

정의 - definition

모집단(분석의 대상이 되는 집단)에서 K개의 샘플을 N회 복원추출한다고 할 때 K가 약 30이상의 너무 적지 않은 개수이고 시행횟수 N이 많아 질수록 K의 평균들이 정규분포에 점점 가까워진다.

다시 풀어서 정리하자면

통계학 - Statistics 에서 랜덤샘플링(Uniform Random Sampling)으로 K개(약 30개이며 많을 수록 좋다) 이상의 데이터 포인트를 추출한다.

2. 샘플에서 평균을 구한다.
3. 반복해서 계속 추출한다. 단 이때 복원추출이다. 즉 앞의 샘플링에 추출된 데이터가 다시 추출될 수도 있다.
4. 다시 평균을 구한다.
5. 계속해서 적당히 반복한다. 최소 100회 이상은 수행해야 하며 많이 할수록 좋다.

위에서 구한 많은 평균값들의 분포는 정규분포이고 위의 평균값의 평균(평균값들에서 다시 평균을 구한 것)은 모집단의 평균과 일치한다.

여러 표본평균들의 평균

용도

용도를 모르기 때문에 흔히 개념만 알고 넘어가기 쉽다. 앞의 설명에 포함된 내용이지만 포괄적인 용도는 다음과 같다.

한 개의 표본으로

1. 적은 샘플수로 모집단의 평균을 매우 정확하게 알아낼 수 있다.
2. 적은 샘플수로 모집단의 분산을 매우 정확하게 알아낼 수 있다.
3. 모든 데이터를 정규분포로 만들 수 있다.
4. 위의 샘플의 평균들은 정규분포하기 때문에 정규분포를 가정하는 통계적 가설검정 및 기법을 여기에 모두 적용해도 아무런 무리가 없다.

· 표본의 크가 클수록 표본평균의 기대값이 모평균에 가까워진다.

· 즉, 표본의 크가 클수록 한 개의 표본평균이 모평균에 더욱 근사하게 된다.

여기에서 가장 중요한 것은 1번이다. 전수검사를 하지 않고 모집단의 평균을 정확히 알아내는 것은 생각보다 매우 어려운 일이다.

첨부해 주신 내용은 중심극한에 대한 이해를 돕기 위해 진행한 시뮬레이션을 진행한 것입니다

무한히 샘플을 뽑을 수 없으니 충분히 큰 숫자로 생각되는 100개를 뽑아서 정규분포를 따르는지 간접적으로 보여 준 것입니다

즉 교육적인 목적으로 진행한 것입니다

분석에서는 충분한 표본수가 있을 경우에는 중심극한 정리가 성립한다고 받아들이고 분석을 합니다

★ 아니면 가능도함수를 그려보기도 하는데, 이것은 기초통계수준을 벗어나는 방법이니 자세한 설명을 하지는 않겠습니다

교수님 답장을 받은 후 계속 공부해보니, '하나의 표본에 포함된 표본의 수가 충분하다면, 표본평균이 모평균에 충분히 가까울 것이라고 본다.'라는 교수님의 답변이 조금씩 이해가 되는 것 같습니다.

즉 표본의 크기가 크면($n > 30$) 표본평균(\bar{x})의 분포가 $N(\mu, s^2/n)$ 을 따르기 때문에(즉, 대부분의 표본평균들이 표본평균의 평균 주위로 많이 몰리는 정규분포를 띄기 때문에),

'한 개의 표본집단에서 나온 표본평균은 표본평균의 평균(결국, 모평균)에 근사할 가능성이 매우 높고, 표본의 수가 충분한 한 개의 표본에서 구해진 표본평균을 모평균으로 여긴다.'라고 생각할 수 있다는 것 맞습니까?

네 맞습니다

특히 s^2/n 이므로 표본수가 커질 수록 분산은 0으로 수렴합니다

즉 표본평균이 모평균에 가까워진다고 볼 수 있습니다