모집단 : 대한민국 성인 남성 (편의상 2천만명으로 침)

모평균 : 진짜 평균. 전수조사 안하면 모르는 값. m이라고 씀.

모표준편차 : 진짜 표준편차. 전수조사 안하면 모르는 값. σ라고 씀.

표본평균 : n명 뽑아서 걔들로 구한 평균. $\bar{X}$라고 씀.

표본표준편차 : n명 뽑아서 걔들로 구한 표준편차. s라고 씀.

1

표본평균 $\bar{X}$는 일종의 확률변수이다

n명 임의추출해서 구할 때마다 표본평균값이 확률적으로 정해지니까 $\bar{X}$는 연속확률변수이다.

우리가 증명할 필요는 없고 수학자들이 알아낸 바에 따르면

표본평균은 (어차피 문제풀땐 상관없이 대체적으로) 기댓값이 m, 표준편차가 $\frac{\sigma}{\sqrt{n}}$인 정규분포를 따른다.

여기서 조심할 건, '표본평균의 표준편차'는 그냥$\frac{\sigma}{\sqrt{n}}$일 뿐이고, '표본표준편차'인 s와는 하등 무관함.

하등 무관함. 하등 무관함.

2

조사한 표본평균 하나로 모평균이 어떤 범위에 들어올 확률을 개략적으로 구하는 방법

표본평균이 정규분포를 따른다는 점을 이용하여

정규분포 확률 구하는 방식을 역이용하여 적절히 수식을 변형하면

$\bar{X} - a\frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + a\frac{\sigma}{\sqrt{n}}$

라는 부등식이 성립할 확률이 '적당한 상수'에 의해 정해진다는 사실을 알 수 있다.

(a = 1.96이면 95% , 2.58이면 99%)

3

2번은 탁상공론. 인생은 실전이다. 대충 되면 그만이야.

2번은 참이긴 하겠지만 현실에 적용할 수 없다. 시그마는 아까 말했듯이 '전수조사 안하면 모르는 값'이니까.

그래서 아예 모르는 값인 시그마 대신에, 지금 알고 있는 값인 S를 대체하여 써도 허용됨. 안그럼 추정 자체를 못하니까.

따라서 현실적으로 우리가 세우는 부등식은

$\bar{X} - a\frac{\sigma}{\sqrt{n}} \le m \le \bar{X} + a\frac{\sigma}{\sqrt{n}}$

인 척 하는

$\bar{X} - a\frac{\mathrm{s}}{\sqrt{n}} \le m \le \bar{X} + a\frac{\mathrm{s}}{\sqrt{n}}$

이고, 이걸로 풀면 됨.

끝.