1.안녕하세요. 저희들은 불로소득과 부채가 총 생활비에 미치는

영향을 연구주제로 선정하여 데이터 분석을 진행하게 되었고요.

저는 앞부분 설명을 맡은 김정훈이라고 합니다.

그럼 발표를 시작하도록 하겠습니다.

2.발표에 앞서 간단하게 팀원 소개를 하겠습니다.

전반적으로 다함께 코드 작성 및 PPT를 만들었습니다만 그 중에서도 각자

비중있는 역할을 맡은 업무는 이름 밑에 기재를 하게되었습니다.

3.목차에 들어가기 전 간단하게 연구 개요에 대해 설명드리겠습니다.

현대사회에서 금융 부채 문제와 소득 양극화가 이슈가 되고 있어

받은 데이터파일을 확인해본 결과, 연구 주제와 관련된 변수를 확인 할 수 있었습니다.

본 연구를 통해 가구별 총생활비를 부채, 불로소득과의 상관관계를 조사하여 도출된

결론을 바탕으로 다양한 분야와 단체의 도움이 되는 참고 자료를

제공하고자 분석을 진행하였습니다.

4.이어서 목차에 대해 설명드리겠습니다.

5.첫번째로 데이터 준비입니다.

준비한 데이터를 불러오고 데이터 형식 확인 및 데이터 전처리를 진행할 예정입니다

6.두번째로 이상치 확인 및 정제입니다.

불러온 데이터에 이상치가 있는지 그래프 및 히스토그램을 이용하여 확인하고

이상치가 있을 경우 처리를 할 예정입니다.

7.세번째로 신뢰구간 확인입니다.

이 부분은 사실 필수사항은 아니지만 조원들간에 이야기를 해 본 결과

공부했던 내용을 리마인드 한다는 느낌으로 추가하게 되었고요.

보시는바와 같이 모집단은 대한민국 전체 가구로 가정하고

표본은 데이터 엑셀 파일에 있는 7865가구가 대상입니다.

8.네번째로 정규성 및 상관관계 확인입니다.

해당 데이터셋이 정규성을 만족하는지 확인해보고 산점도행렬 및 히트맵으로

변수들간 상관관계를 확인해볼 예정입니다.

9.다음으로 주성분 분석입니다.

주성분 분석을 하기 전 먼저 데이터 정규화를 진행하고 이후

pca분석을 진행하여 해석력이 약한 변수를 확인해볼 예정입니다.

10.여섯번째로 다중선형회귀분석입니다.

선생님께 배웠던 get\_df 함수를 응용하여 경우의 수를 전부 계산해보고

비교 및 분석을 진행할 예정입니다.

11.마지막으로 결과입니다.

분석된 결과를 토대로 최종 결론 도출 및 시각화해서 어느정도 예측하는지

확인할 예정입니다.

12.처음으로 확인할 것은 변수 선정입니다.

저희들 같은 경우에는 총 15개의 변수를 선정하게 되었고요.

이 중에서 총생활비를 종속변수로 하고 그 외에는 독립변수로 선정했습니다.

독립변수 같은경우에는 처음에 말씀드린것처럼 불로소득과 부채

2가지 카테고리로 나눠집니다. 또한 모든 변수들의 단위는 만원으로 통일되어

있는 점 참고 부탁드립니다.

13.다음으로 변수 선정을 하던 중 저희들이 선정한 재난지원금,공적연금 같은 것도

불로소득에 포함되는지 의문을 갖게되었습니다. 왜냐하면 연구주제와 상관없는

변수를 포함하여 데이터 분석을 하게 될 경우 잘못된 결과를 도출할 수 있기

때문인데요. (다음으로 넘겨주세요)

14.먼저 불로소득에 대해 간단히 설명드리자면

말 그대로 일을 하지 않고 얻는 소득을 의미합니다. 불로소득에 범주로는

보시는 바와 같이 여러가지가 포함되는데요. 여기서 공적연금과 기초연금은

연금속성으로 간주하고 재난지원금은 복지제도의 혜택으로 간주해서

이번 조사 대상에 같이 포함하게 되었습니다.

15.네, 제 설명은 여기까지이고요. 나머지 부분은 재현님이 이어서

말씀드릴 수 있도록 하겠습니다. 감사합니다.