

어드벤처디자인:빅데이터기초분석과정정보활용

5주차 과제

금융정보통계학과 20173204 곽명빈

Q1. \bar{X} 가 불편추정량임을 보여라

$$\begin{aligned} E(\bar{X}) &= E\left(\sum_{i=1}^n \frac{x_i}{n}\right) \\ &= E\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) \\ &= \frac{\mu + \mu + \cdots + \mu}{n} \\ &= \mu \end{aligned}$$

Q2. 다음을 논하라

연습1-3에서 모집단, 표본을 정의하여라.

만약 모집단이 명확하지 않거나 표본이 잘못 표기되어 있는 경우는 다시 정정하여 주세요.

연습1]

동아일보 1988년 11월 전두환 전 대통령의 대 국민 사과 성명에 대한 여론 조사를 실시하였다.
이 조사는 서울, 부산, 인천, 광주, 대전 등 6대 도시의 만 20세 이상 성인 남녀 1,600명을 전화번호부에서 추출하여 실시되었다.

모집단은 전국민이고,

표본은 5대 도시의 만 20세 이상 성인남녀 1600명이다.

정치적 문제이기 때문에 발신자가 편향된 정치성향을 가지고 있다면 편향된 답으로 질문을 유도할수 있기 때문에 편의가 발생할 수 있다.

연습2]

스포츠서울은 매주 지난 일주일 동안 방영된 국내 TV 프로 중에서 가장 재미있고 유익한 프로를 독자투표로 선행하여 "TV 프로 top 20"을 발표하고 있다.
투표요령은 신문에 인쇄된 투표 용지에 좋았다고 생각한 프로 5개를 기록한 뒤 오려서 우편엽서에 붙여서 우송하는 것이다.
1988년 12월에 실시된 35회 투표에는 6,211명이 참가했다.

모집단은 전국민 이다.

표본은 1988년 12월에 실시된 35회 투표자 6211명이다.

이것이 잘못된 이유는 투표할수 있는사람이 스포츠서울 신문을 구독하는 사람들이기 때문에 스포츠에 관련된 프로의 답변이 나올확률이 굉장히 높아 편의가 발생할수 있다. 따라서 옳은 여론 조사가 되려면, 국내 스포츠 TV프로 중 가장 재미있고 유익한 프로를 독자투표로 선행하여 TV 프로 TOP20을 정해야 한다.

연습3)

여론조사를 값싸게 그리고 빠른 시간에 할 수 있는 방법이 있다.

여러 개의 전화번호를 준비하여 그 중 반을 찬성에 나머지 반은 반대에 배정한다. TV를 통하여 질문을 하고, 찬·반에 해당하는 전화번호를 돌리기만 하면 되며 조사자는 각 번호에 걸려온 전화횟수만 기록하면 되는 것이다.

실제로 미국의 ABC-TV는 1980년 레이건과 카터의 TV 토론 후에 1,800개의 전화번호를 사용하여 여론조사를 해 보았다.

그 결과 레이건의 2 대 1로 토론에서 승리한 것으로 나왔다. 그러나 CBS 뉴스가 실시한 표본조사는 레이건의 승리 쪽에 44%, 카터의 승리 쪽에 36%라는 결과를 보였다.

전화를 이용한 위의 조사방법에는 왜 편의가 발생할 가능성이 높은지를 설명하시오.

모집단은 전국민 이다

표본은 ABC-TV토론을 본 시청자 1800명이다.

TV프로 특성상 중간부터 보기 시작하면, 처음에 어떤 내용에 대해 토론을 했는지 볼 수 없기 때문에, 중간부터 본 시청자는 특정 후보자에게 쏠릴 가능성이 존재한다. 특정 시간 이후에 한 후보자가 말을 잘 했다면, 중간부터 본 시청자들은 처음에 어떤 일들이 있었는지 모르고 그 지지자에게 호감을 표시하여 전화를 했을 것이다. 그러므로 TV프로를 통한 여론조사는 옳지 않다고 할 수 있다.

2-2.각자 2개의 추가 사례를 조사하여 만들어보세요.

1)1936년 미국 대통령 선거에서 공화당 후보 알프레드 랜던(Alfred Landon) vs 민주당 후보 프랭클린 루스벨트(Franklin Roosevelt) 선거에서 리터러리 다이제스트라는 곳에서 1000만명에게 설문조사를 하여 240만 명에게서 응답을 받았다.

이때 모집단은 미국국민이고

표본은 리터러리 다이제스트 구독자 240만명이다.

당시 상황을 보면 잘못된 것을 이해할 수 있는데, 당시 소득이 낮은 유권자는 민주당, 소득이 높은 유권자는 공화당을 선호하였다. 하지만 정기 구독자와 전화와 자동차 등 당시 고가의 제품을 소유한 덕에 마케팅 담당자의 명단에 있던 사람들로 중산층 이상의 사람들로만 조사가 된 것이다. 이 때문에 소득이 낮은 유권자들의 의견을 제대로 조사에대해 반영 시킬수 없었기 때문에, 이 조사는 잘못된 것이다.

2) 임금근로일자리별 소득 분포 분석

본 자료는 2015년 기준 일자리행정통계를 보자.

남자의 평균 소득은 390만원, 중위소득은 300만원이며, 여자의 평균소득은 236만원, 중위소득은 179만원임.



모집단은 노동을 하는 전국민의 소득이고이고

표본은 보험에 가입한 임금근로자가 점유한 약 1500만개 일자리에 대한 소득이다.

임금에 미치는 다른 요소들을 통제하지 않은 채 단순히 평균소득과 중위소득을 통해 남녀 두 모집단의 소득을 비교하였다. 남자와 여자가 일을 하는 직업도 다르고, 노동 강도도 다르기 때문이다. 올바른 조사가 되려면 남녀 사이에 공통된 조건을 가지고 조사를 해야 한다. 표본을 같은 직장에서 같은 일을 하는 남녀로 정정을 해야 올바른 조건 이 된다. 그렇지 않고 이렇게 단순히 비교하면, 남녀 의 수입의 차이가 볼 수 있게 되어 오류가 생긴다.

