회귀분석실습 중간고사

2015. 10. 22.

1. test.txt 파일 사용하여 test 데이터 셋을 생성하고, 다음 물음에 답하여라.

(1) 만약 section이 1이라면 test1과 test2의 점수를 각각 10점, 14점씩 감점하고, section이 2라면 test1과 test2의 점수를 각각 3점, 5점씩 올려 test5, test6이라는 변수를 생성하여라. [10점]

(2) section이 1인 경우만을 추출하여 test1부터 test6까지의 총합을 ARRAY문을 사용하여 구하여라. [15점]

2. preg.txt와 expo.txt 파일을 사용하여 preg, expo 데이터 셋을 생성하고, 다음 물음에 답하여라.

(1) 두 데이터 셋을 DATE변수의 이름을 각각 PREG\_DATE, EXPO\_DATE로 변경한 후, name에 따라 MERGE하여 bothid 데이터 셋을 생성하여라. [10점]

(2) (1)의 데이터 셋을 생성할 때, PREG\_DATE와 EXPO\_DATE가 결측치가 아닌 값만 저장하여라. [10점]

3. timerec.xlsx 파일을 사용하여 timerec 데이터 셋을 생성하고, 다음 물음에 답하여라.

(1) employee, phase, hours 변수에 각각 ‘고용인’, ‘분석단계’, ‘시간’ 이라는 라벨을 생성하여라. [5점]

(2) employee가 Chen이면 emp1이라는 데이터 셋에 저장하고, employee가 Stewart라면 emp2라는 데이터 셋에 저장하고 employee 변수는 지워라. [10점]

(3) emp1의 데이터 셋으로부터 RETAIN문을 사용하여 각각의 분석단계에서 걸린 시간의 합을 구하여라. 걸린시간의 합은 hoursum으로 나타내고, 분석단계별로 계산된 걸린 시간의 최종 합만을 저장하여라. Hour 변수는 지워라. [20점]  
(Hint : last.phase를 사용하여 저장하고, hoursum을 초기화시킨다.)

4. grunfeld.txt 파일을 사용하여 grunfeld 데이터 셋을 생성하고, 기술통계량을 구하는 프로시저를 사용하여 sort에 따른 invest, stock, value의 기술통계량을 구하여라. [10점]  
(변수명은 sort, invest, stock, value 다.)

5. 평균이 10이고 분산이 4인 정규분포의 PDF를 구하고 그림으로 나타내어라. 그 때의 x값은 1부터 20까지다. (plot 생성시 옵션은 symbol i=spline v=none; 으로 지정한다) [10점]