함수 곡선

σ²이 작아질 때.

σ²에 의존.

σ²이 커질때



$$G(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

(가우스 함수)

$X \sim N(\mu, \sigma^2)$

정규분포

예) 20 ~ 24세 한국 남성들의 키를 가리키는 변수를 X라 할때,

$$X \sim N(173.5,\ 5.17^2)$$

$$P(X > 180) = ?$$

예) 중간고사 점수.

$$Y \sim N(61.6,\ 16.3^2) \quad \text{(가정)}$$

질문 : 내점수 = 75 점.

$$P(Y > 75) = ?$$

- 정규분포 확률밀도함수의 특징.

① σ^2 의 커지면 곡선이 낮고 평평해짐.
② σ^2 〃 작아 〃 〃 좁고 뾰족해짐.

③ 곡선은 x축과 절대로 만나지 않음.
하지만 평균값에서 멀어질 수록 점점 x축에 가까워 짐.

④ 곡선과 x축 사이의 공간의 면적은 항상 1 이다.

- 정규분포가 주어졌을 때, 즉, $X \sim N(\mu, \sigma^2)$,

$$\left.\begin{array}{l} P(X > a) = ? \\ P(X < a) = ? \\ P(a < X < b) = ? \end{array}\right\} \text{정규확률}$$

- 정규확률을 쉽게 구하는 3 단계

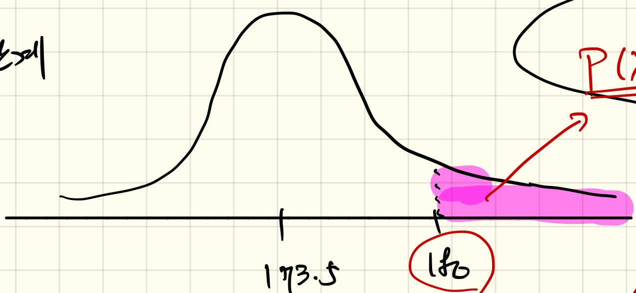
① 분포 결정 : μ 와 σ^2 확인
② 표준화 하기 : 표준점수 구하기
③ 확률 찾기(검색) : 확률 테이블 활용

예제) 20~24세 한국 남학생들의 키

$$X \sim N(173.5,\ 5.17^2)$$

$$P(X > 180) = ?$$

① 단계



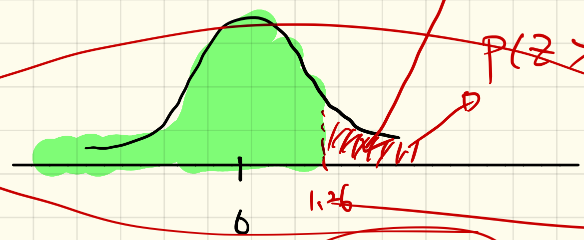
$P(X > 180)$

② 단계

$$z = \frac{x - \mu}{\sigma}$$

$$\frac{180 - 173.5}{5.17} = 1.26$$

$P(Z > 1.26) = ?$



$$Z \sim N(0,\ 1^2)$$

↓

표준정규분포.

$P(Z<a)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511966	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555670	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621720	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802337	0.805105	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823814	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.879000	0.881000	0.882977
1.2	0.884930	0.886861	0.888768	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903200	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935745	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959070	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965620	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999534	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999651
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999822	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

$$P(X>180) = P(Z>1.26)$$
$$= 1 - \underline{0.896165}$$
$$\simeq 0.104$$

$$P(180<X<185) = ?$$
$$= P(\underline{1.26}<Z<2.22) =$$

$$\frac{185-173.5}{5.17} = 2.22$$

$$P(1.26 < Z < 2.22)$$

$$= P(Z < 2.22) - P(Z < 1.26)$$

$$= 0.986791 - 0.896165$$

$$\simeq 0.091$$

9장 정규분포 알아보기 Ⅱ

- 주요 내용

정규분포 연산	정규분포 활용
- 합과 차이 - 선형변환 - 독립관측	- 이항분포

- 정규분포들의 합

예제 : 신부의 체중 $Y \sim N(150, 400)$
신랑의 체중 $X \sim N(190, 500)$

⇒ 신랑 + 신부 체중

$$X+Y \sim N(150+190, 400+500)$$

$$\sim N(340, 900)$$

일반화

$$X \sim N(\mu_x, \sigma_x^2) \qquad Y \sim N(\mu_y, \sigma_y^2)$$

$$\Rightarrow \quad X+Y \sim N(\mu_x+\mu_y, \sigma_x^2+\sigma_y^2)$$

<u>문제</u> 신랑과 신부의 체중의 합이 380 파운드 이하일 확률?

$$P(X+Y < 380) = P\left(Z < \frac{380-340}{30}\right)$$
$$= P(Z < 1.33)$$
$$= 0.908$$

· <u>정규분포의 차이</u>

인치

예제) 남자의 키 $X \sim N(71, 20.25)$
여자의 키 $Y \sim N(64, 16)$

⇒ 남·여 키의 차이의 분포

$$X - Y \sim N(7, \underline{36.25})$$

일반화 :

$$X \sim N(\mu_x, \sigma_x^2) \quad Y \sim N(\mu_y, \sigma_y^2)$$

$$\Rightarrow X - Y \sim N(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$$

문제 : 남·여 키의 차이가 5 인치 이상일 확률?

$$P(X-Y > 5) = P\left(Z > \frac{5-7}{6.02}\right)$$
$$= P(Z > -0.33)$$

$$P(Z > -0.33) = P(Z < 0.33)$$

$$= 0.629$$

p.416

- 정규분포 선형변환.

X가 한국인 성인 한 명의 체중을 가리킴.

$$X \sim N(\mu, \sigma^2)$$

전제: 한국인 전체가 달나라로 이주했음.

$$Y \sim N\left(\frac{1}{6}\mu, \frac{1}{36}\sigma^2\right)$$

$$Y = \frac{1}{6}X$$

일반화: $X \sim N(\mu, \sigma^2)$

$$Y = aX + b$$

$$\Rightarrow Y \sim N(a \cdot \mu + b, a^2 \cdot \sigma^2)$$

p.417

- 정규분포의 독립관측.

$$X \sim N(\mu, \sigma^2)$$ (한국인 체중) 성인

질문: 한국인 성인 4명을 임의로, 그리고 독립적으로, 선발했을 때, 선택된 4명의 체중의 합의 확률분포는?

$$Y = X_1 + X_2 + X_3 + X_4$$

$$X_i = X \ !$$

$$\Rightarrow Y \sim N(4\mu, 4\sigma^2)$$

일반화:

$$X \sim N(\mu, \sigma^2)$$

$$X_i = X$$

$$Y = X_1 + \cdots + X_n$$

$$\Rightarrow Y \sim N(n \cdot \mu, n \cdot \sigma^2)$$

p.423

- 이항분포를 정규분포로 대체하기

예제

4지선다 문제를 40개 풀어야 하는 상황.

40문제를 모두 찍어서, 30문제 이상 맞출확률?

- $\left(\frac{3}{4}\right)^{40} = (0.75)^{40}$

- $P(X \geq 30) = ?$

X가 맞히는 정답 개수를 가리키는 확률변수.

7장 이항분포 (p330 ~345)

동일한 시도를 n번 반복할때, 특정 사건이

r번 발생할 확률들의 분포

예제 : 4지 선다 문제 40개 찍기

r	0	1		r		39	40
$P(X=r)$	$\left(\frac{3}{4}\right)^{40}$	$40 \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^{39}$	$\cdots$	$P(X=r) = {}_nC_r \cdot \left(\frac{1}{4}\right)^r \cdot \left(\frac{3}{4}\right)^{n-r}$		$40 \cdot \left(\frac{1}{4}\right)^{39} \cdot \left(\frac{3}{4}\right)$	$\left(\frac{1}{4}\right)^{40}$

$40 \to {}_{40}C_1$

$40 \to {}_{40}C_{39}$

$$P(X=r) = {}_nC_r \cdot \left(\frac{1}{4}\right)^r \cdot \left(\frac{3}{4}\right)^{n-r}$$

${}_nC_r$ = n개 중에서 r개를 뽑는 방법의 수 (순서는 무시)

$$= \frac{n!}{r! \cdot (n-r)!}$$

10만

$$P(X=30) = {}_{40}C_{30} \cdot \left(\frac{1}{4}\right)^{30} \cdot \left(\frac{3}{4}\right)^{10}$$

$$\frac{40!}{\cancel{30!} \cdot 10!} = \frac{40 \cdot 39 \cdot \cdots \cdot 31}{10 \cdot 9 \cdot 8 \cdots \cdot 1}$$

$${}_{\text{10만}}C_{100} = \left(\frac{\text{10만}!}{100! \cdot (\text{10만} - 100)!} \right)$$

P

이항분포

- 이산확률분포의 특별한 경우.
- 특정 사건이 발생하는지, 안하는지 딱 두 사건만 대상으로 확률 따짐.
- 동일 행위를 반복적으로, 독립적으로 실행하여 얻은 분포.

$$X \sim B(\underbrace{n}_{\text{반복횟수}},\ \underbrace{p}_{\text{특정 사건이 발생할 확률}})$$

$$X \sim B\left(100, \frac{1}{2}\right)$$

⇐ · 동전을 던져서 숫자가 나오는 사건의 분포
 · 주사위를 던져서 짝수가 〃 〃 〃

$$X \sim B\left(40, \frac{1}{4}\right)$$

r	0	1		r		39	40
$P(X=r)$			..				

$X \sim B(40, \frac{1}{4})$

$P(X=r) = {}_nC_r \cdot \left(\frac{1}{4}\right)^r \left(\frac{3}{4}\right)^{40-r}$

n개 r개를 선택하는 경우 (조합)

선택된 r개의 문제를 모두 맞출 확률

나머지 n-r개의 문제를 모두 틀릴 확률.

$${}_nC_r = \frac{n!}{r! \cdot (n-r)!}$$

① n개 중에 n　1개를 선택하는 경우의 수 : n

② (n-1)　"　" : n-1

③ (n-2)　"　" : n-2

$n \cdot (n-1) \cdot (n-2)$

40 문제 중에 3 문제를 선택할 수 있는 경우의 수:

$$\frac{40 \cdot 39 \cdot 38}{3!} = \frac{40!}{3! \cdot 37!}$$

1	5	7
5	1	7
7	5	1
1	7	5
5	7	1
7	1	5

$$_nC_r = \frac{n!}{r!\,(n-r)!}$$

$$X \sim B\left(40, \frac{1}{4}\right)$$

$$P(X \geq 30) = P(X=30) + P(X=31) + \cdots + P(X=40)$$

$$= {}_{40}C_{30} \cdot \left(\frac{1}{4}\right)^{30} \cdot \left(\frac{3}{4}\right)^{10} +$$

$${}_{40}C_{31} \cdot \left(\frac{1}{4}\right)^{31} \cdot \left(\frac{3}{4}\right)^{9} +$$

$$\vdots$$

$${}_{40}C_{40} \cdot \left(\frac{1}{4}\right)^{40} \cdot \left(\frac{3}{4}\right)^{0}$$

예) ① $X \sim B\left(1000, \frac{1}{4}\right)$

$P(X \geq 300) = ?$

② $X \sim B\left(10^5, \frac{1}{4}\right)$

$P(X \geq 30000) = ?$

10만!

- - -

9장의 423쪽

- 이항분포를 정규분포로 대체하기

$X \sim B(40, \frac{1}{4})$ | $Y \sim B(10^5, \frac{1}{4})$

$P(X \geq 30) = ?$ | $P(Y \geq 30{,}000) = ?$

- 해결책 : 정규분포

$X \sim B(n, p)$ 이고 $\begin{cases} n \cdot p > 5 \text{ 고}, \\ n \cdot (1-p) > 5 \end{cases}$

$\Rightarrow$ X는 정규분포를 따름.

$X \sim N(\mu, \sigma^2)$

$\mu = n \cdot p$, $\sigma^2 = n \cdot p \cdot (1-p)$

r	0	1		r			n
P(X=r)			..				

$\Rightarrow E(X) = n \cdot p$

$Var(X) = n \cdot p \cdot (1-p)$

$$X \sim B(1, p)$$

$r=0$	0	1
$P(X=r)$	$1-p$	p

$$\Rightarrow \begin{cases} E(X) = 0\cdot(1-p) + 1\cdot(p) = p \\ Var(X) = p\cdot(1-p) \end{cases}$$

$$\Downarrow$$

n번 독립관측 $\quad X_i = X$

$$E(X_1 + \cdots + X_n) = n \cdot E(X)$$
$$= n \cdot p$$

$$Var(X_1 + \cdots + X_n) = n \cdot Var(X)$$
$$= n \cdot p \cdot (1-p)$$

$$=$$

$$X \sim B(n, p), \quad n\cdot p,\ n\cdot(1-p) > 5$$

$$\Rightarrow X \sim N(np,\ n\cdot p\cdot(1-p))$$

$$X \sim B\left(40, \frac{1}{4}\right) \Rightarrow \begin{array}{l} 40 \cdot \frac{1}{4} > 5 \\ 40 \cdot \frac{3}{4} > 5 \end{array}$$

$$\Rightarrow X \sim N(10, 7.5)$$

?

$$P(X \geq 30) = P\left(Z \geq \frac{30-10}{\sqrt{7.5}}\right)$$

$$= \cdots !$$ (정규분포표에서 확인)

조심해야 하는 부분이 있음.

p. 430

예) $X \sim B(12, \frac{1}{2})$

$P(X<6) = ?$

$X \sim N(6, 3)$

정규분포화

$P(X<6) = P(Z < \frac{6-6}{\sqrt{3}})$

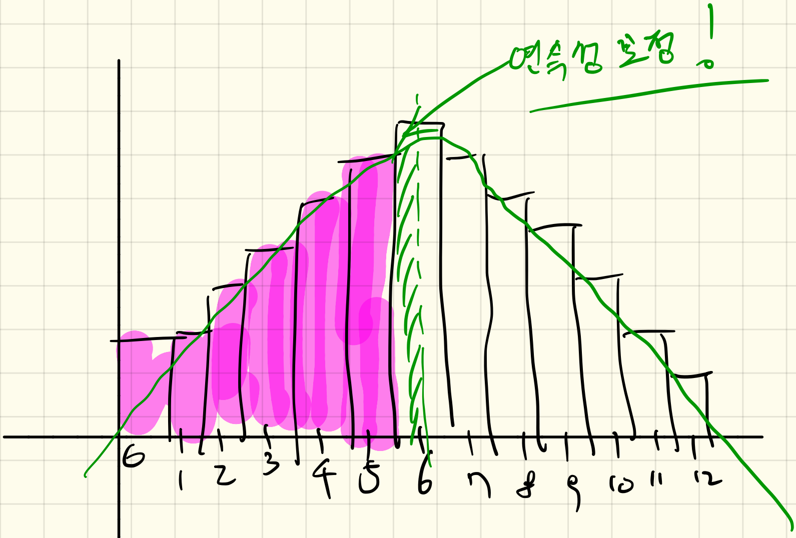
$= P(Z<0)$

$= 0.5$

$P(X=0) + P(X=1) + \cdots + P(X=5)$

$= {}_{12}C_0 \cdot (\frac{1}{2})^{12} + {}_{12}C_1 \cdot (\frac{1}{2})^{12} + \cdots + {}_{12}C_5 \cdot (\frac{1}{2})^{12}$

$= 0.387$



$P(X<6)$

⇒ 연속성 보정:

$$P(X<5.5\cancel{6}) = P\left(Z<\frac{5.5-6}{\sqrt{3}}\right)$$

$$= P(Z<-0.29)$$

$$= 0.3859$$

$$P(X \geq 30) = P(X=30) + P(X=31) + \cdots + P(X=40)$$

r	0	1		r		39	40
$P(X=r)$			..				

$$P(X=r) = {}^{n}C_r \cdot p^r \cdot (1-p)^{n-r}$$

7장: 이항분포. (p.333 ~ 345)

동일한 시도를 n번 할 때, 특정 사건이 r번 나올 확률들의 분포.

예제) 4지 선다.

$p = \frac{1}{4}$

$X \sim B(40, \frac{1}{4})$

$X \sim B(n, \frac{1}{2})$

이항분포는 이산 확률분포에 속함.

$P(X = r) = ?$

$P(X \geq r) = ?$

① $X \sim B(n, p)$

② $n \cdot p > 5$

③ $n \cdot (1-p) > 5$

$\Big\}$ $X \sim N(E(X), Var(X))$

$X \sim B(40, \frac{1}{4}) \Rightarrow X \sim N(10, 10 \cdot \frac{3}{4})$

$10 \cdot \frac{3}{4} = 7.5$

$40 \cdot \frac{1}{4} = 10 > 5$

$40 \cdot \frac{3}{4} > 5$

$X \sim N(n \cdot p, n \cdot p \cdot (1-p))$

?

$X \sim \underline{B(n, p)}$

$X \sim B(40, \frac{1}{4})$

$Y \sim B(1, \cancel{\tfrac{1}{4}}^{\,p})$

r	0	1
$P(Y=r)$	$\cancel{\frac{3}{4}}$ $1-p$	$\cancel{\frac{1}{4}}$ p

$$E(Y) = 0 \cdot \frac{3}{4} + 1 \cdot \frac{1}{4} = 1 \cdot \frac{1}{4}$$

$$Var(Y) = E(Y^2) - (E(Y))^2$$
$$= 1 \cdot \frac{1}{4} \cdot \frac{3}{4}$$

일반화

$$E(Y) = p = 0 \cdot (1-p) + 1 \cdot p$$

$$Var(Y) = p \cdot (1-p) = \quad - \quad -$$

$$E(X) = E(Y_1 + \cdots + Y_{\cancel{40}^{\,n}}), \quad Y_i = Y$$

$$= 40 \cdot E(Y) = \cancel{40}^{\,n} \cdot \cancel{\tfrac{1}{4}}\, \underline{p}$$

$$Var(X) = Var(Y_1 + \cdots + Y_{40})$$

$$= 40 \cdot Var(Y) = \cancel{40}^{\,n} \cdot \cancel{\tfrac{1}{4}}^{\,p} \cdot \cancel{\tfrac{3}{4}}^{\,1-p}$$

$$X \sim B(n, p)$$

$$\Downarrow$$

$$E(X) = n \cdot p$$

$$Var(X) = n \cdot p \cdot (1-p)$$

$$X \sim B(n, p)$$

$$n \cdot p, \quad n \cdot (1-p) > 5$$

$$\Rightarrow X \sim N(\underline{n \cdot p}, \underline{n \cdot p \cdot (1-p)})$$

예)

$$X \sim B\left(600, \frac{1}{6}\right)$$

$$600 \cdot \frac{1}{6} \cdot \frac{5}{6} = 100 \cdot \frac{5}{6} = \frac{500}{6} = 83$$

$$\sqrt{83} = 9 \ldots$$

$$\Rightarrow X \sim N(100, 83)$$

$\Rightarrow$ 숫자 1이 성공횟수

$$[100 - 9, 100 + 9]$$

사이에서 발생.

$$X \sim B\left(40, \frac{1}{4}\right)$$

질문 : $P(X \geq 30) = ?$

① $40 \cdot \frac{1}{4}, \ 40 \cdot \frac{3}{4} > 5$

$$\Rightarrow X \sim N(10, 7.5)$$

$$\Rightarrow P(X \geq 30) \simeq P\left(Z \geq \frac{30-10}{\sqrt{7.5}}\right)$$

$= \ldots$! (정규분포표 확인)

근사치 정도가 얼마인가?

예제)

$$X \sim B\left(12, \frac{1}{2}\right)$$

질문 : $P(X<6) = ?$

$$= \left(P(X=0) + P(X=1) + \cdots + P(X=5) \right)$$

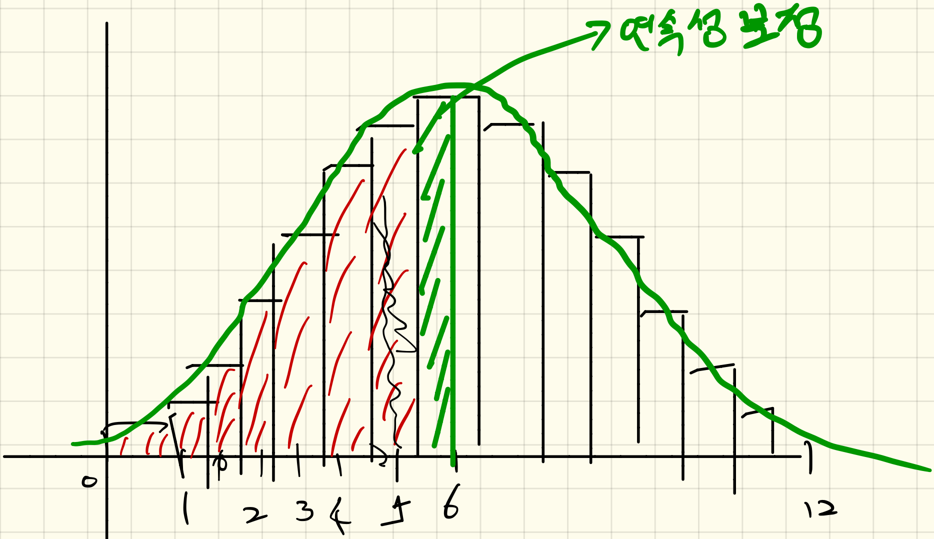
이산 $\Rightarrow$

$$= {}_{12}C_0 \cdot \left(\frac{1}{2}\right)^{12} + {}_{12}C_1 \left(\frac{1}{2}\right)^{12} + \cdots + {}_{12}C_5 \left(\frac{1}{2}\right)^{12}$$

$$= 0.387$$

$$12 \cdot \frac{1}{2} > 5 \Rightarrow X \sim N(6, 3)$$

$$P(X<6) = P\left(Z < \frac{6-6}{\sqrt{3}}\right) = P(Z<0) = 0.5$$



$X \sim B(12, \frac{1}{2})$

$P(X < 6)$

$\Rightarrow X \sim N(6, 3)$

$P(X < 5.5) = P\left(Z < \frac{5.5 - 6}{\sqrt{3}}\right)$

$= P(Z < -0.29)$

$= 0.3859$

예)　$X \sim B(12, \frac{1}{2})$

$P(X > 6) = ?$

$\Rightarrow$ $X \sim N(6, 3)$

$P(X > 6.5) = !$