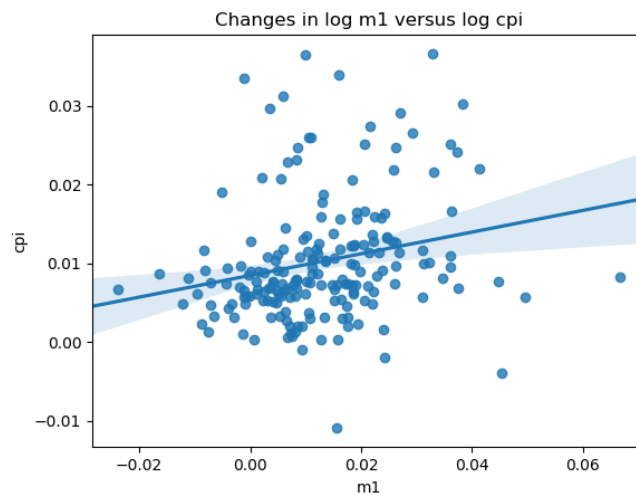


- (10점) "I\_have\_a\_dream.txt"를 utf-8으로 읽고 철자의 길이가 가장 긴 단어를 출력하시오. 같은 길이의 단어가 여러 개인 경우 모든 단어를 전부 출력하시오. (구두점은 제외하고 순수한 알파벳으로 구성된 단어이어야 함)
- (10점) 표준정규분포 (평균=0, 표준편차=1)에서 500개의 난수를 생성하고 크기 순서로 50개씩 10개의 그룹으로 나눈 다음 각 그룹의 평균과 표준편차를 계산해서 출력하시오.
- (10점) 데이터에서 "macrodata.csv"에서, "m1"의 (분기)증가율이 4분기 뒤의 "cpi"의 (분기)증가율에 미치는 영향을 나타내는 다음의 회귀식을 포함한 그래프를 그리시오. 증가율은 모두 log 증가율을 의미한다. 그림의 제목을 반드시 넣으시오.



- (20점) 1부터 5까지 5개의 도시가 있다. 각 도시와 도시를 연결하는 항공노선의 가능한 총 개수는 10개이다. 각 노선은 두 도시의 번호를 가지는 tuple로 나타낼 수 있으며, 노선의 출발지와 도착지의 구분은 없다. 예를 들어, 도시 1과 3을 연결하는 노선은 (1,3)으로 표현되며, (1,3)과 (3,1)은 동일한 노선을 의미한다.
  - (10점) 노선들의 집합을 list로 전달받아서 모든 도시가 하나로 연결되었는지 판단하는 함수를 만드시오. (모든 도시가 연결되었다면 True를 아니면 False를 리턴 한다)
 

예를 들어, [(1,2),(3,4),(4,5),(3,5)] 가 입력되면 False를 리턴하고, [(1,2),(2,3),(2,4),(2,5)]가 입력되면 True를 리턴함.
  - (10점) 모든 도시를 하나로 연결하는 최소의 노선의 개수는 4개이다. 4개의 노선으로 모든 도시를 연결하는 모든 경우의 수를 출력하시오. (1,2)와 (2,1) 같이 중복되지 않도록 반드시 노선의 앞자리는 작은 숫자의 도시를 나타내도록 하시오.

5. (20점) "SPX.csv"와 "NKY.csv"는 각각 S&P500과 니케이225 지수의 일간 데이터이다. 이를 이용해서 다음을 구하시오.
- A. (5점) 두 데이터 파일을 읽어서 하나의 dataframe으로 결합하고, (두 데이터 중 하나라도 포함된 모든 날짜를 포함시킴) 각 지수의 결측치의 개수를 구하시오.
- B. (5점) 2016년~2018년 사이의 매월 주가지수의 최대값과 최소값의 차이를 계산하시오.
- C. (5점) 2000년 1월 1일 이후 두 지수의 월간 수익률 (매월 2번째주 목요일 종가 기준)의 상관계수를 구하시오. 해당일에 주가가 없으면 해당일 이전 가장 최근 거래일의 주가를 이용하시오.
- D. (5점) 2000년 1월 1일 이후 두 지수 중 하나라도 결측치가 있으면 해당일의 데이터를 제거하고 나서, 두 지수의 일간수익률의 250일 윈도우의 이동 표준편차를 계산하고 하나의 axes에 선그래프로 함께 도시하시오.
6. (30점) "flight\_data.csv"는 2009년도 미국의 국내선 항공 운행 기록이고, 11개의 컬럼은 다음과 같다. 출발공항과 도착공항 간에 월간 운행기록과 출발 도시와 도착 도시의 인구수 정보를 담고 있다.

Col	Short name	Type	Description
1	Origin	String	Three letter airport code of the origin airport
2	Destination	String	Three letter airport code of the destination airport
3	Origin City	String	Origin city name
4	Destination City	String	Destination city name
5	Passengers	Integer	Number of passengers transported from origin to destination
6	Seats	Integer	Number of seats available on flights from origin to destination
7	Flights	Integer	Number of flights between origin and destination (multiple records for one month, many with flights > 1)
8	Distance	Integer	Distance (to nearest mile) flown between origin and destination
9	Fly Date	Integer	The date (yyyymm) of flight
10	Origin Population	Integer	Origin city's population as reported by US Census
11	Destination Population	Integer	Destination city's population as reported by US Census

- A. (5점) 데이터에 포함된 출발공항과 도착공항 별로 관측치의 개수는 각각 몇 개인가? 또한 출발은 있지만 도착은 없는 공항과 반대로 출발은 없지만 도착은 있는 공항을 각각 출력하시오.
- B. (5점) 2009년 1년간 합계로 가장 많은 승객(Passenger)이 탑승한 노선 상위 5개와 각각의 Passenger 합계를 출력하시오.
- C. (5점) 월별로 전체 노선에 대한 탑승률 (=승객수/좌석수)을 구하고, 탑승률이 가장 높은 월과 가장 낮은 월을 구하시오.
- D. (5점) 출발 공항 별 운항회수(Flights)의 연간 합계와 출발도시의 인구수의 상관계수를 구하시오.

- E. (10점) 출발과 도착 도시명에 포함된 주(state)의 약자 (예를 들어 뉴욕주는 NY) 를 이용해서 출발 주와 도착 주를 확인할 수 있다. 출발 주와 도착 주 별로 연간 탑승객 수의 합계를 계산하고, 출발과 도착을 각각 세로, 가로 축으로 하는 테이블을 만드시오. 결과를 "result.csv"로 저장하도록 하시오.

테이블 일부 예시:

	NY(도착)	CA	TX
NY(출발)	xxx,xxx	xxx,xxx	xxx,xxx
CA	xxx,xxx	xxx,xxx	xxx,xxx
TX	xxx,xxx	xxx,xxx	xxx,xxx