

텍스트이(가) 표시된 사진

자동 생성된 설명

1. 국가통계의 이용을 개인, 기업, 정부의 측면에서 설명하라.

정부에서 특정 목적을 가지고 국가적인 정보를 주기적으로 만드는데 이를 국가통계라고 한다. 국가통계를 활용하여 개인과 기업 그리고 정부는 합리적인 의사결정 또는 선택을 할 수 있다. 개인은 국가통계를 활용하여 합리적인 판단을 할 수 있다. 예를 들면 기대수명을 조사한 국가통계를 활용하여 은퇴시점을 결정할 수 있다. 기업은 국가통계를 활용하여 시장을 파악할 수 있고 이를 이를 통해 전략을 세울 수 있다. 예를 들어 혼인 및 이혼 자료 혹은 인구 총 조사를 활용하여 1인 가구를 위한 제품의 생산을 늘리는 판단을 할 수 있다. 정부는 국가통계를 보고 현재 상태를 파악하고 이를 토대로 정책을 추가적으로 만들 수 있다. 정책에 대한 국민의 합의를 도출 할 수 있을 뿐 아니라 정책에 실제로 효과가 있는지 판단할 수 있다. 예를 들어 정부가 혼인율과 출산율이 감소하고 있다는 통계를 본다면, 인구 수를 유지하기 위해 혼인 및 출산을 장려하는 정책을 펼칠 수 있다. 또한 출산율 통계를 추적 관찰하여 출산 장려정책의 실효성에 대해서도 파악할 수 있다.

2. 통계학의 세가지 역할을 설명하라.

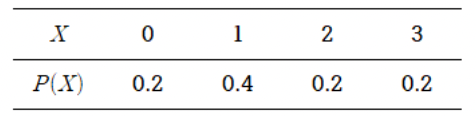
첫번째로는 통계학을 바탕으로 자료를 수집할 수 있다. 시청률이나 여론조사 뿐만 아니라 인사이트를 얻기 위해 통계를 활용하여 연구를 할 수 있다. 두가지 약이 있을 때 어떠한 약의 효과가 우월한지, 약의 효과가 정말 있는 것인지 실험을 하고 통계를 활용하여 이에 대해 정보를 얻을 수 있다.

두번째로는 자료를 요약하고 설명할 수 있다. 예를 들어 기업에 지원하는 사람이 상당수 있을 때 남녀 비율은 어떻게 되는지, 출신 대학교는 어떻게 되는지, 어떠한 연령대로 분포해 있는지 요약해서 일목요연하게 알아볼 수 있다. 수집한 자료를 그래프나 수치를 통해 보여줌으로써 다른 사람에게 편하게 전달할 수 있다.

세번째로는 자료로부터 인사이트를 도출하거나 미래를 예측하고 대비할 수 있다. 개인

, 정부, 기업은 데이터를 통해서 다음 스텝을 어떻게 밟아야 할지 생각해볼 수 있다. 예를 들어 대통령은 본인의 지지율을 바탕으로 다음 이대로의 행보를 이어갈지 아니면 변화를 줄지 판단할 수 있다. 기업의 경우에는 세계 경제의 동향을 파악하여 공격적으로 사업을 늘려 나갈지 긴축을 할지 판단할 수 있다. 국가 역시 물가상승률을 파악하여 금리를 올릴지 동결할지 내릴지를 정할 수 있다.

3. 캠핑카 판매지점의 판매량이 다음과 같은 확률분포를 가질 때 기대되는 평균 판매기대수 E(X)를 구하라.



0x0.2+1x0.4+2x0.2+3x0.2=1.4

E(X)=1.4

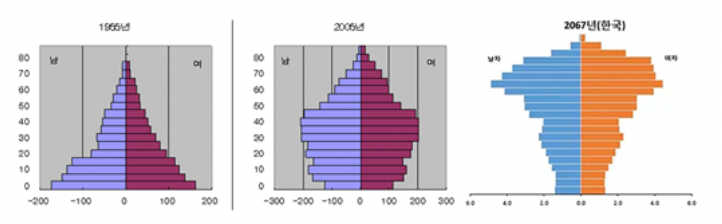
4. 다음 자료로 정규분포의 모수에 의한 변화를 같은 X축상의 분포도 그림과 함께 설명하라.

쇼지이(가) 표시된 사진

자동 생성된 설명

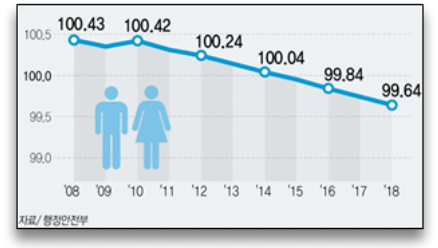
B군은 A군보다 평균이 높기 때문에 X축 상에서 더 오른쪽에 위치한다. 또한 B군의 분산이 A군보다 낮기 때문에 평균을 중심으로 더 밀집한 형태를 보인다. A군은 평균 20을 전후로 상대적으로 평평한 모습을 보이는 반면 B군은 상대적으로 가파른 모습을 보이고 있다. 또한 모집단의 수가 커질수록 점차 연속적인 모습을 보인다. n=30일 때에는 연속적인 모습을 보이지 않으나 n수가 점차 커지면서 그래프가 연속적인 모습을 보인다. 또한 n수가 커질수록 대칭적인 모습을 보이며 봉우리가 뚜렷하게 보인다. 또한 동떨어진 듯한 자료가 상대적으로 적다.

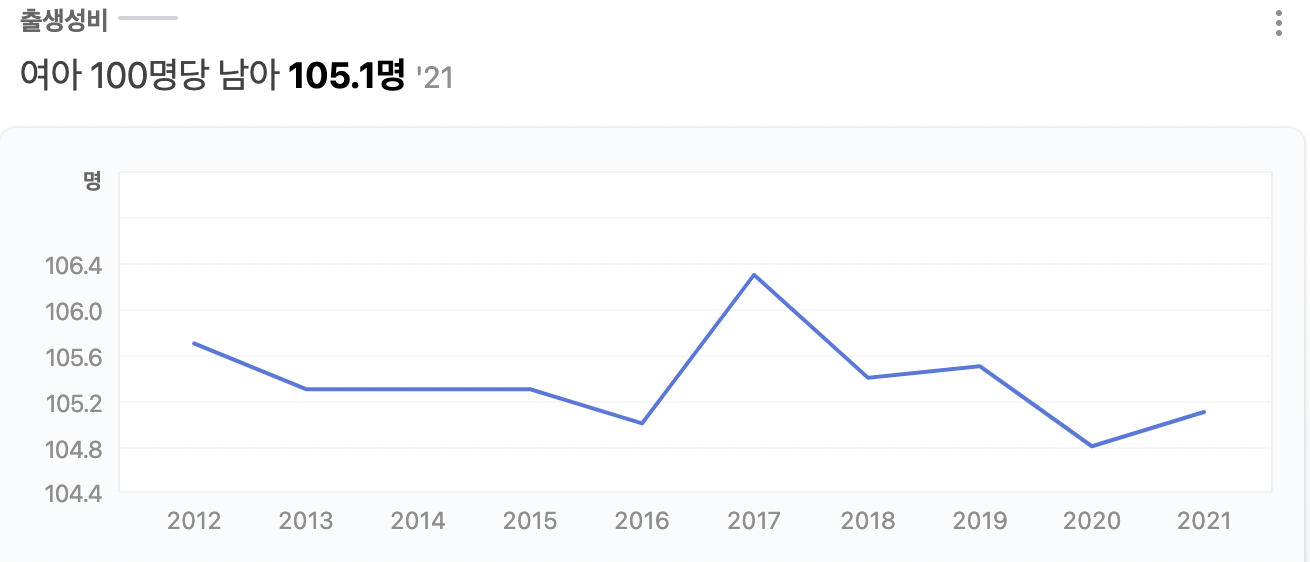
5. 다음의 인구피라미드에 대해 비교 설명하라



왼쪽은 과거의 우리나라 인구피라미드 가운데는 현재의 우리나라 인구피라미드 오른쪽은 미래의 우리나라 인구피라미드를 나타내고 있다. 왼쪽의 인구피라미드는 시대가 전후이기 때문에 사망인구는 많았다. 그렇기에 나이가 많아질수록 인구가 줄어드는 모습을 보이며, 1960년도는 베이비 붐으로 출생자수가 크게 증가하였기 때문에 아래는 넓은 모습을 보여주고 있다. 그렇기 때문에 위가 좁고 아래가 넓은 피라미드 모양의 인구형태를 보이게 된다. 시간이 점차 지나면서 위생 개념이 증가하고 의료기술이 발달함으로써 점차 수명이 증가하게 된다. 1970년도에는 농경시대가 접어들고 산아제한을 하면서 출생자수가 증가하는 모습을 보이지 않고있다. 현재에는 여성의 사회진출 그리고 결혼을 기피하려는 문화가 합쳐져 급격한 저출산의 모습을 보이고 있기에 가운데의 그래프는 방추형의 모습을 보이고 있다. 평균수명이 늘어나고 출생률이 저하되었기 때문이다. 마지막으로 오른쪽 그래프는 한국의 미래를 예측한 인구피라미드이다. 현재 출산율이 0.8명까지 내려갔기 때문에 급격하게 인구가 줄어드는 모습을 보이고 있다. 현재에는 사망자 수가 출생아 숫자보다 단지 많아지는 수준이지만 그 격차를 계속해서 벌릴 것으로 보인다. 늦게 낳고, 안 낳고, 덜 낳기 때문에 어떠한 변화가 없는 경우 오른쪽 같은 그래프의 모습을 보일 것이라고 기대된다.

6. 출생성비의 정의와 최근 성비 변화에 대한 다음의 통계자료에 대해 설명하라.





출생성비와 성비는 다른 개념이다. 출생성비는 여아 100명당 남아의 수를 말하며 (남아출생아/여아출생아)x100으로 나타낸다. 생물학적으로는 103에서 107사이가 정상이라고 한다. 예전에는 남아선호상으로 인해 출생성비가 115 이상까지 올랐다고 한다. 특히 셋째 이상의 성비는 200이 넘었다. 허나 현재는 그러한 문화적인 기조가 변했기 때문에 105를 기점으로 평균회귀하는 모습을 보이고 있다.



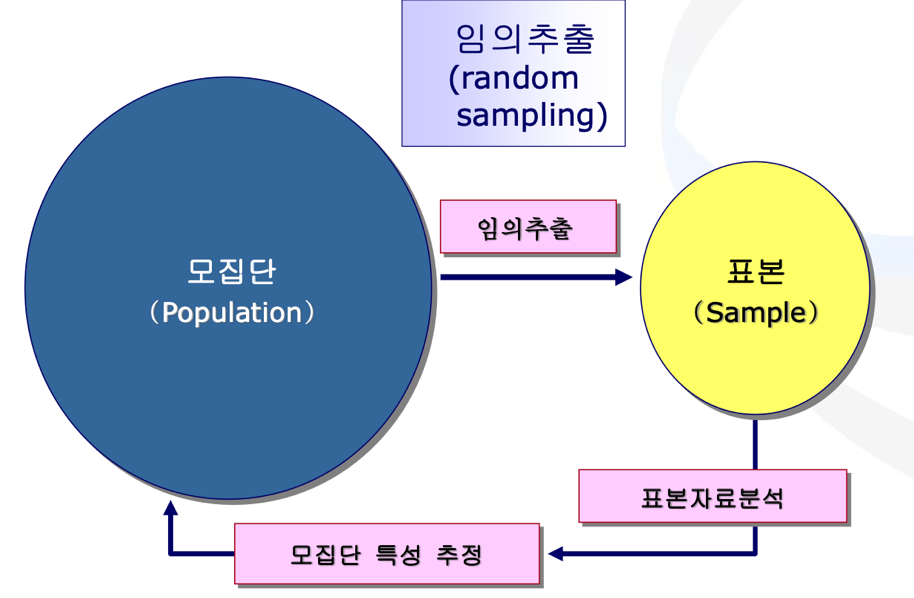
성비는 여자 100명당 남자의 수를 나타낸다. 출생성비는 105명인 반면, 성비는 100명을 유지하고 있다. 출생아의 남아 사망률이 더 높기도 하고 남성의 평균수명이 여성의 평균수명보다 6년정도 낮기 때문에 발생하는 현상이다. 우리나라의 경우 100아래로 내려갔지만 외국인까지 고려한 경우에는 100을 넘어 남성이 여성보다 조금 많다고 한다.

7. 통계분석 초기에 자료의 이상치를 파악하기 위한 검토 방법을 설명하라.

변수가 하나인 경우에는 표준화 점수, 통계적 가설검정, 사분위수 범위를 사용해서 이상치를 탐색할 수 있다. 표준화 점수는 평균을 중심으로 표준편차가 얼마나 떨어져 있는지 확인하는 방법이며, 표준편차의 절대값이 3보다 큰 경우에는 이상치라고 대개 판단한다. 평균과 표준편차를 활용하기 때문에 이상치의 값을 사용한다는 단점이 있어 중앙값과 절대편차를 활용하는 방법도 있다. 통계적 가설검정은 이상치로 예상되는 값을 제거해 나가는 과정을 반복적으로 수행하는 방법이다. 대표적으로 카이제곱 검정, 딕슨 Q검정이 있다. 사분위수 범위의 1.5배를 넘는 경우 대개 이상치라고 판단하며, 상자그림을 이용해서 쉽게 알 수 있다.

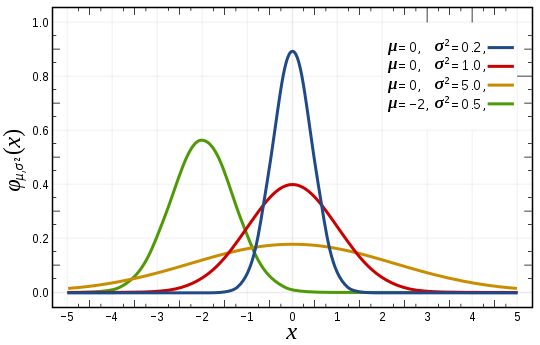
변수가 두개인 경우에는 회귀진단, 미할라노비스 등이 있으며 회귀진단은 회귀식에 대한 전반적인 검토를 하고 영향을 미치는 극단치를 알아보는 것이다. 마할라노비스 거리는 자료 간의 거리를 활용하여 관측치가 평균보다 벗어나는 것을 측정한다. 변수의 개수를 고려하여 임계점을 넘어가는 경우 이상치로 생각한다.

8. 모집단과 표본의 관계와 임의추출을 도표로 설명하라.



조사대상이 되는 전체 집단을 모집단이라고 하며, 모집단에서 추출되어 모집단을 대표하는 일부분을 표본이라고 한다. 추출에는 다양한 방법이 있지만 가장 기본적인 방법은 임의추출이다. 말그대로 무작위로 추출하는 방법이며, 추출하는 연구자의 주관을 개입시키지 않는다. 연구자는 모집단의 전체를 조사할 수 없기 때문에 표본을 연구함으로써 모집단의 특성을 엿볼 수 있다. 모집단 전체를 연구하게 되면 많은 시간과 노력이 들기 때문이다. 표본이 너무 적게 되면 통계적으로 유의하지 않기 때문에 유의할 만큼의 표본을 선발하게 된다. 모평균, 모분산, 모표준편차를 알기에는 힘들기에 표본평균, 표본분산, 표본표준편차인 표본자료분석을 하여 모집단을 알 수 있다.

9. 정규분포의 특징에 대해 설명하라.



정규분포의 특징은 다음과 같다. 정규분포는 평균을 중앙으로 하여 좌우 대칭이다. 정규분포의 모양은 평균과 표준편차인 두가지 변수만을 통해서 나타낼 수 있다. 대부분의 관측치는 평균에 몰려 있으며, 하나의 극단을 가진다. 표본평균은 표본이 어느정도 크기만 하면 정규분포를 따르게 된다. 정규분포는 이항분포의 확률을 계산하는데 유용하다. 양극단으로 갈수록 X축에 접근하지만 닿지는 않는다.

10. 자료의 그래프에 의한 표현의 장점과 그 종류를 설명하라.

정보를 글자로만 나타나게 되면 한눈에 파악하기 힘들다. 점이나 선 또는 색깔을 사용하여 가시성을 높이게 되면 정보를 한눈에 파악하기 쉽다. 그래프의 변화를 보여주는 경우에는 관측치의 변화와 추세를 직관적으로 알아볼 수 있다. 정보를 전달하는 것에도 편하다. 사람을 설득하는 데에도 용이하다. 수많은 말보다 관측치를 그래프로 나타내는 경우 상대방은 일목요연하게 파악할 수 있어 정보전달이 쉽다. 자료에 따라 적절한 그래프의 형태가 있다. 대표적으로 막대그래프, 꺾은선그래프, 원그래프, 산점도, 방사형 그래프 등이 있다.

막대그래프는 관측 수치를 막대모양의 길이로 변환한 그래프이다. 크고 작음을 한눈에 알아볼 수 있기 때문에 이해하기 편하나 시간의 흐름을 나타낼 수 없다는 단점이 있다. 반면 꺾은선형 그래프는 관측 수치를 점을 표시하고 그러한 점을 잇는 그래프 이며 시간에 따라 변화하는 것을 기록할 때 유용하다. 원그래프는 각 관측치에 대한 백분율을 나타내거나 백분율을 단순 비교할 때 유용하게 활용할 수 있다. 방사형 그래프는 다각형을 통해 나타내는 그래프로 세 개 이상의 측정값의 치수를 시각적으로 나타내는 경우 유용하다. 산점도는 직교 좌표계를 이용해서 점을 찍는 방식의 그래프이다. 두 개의 변수의 관계를 나타내는 데에 유용하다. 변수 간의 관계를 수학적 모델로 확인해서 상관관계를 파악해 볼 수 있다.

Reference

1. 이태림, 생활과 통계(Statistics in Life), 한국방송통신대학교출판문화원(2015)

2. 김진휘, 이상치 탐색을 위한 통계적 방법(Statistical methods for outlier detection))

3. 위키백과, <https://ko.wikipedia.org/wiki/%EC%A0%95%EA%B7%9C_>

%EB%B6%84%ED%8F%AC#:~:text=%EC%A0%95%EA%B7%9C%EB%B6%84%ED%8F%AC%EB%8A%94%20%ED%8F%89%EA%B7%A0%EA%B3%BC,%EC%B6%95%EC%97%90%20%EB%8B%BF%EC%A7%80%EB%8A%94%20%EC%95%8A%EB%8A%94%EB%8B%A4.