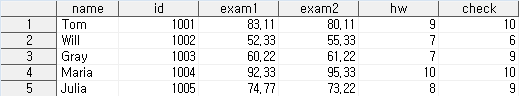
**2013년도 회귀분석1 실습 중간고사**

|  |  |  |
| --- | --- | --- |
| 학과: 통계학과 | 학번: 22121423 | 이름: 김정환 조교 |

**1. 원본 데이터와 주어진 조건을 이용하여 수정된 데이터를 생성하는 프로그램을 작성하라. (20점)**

원본 데이터는 성적에 대한 정보를 가지고 있다. 아래를 참고하여 **직접 원본 데이터를 생성**하시오.



조건)

① score = exam1\*0.3 + exam2\*0.5 + hw + check

② grade

: score≥ 90이면 A, 80≤score<90이면 B, 70≤score<80이면 C, 60≤score<70이면 D, score<60이면 F

③ 변수 레이블 수정.

exam1->중간고사, exam2->기말고사, hw -> 과제, check->출석, score->최종성적, grade->등급

수정된 데이터)



**data** temp\_dataset1;

input name $ id exam1 exam2 hw check;

cards;

Tom 1001 83.11 80.11 9 10

Will 1002 52.33 55.33 7 6

Gray 1003 60.22 61.22 7 9

Maria 1004 92.33 95.33 10 10

Julia 1005 74.77 73.22 8 9

;

run;

**data** dataset1;

set temp\_dataset1;

score = exam1\***0.3** + exam2\***0.5** + hw + check ;

if score >=**90** then grade='A';

else if score >=**80** then grade='B';

else if score >=**70** then grade='C';

else if score >=**60** then grade='D';

else grade='F';

label exam1 = '중간고사' exam2 = '기말고사'

hw = '과제' check = '출석' score = '최종성적' grade = '등급';

run;

**2. 원본 데이터와 주어진 조건을 이용하여 수정된 데이터를 생성하는 프로그램을 작성하라. (20점)**

원본 데이터는 농구 경기 기록에 대한 정보를 가지고 있고, 변수 정보는 표에 제시되어있다.



|  |  |  |  |
| --- | --- | --- | --- |
| team: 팀 이름 | name : 선수이름 | posi : 포지션 | backn : 등 번호 |
| reb : 리바운드 횟수 | assist : 어시스트 횟수 | foul : 반칙 횟수 | time : 총 출장시간 |
| freeth : 자유투 성공횟수 | freetry : 자유투 시도횟수 | suc2p : 2점슛 성공횟수 | try2p : 2점슛 시도횟수 |
| suc3p : 3점슛 성공횟수 | try3p : 3점슛 시도횟수 |  |  |

조건)

① 먼저, **‘총 득점’**에 대한 변수를 생성하시오. (변수 이름은 자유)

총 득점 = 자유투에 의한 득점 + 2점 슛에 의한 득점 + 3점 슛에 의한 득점

(자유투는 성공 시 1점, 2점 슛은 성공 시 2점, 3점 슛은 성공 시 3점으로 계산된다.)

② 다음으로, **‘슛 정확도’**에 대한 파생변수를 생성하시오. (변수 이름은 자유)

슛 정확도 = 총 득점 / (자유투 시도횟수 + 2점 슛 시도횟수 + 3점 슛 시도횟수)

(슛 정확도는 소수점 2째 자리까지만 표현됨.)

③ 새로운 변수에 대한 레이블 처리를 반드시 하고, 수정된 데이터에 표시된 변수들만 나타내시오.

수정된 데이터)



/\*데이터 읽기\*/

libname MIDTEST 'D:\MIDTEST';

/\*개인별 총 득점\*/

**proc** **sql**;

create table temp\_dataset2 as

select \*, (freeth\***1** + suc2p\***2** + suc3p\***3**) as score

from MIDTEST.basket;

**quit**;

/\*슛 정확도\*/

**proc** **sql**;

create table dataset2 as

select team, name, posi, score, score/(freetry+try2p+try3p) as exact format **4.2** label='슛 정확도'

from temp\_dataset2 ;

**quit**;

**3. 원본 데이터와 주어진 조건을 이용하여 수정된 데이터를 생성하는 프로그램을 작성하라. (30점)**

2개의 원본 데이터는 의료기관 정보를 가지고 있고, 각 데이터의 변수 정보는 아래 표에 제시되어있다.



|  |  |  |
| --- | --- | --- |
| YNO : 요양기관 고유번호 | NO : 명세서 고유번호 | MONEY : 진찰비 |



|  |  |
| --- | --- |
| YNO : 요양기관 고유번호 | DOCTOR : 의사수 |

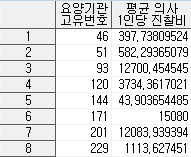
요양기관 고유번호는 각 요양기관을 구별하는 역할을 한다. 마찬가지로 명세서 고유번호는 명세서를 구별하는 역할을 한다. medical\_a.sas7bat는 한 요양기관 별 다수의 명세서를 가지고 있는 형태이다. medical\_b.sas7bat는 한 요양기관 별 단일 관측값 만을 가지고 있다.

조건)

① 데이터 셋을 이용하여, 요양기관 별로 **‘의사 1인당 진찰비’**의 **평균**을 계산하여 아래의 결과가 되도록 데이터를 수정하시오. 단, 두 데이터 셋에 대해 동일한 요양기관 만을 사용하여 분석하시오.

② 새로운 변수에 대한 레이블 처리를 반드시 하시오.

수정된 데이터)



(이하 생략, 총 1906개 관측치)

/\*데이터 읽기\*/

libname MIDTEST 'D:\MIDTEST';

/\*요양기관 별 의사 1인당 진찰비 평균\*/

**proc** **sql**;

create table dataset3 as

select distinct t1.YNO, avg(t1.MONEY/t2.DOCTOR) as AVG\_MON\_DOC label='평균 의사 1인당 진찰비'

from MIDTEST.medical\_a as t1, MIDTEST.medical\_b as t2

where t1.YNO = t2.YNO

group by t1.YNO;

**quit**;

**4. 원본 데이터와 주어진 조건을 이용하여 수정된 데이터를 생성하는 프로그램을 작성하라. (30점)**

원본 데이터는 소비자의 구매 내역에 대한 정보를 갖고 있고, 변수 정보는 표에 제시되어있다.



|  |  |  |
| --- | --- | --- |
| CUST\_ID : 고객 고유번호 | GENDER : 고객 연령 | AGE : 나이 |
| JOB\_CD : 직업 종별코드 | ITEM\_CATE : 상품 분류 | PRICE : 상품 가격(구매액) |

조건)

① 고객 별 구매액 평균이 30,000원 이상인 고객만 추출하여 우수 고객으로 삼고자 한다.

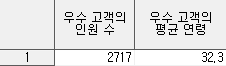
추출된 우수 고객의 **인원 수**와 **평균 연령**을 계산하여 아래의 결과가 되도록 데이터를 수정하시오.

② 새로운 변수에 대한 레이블 처리와 소수점 처리를 반드시 하시오.

참고)

SQL문에서 요약통계량 계산시 관측치의 카운트(개수 세기)는 N(변수명)을 이용하여 계산 가능함.

수정된 데이터)



/\*데이터 읽기\*/

libname MIDTEST 'D:\MIDTEST';

/\*고객 별 구매액 평균\*/

**proc** **sql**;

create table temp1\_dataset4 as

select DISTINCT CUST\_ID, AGE, avg(PRICE) as AVG\_PRICE

from MIDTEST.item

group by CUST\_ID;

**quit**;

/\*우수 고객만 추출\*/

**proc** **sql**;

create table temp2\_dataset4 as

select \*

from temp1\_dataset4

where AVG\_PRICE >= **30000**;

**quit**;

/\*우수 고객의 요약통계량 계산\*/

**proc** **sql**;

create table dataset4 as

select count(CUST\_ID) as N\_CUST label='우수 고객의 인원 수',

avg(AGE) as AVG\_AGE format **4.1** label='우수 고객의 평균 연령'

from temp2\_dataset4;

**quit**;