확률과 통계 - 통계

2020.11.24

실상빈

확률과 통계 - 통계

1. 이산확률분포의 확률분포

∘ 확률변수

쉽게 표현하자면 각 사건에 하나의 실수값을 대응시키는 함수

확률변수 X가 어떤 값 x를 가질 확률을 $P(X=x)$ 라고 나타낸다.

확률 X가 x인 값.

∘ 이산확률변수

확률변수 X가 가지는 값이 유한 개이거나 셀 수 있을 때

유한 개가 아니면

X를 이산확률분포 라고 한다.

∘ 확률질량함수

이산확률변수 X와 X가 가지는 확률 사이의 대응 관계를 나타내는 함수

↳ 경우의 수에 대한 수를 대입하면 확률이 나온다.

함숫값 = 확률 → 모두 셀 수 있고 다 더하면 1

$(0 \le p \le 1)$ 때! ↳ 모든 경우의 확률 합은 1이다

∘ 이산확률분포의 기댓값(평균), 분산, 표준편차.

- 기댓값(평균) → 내가 이 일로 얼마의 결과를 예상할 수 있는가 = 평균

$E(X)$라고 쓰고 (확률변수 x 각각의 확률)을 모두 더해서 구한다.

Expectation.

- 분산 → 얼마나 흩어져있는가?

$V(X)$라고 쓰며 각각의 확률과 평균의 차이를 제곱한 값의 평균이다.

편차. ↳ 편차를 다 더하면 0 이니 제곱하여

※ 제곱의 평균 과 평균의 제곱을 알면 분산을 구할 수 있다.

$E(X^2)$ $\{E(X)\}^2$

$$\boxed{E(X^2) - \{E(X)\}^2 = V(X)}$$

→ 많이 쓰니 꼭 기억!

제곱의 평균이 먼저!

- 표준편차 $(\sqrt{분산})$ → 분산은 제곱한 값이니 다시 복구.

표준편차는 한 번에 구하기는 어렵고 분산을 통해 구한다

$\sigma(X)$ 라고 쓴다.

∘ $aX+b$ 의 형태에서의 값

$E(aX+b) = aE(X)+b$: 모든 X가 a배가 되고 b만큼 증가하므로 평균도 따라감

$V(aX+b) = a^2V(X)$: 모든 X가 a배가 되면 분산은 a^2배 (제곱의 평균) 다 같이 +b면 퍼진정도 차이는 X

$\sigma(aX+b) = |a|\sigma(X)$: a가 음수라도 $|a|$배 퍼지는 거니까 절댓값으로

∘ 이항분포

$B(n, p)$로 나타낸다.

횟수 확률

이항분포 $B(n,p)$를 보고 바로 확률질량함수를 쓸 수 있다.

$$P(X=x) = {}_nC_x \cdot p^x (1-p)^{n-x}$$

∘ 이항분포로 평균, 분산, 표준편차 구하기

$B(n,p)$에서

평균 : $n \times p$

분산 : $n \times p \times (1-p)$

표준편차 : $\sqrt{n \times p \times (1-p)}$

2. 연속 확률분포의 확률분포

◦ 연속확률변수

확률변수 X가 모든 실수의 값을 가질때

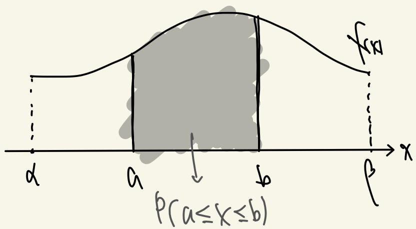
X를 연속확률변수라고 한다.

◦ 확률밀도함수

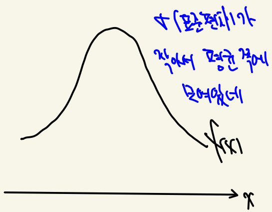
$\alpha \le X \le \beta$ 에서 모든 실수의 값을 가지는 연속확률변수 X에

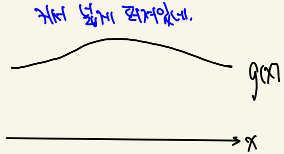
대하여 이 구간에서 정의된 함수 $f(x)$를 확률밀도함수라고 한다.

넓이 : 확률



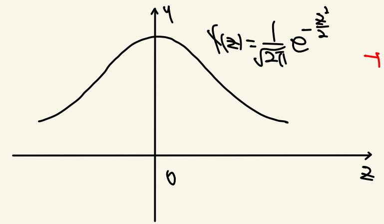
※ 표준편차의 대소에 의한 그래프의 차이





◦ 표준정규분포

평균이 0이고 표준편차가 1인 정규분포를 표준정규분포라고 한다.



$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

→ 알 필요 절대 X

확률변수 X가 (m, σ^2)을 따를 때, 확률변수

$$Z = \frac{X - m \text{ 편차}}{\sigma \text{ 표준편차}}$$ 은 표준정규분포 $N(0,1)$을 따른다.

차이가 표준편차의 몇 배인가?

└ 편차가 표준편차의 몇 배인지가 핵심

ex) 확률변수 X가 정규분포 $N(9, 3^2)$을 따를 때

$P(6 \le X \le 12)$의 값을 구해보자.

A. 6과 12는 모두 평균인 9와 3 차이가 난다.

3은 표준편차 ×1 이므로 이를 표준정규분포로 바꾸면

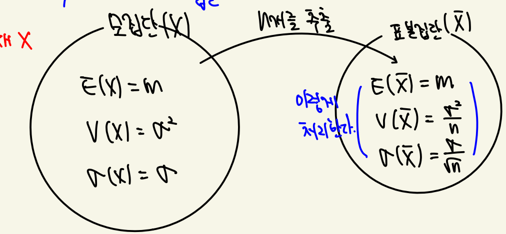
$$P\left(\frac{6-9}{3} \le Z \le \frac{12-9}{3}\right) = P(\underbrace{-1}_{-1\text{표준편차}} \le Z \le \underbrace{1}_{+1\text{표준편차}})$$

$$= 2P(0 \le Z \le 1) = 2 \times 0.3413 = 0.6826$$

3. 통계적 추정

◦ 모집단과 표본집단

母 뽑은



이렇기에 모집단이 $N(m, \sigma^2)$를 따른다면

n명을 뽑은 표본집단은 $N(m, \frac{\sigma^2}{n})$를 따른다.

◦ 모평균의 추정

표본의 자료를 보고 모집단의 정보를 추정하는 것이다.

그렇기에 표본의 표준편차를 모집단에 사용한다.

σ ← 모집단의 표준편차.

$\sqrt{n}$ ← 표본의 표준편차.

- 95% → k = 1.96

$$\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \le m \le \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

신뢰구간 : $2 \times 1.96 \times \frac{\sigma}{\sqrt{n}}$

- 99% → k = 2.58

$$\bar{X} - 2.58 \times \frac{\sigma}{\sqrt{n}} \le m \le \bar{X} + 2.58 \cdot \frac{\sigma}{\sqrt{n}}$$

신뢰구간 : $2 \times 2.58 \times \frac{\sigma}{\sqrt{n}}$

6월 모의고사에서 64점을 받고 그 뒤 2개월 동안 공부를 하지 않았다. 아니 할 수 없었다.

어느 사람이 28점이 떨어지고도 평온할 수 있을까

그래서 7월 모의고사 날에 학교에 시험을 보러 가지 않았다. 출석은 위해서 3시 30분에 가서 아프다는 말밖에 할 수 없었다.

9월 모의고사도 69점을 받았고 담임과의 상담에서는 그나마 잘 할 줄 아는 축구를 해보라는 말까지 나왔다. 친구들이 드릴 한다고 하니 삼.

그래도 나는 축구로는 최고가 될 수 없다는 것을 알기에 거절하고 인강을 결제한다.

단순 수업 수로만 90개 X 1시간의 수업과 300페이지가 넘는 책을 10월 한 달에 끝냈다.

심지어 시립대 논술 보러 가서 시립대 기출 안 풀고 뉴런 풀었음...

10월 모의고사도 60점대를 받고 결국 2018. 11. 15, 즉 수능을 보러 갔다.