/\*IV值\*/

/\*\*\*\*\*\*\*\*\*\*\*跑IV前需要進行相應設置\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

%LET INDATA=short\_list;\*原數據集;

%LET EXP\_VAR='NO','Bad2';\*不參與計算的變量如ID,好壞標籤等,注意大小寫;

%LET SPE\_VAR=.,-9999,9999,-9993;\*數值變量特殊值，需要做更改;

%LET N=5;\*除特殊值外，數值變量最大分組數;

%LET GB=Bad2;\*好壞標籤字段，用0,1表示;

/\*\*\*\*\*\*\*\*\*\*\*跑IV前需要進行相應設置\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

%MACRO IV;

PROC CONTENTS DATA=&INDATA

OUT=VARLIST(KEEP=NAME TYPE WHERE=(NAME NOT IN (&EXP\_VAR))) NOPRINT;

RUN;

DATA DATA\_CHAR DATA\_NUM;

SET VARLIST;

IF TYPE=1 THEN OUTPUT DATA\_NUM;

ELSE OUTPUT DATA\_CHAR;

RUN;

%LET DISD=%SYSFUNC(OPEN(DATA\_CHAR));

%LET NCHAR=%SYSFUNC(ATTRN(&DISD,NOBS));

%PUT &NCHAR;

%LET C=%SYSFUNC(CLOSE(&DISD));

%LET DISD=%SYSFUNC(OPEN(DATA\_NUM));

%LET NNUM=%SYSFUNC(ATTRN(&DISD,NOBS));

%PUT &NNUM;

%LET C=%SYSFUNC(CLOSE(&DISD));

/\*數值變量計算IV\*/

%IF &NNUM>0 %THEN %DO;

DATA \_NULL\_;

SET DATA\_NUM END=LAST;

CALL SYMPUT(COMPRESS('NVAR'||\_N\_),NAME);

IF LAST THEN CALL SYMPUT('N\_NOBS',\_N\_);

RUN;

%MACRO NUM\_IV;

DATA IV\_NUM\_VAR;

FORMAT LABEL $100. NAME $32. IV;

STOP;

RUN;

%DO I=1 %TO &N\_NOBS;

DATA DATA\_REG DATA\_NREG;

SET &INDATA(KEEP=&&NVAR&I &GB);

IF &&NVAR&I IN (&SPE\_VAR) THEN OUTPUT DATA\_NREG;

ELSE OUTPUT DATA\_REG;

RUN;

%LET TDISD=%SYSFUNC(OPEN(DATA\_NREG));

%LET TNNUM=%SYSFUNC(ATTRN(&TDISD,NOBS));

%PUT &TNNUM;

%LET TC=%SYSFUNC(CLOSE(&TDISD));

PROC RANK DATA=DATA\_REG OUT=RANK GROUPS=&N;

VAR &&NVAR&I;

RANKS BIN;

RUN;

PROC MEANS DATA=RANK N MAX MIN NOPRINT;

CLASS BIN;

VAR &&NVAR&I;

OUTPUT OUT=RANK1;

RUN;

PROC TRANSPOSE DATA=RANK1(WHERE=(NOT MISSING(BIN) AND \_STAT\_ IN ('MAX','MIN')))

OUT=RANK2(DROP =\_NAME\_ RENAME=(MIN=START MAX=END));

BY BIN;

ID \_STAT\_;

VAR &&NVAR&I;

RUN;

DATA RANK3;

SET RANK2;

LENGTH BIN\_GROUP $30.;

BIN\_GROUP=COMPRESS(\_N\_||'、'||START||'-'||END);

RUN;

PROC SQL;

CREATE TABLE RANK4 AS

SELECT

BIN,

SUM(CASE WHEN &GB=0 THEN 1 ELSE 0 END) AS G\_CNT,

SUM(CASE WHEN &GB=1 THEN 1 ELSE 0 END) AS B\_CNT

FROM RANK

GROUP BY

BIN;

QUIT;

DATA RANK5;

FORMAT BIN \_LABEL\_ BIN\_GROUP;

MERGE RANK3 RANK4;

BY BIN;

RUN;

%IF &TNNUM>0 %THEN %DO;

PROC SQL NOPRINT;

SELECT DISTINCT &&NVAR&I INTO :NREG\_VAR SEPARATED BY ',' FROM DATA\_NREG;

CREATE TABLE DATA\_NREG1 AS

SELECT

"0、&NREG\_VAR" AS BIN\_GROUP FORMAT $30.,

SUM(CASE WHEN &GB=0 THEN 1 ELSE 0 END) AS G\_CNT,

SUM(CASE WHEN &GB=1 THEN 1 ELSE 0 END) AS B\_CNT

FROM DATA\_NREG

GROUP BY

"0、&NREG\_VAR";

QUIT;

DATA RANK6;

FORMAT \_LABEL\_ BIN\_GROUP $30. START END G\_CNT B\_CNT;

SET DATA\_NREG1 RANK5;

KEEP \_LABEL\_ BIN\_GROUP START END G\_CNT B\_CNT;

RUN;

%END;

%ELSE %DO;

DATA RANK6;

FORMAT \_LABEL\_ BIN\_GROUP $30. START END G\_CNT B\_CNT;

SET RANK5;

KEEP \_LABEL\_ BIN\_GROUP START END G\_CNT B\_CNT;

RUN;

%END;

DATA RANK7;

SET RANK6 END=LAST;

IF \_N\_=1 THEN DO;

CUM\_G\_CNT=0;

CUM\_B\_CNT=0;

END;

CUM\_G\_CNT+G\_CNT;

CUM\_B\_CNT+B\_CNT;

IF LAST THEN DO;

CALL SYMPUT('TOTG',CUM\_G\_CNT);

CALL SYMPUT('TOTB',CUM\_B\_CNT);

END;

RUN;

DATA &&NVAR&I;

SET RANK7 END=LAST;

G\_CNT\_PCT=ROUND(G\_CNT/&TOTG,0.0001);

B\_CNT\_PCT=ROUND(B\_CNT/&TOTB,0.0001);

BAD\_RATE=ROUND(B\_CNT/(B\_CNT+G\_CNT),0.0001);

IF B\_CNT\_PCT^=0 AND G\_CNT\_PCT^=0 THEN WOE=ROUND(LOG(G\_CNT\_PCT/B\_CNT\_PCT),0.0001);

ELSE WOE=0;

IV\_SEG=ROUND((G\_CNT\_PCT-B\_CNT\_PCT)\*WOE,0.0001);

IF \_N\_=1 THEN IV=0;

IV+IV\_SEG;

KEEP \_LABEL\_ BIN\_GROUP START END G\_CNT B\_CNT BAD\_RATE WOE IV;

RENAME \_LABEL\_=LABEL;

RUN;

DATA TEMP\_IV;

FORMAT LABEL NAME $32. IV;

SET &&NVAR&I END=LAST;

NAME="&&NVAR&I";

IF LAST;

KEEP LABEL NAME IV;

RUN;

DATA IV\_NUM\_VAR;

SET IV\_NUM\_VAR TEMP\_IV;

RUN;

%END;

%MEND;

%NUM\_IV;

%END;

/\*字符變量計算IV\*/

%IF &NCHAR>0 %THEN %DO;

DATA \_NULL\_;

SET DATA\_CHAR END=LAST;

CALL SYMPUT(COMPRESS('CVAR'||\_N\_),NAME);

IF LAST THEN CALL SYMPUT('C\_NOBS',\_N\_);

RUN;

%MACRO CHAR\_IV;

DATA IV\_CHAR\_VAR;

FORMAT LABEL $100. NAME $32. IV;

STOP;

RUN;

%DO I=1 %TO &C\_NOBS;

DATA DATA\_REG;

SET &INDATA(KEEP=&&CVAR&I &GB);

RUN;

PROC SQL;

CREATE TABLE RANK1 AS

SELECT

&&CVAR&I,

SUM(CASE WHEN &GB=0 THEN 1 ELSE 0 END) AS G\_CNT,

SUM(CASE WHEN &GB=1 THEN 1 ELSE 0 END) AS B\_CNT

FROM DATA\_REG

GROUP BY

&&CVAR&I;

QUIT;

DATA RANK2;

FORMAT LABEL BIN\_GROUP;

SET RANK1 END=LAST;

LABEL=VLABEL(&&CVAR&I);

BIN\_GROUP=COMPRESS(\_N\_||'、'||&&CVAR&I);

IF \_N\_=1 THEN DO;

CUM\_G\_CNT=0;

CUM\_B\_CNT=0;

END;

CUM\_G\_CNT+G\_CNT;

CUM\_B\_CNT+B\_CNT;

IF LAST THEN DO;

CALL SYMPUT('TOTG',CUM\_G\_CNT);

CALL SYMPUT('TOTB',CUM\_B\_CNT);

END;

RUN;

DATA &&CVAR&I;

SET RANK2 END=LAST;

G\_CNT\_PCT=ROUND(G\_CNT/&TOTG,0.0001);

B\_CNT\_PCT=ROUND(B\_CNT/&TOTG,0.0001);

BAD\_RATE=ROUND(B\_CNT/(B\_CNT+G\_CNT),0.0001);

IF B\_CNT\_PCT^=0 AND G\_CNT\_PCT^=0 THEN WOE=ROUND(LOG(G\_CNT\_PCT/B\_CNT\_PCT),0.0001);

ELSE WOE=0;

IV\_SEG=ROUND((G\_CNT\_PCT-B\_CNT\_PCT)\*WOE,0.0001);

IF \_N\_=1 THEN IV=0;

IV+IV\_SEG;

KEEP \_LABEL\_ BIN\_GROUP G\_CNT B\_CNT BAD\_RATE WOE IV;

RUN;

DATA TEMP\_IV;

FORMAT LABEL NAME $32. IV;

SET &&CVAR&I END=LAST;

NAME="&&CVAR&I";

IF LAST;

KEEP LABEL NAME IV;

RUN;

DATA IV\_CHAR\_VAR;

SET IV\_CHAR\_VAR TEMP\_IV;

RUN;

%END;

%MEND;

%CHAR\_IV;

%END;

%MEND;

%IV;