A SAS MACRO FOR DETERMINING THE CAUSES OF INTERACTION IN A TWO WAY MODEL

Edward P. Barry, Schering-Plough
Dar-Shong Hwang, Sandoz, Inc.

1. Introduction

In the analysis of clinical trials examining treatment by investigator interaction is a desirable step before pooling the results across investigators. When the interaction is statistically significant, the statistician would like to know which investigators are responsible for the interaction and how much can be attributed to them. This macro presents two methods for answering these questions.

2. Contrast Approach

One method of determining which investigators are different from the others is to form pairwise contrasts comparing each investigator to all others. We refer to this as pairwise interaction. With two treatments, the pairwise interaction contrast between investigators i and j is made by calculating

$$t_{ij} = \frac{(\bar{y}_{1i} - \bar{y}_{2i}) - (\bar{y}_{1j} - \bar{y}_{2j})}{s\,(1/n_{1i} + 1/n_{2i} + 1/n_{1j} + 1/n_{2j})^{1/2}},$$

a t-statistic with the error variance and degrees of freedom from a two way ANOVA model. One can then find how many times an investigator significantly interacts with the other investigators. Those who interact most frequently with the others are presumably the chief contributors to the overall interaction.

The significance level used is 5%. We prefer the non-multiple comparison procedure so that no adjustment of the p-value is needed. However, the p-level can be changed if adjustment for Bonferroni-type multiple comparisons is preferred by the user.

The SAS code for such a procedure is simply the CONTRAST statement in GLM. SAS is used to generate the appropriate contrasts, since the number of investigators used in the analysis is not usually constant from variable to variable and/or time point to time point.

3. Percent Contribution

When the investigators who are significantly different from the others have been determined, the next question asked is how much of the total interaction sums of squares are they responsible. For the two treatment case, the calculations are made as follows:

let n_{1j} and n_{2j} be the cell sizes , and

$\bar{y}_{1j}$ and $\bar{y}_{2j}$ be the cell means. Then

$$D_j = (\bar{y}_{1j} - \bar{y}_{2j}),$$

$$W_j = (n_{1j} * n_{2j}) / (n_{1j} + n_{2j}),$$

$$\bar{D} = \Sigma W_j * D_j / \Sigma W_j,$$

$$C_j = W_j (D_j - \bar{D})^2,$$

$$PC_j = C_j / \Sigma C_j * 100.$$

the term PC_j is the percent contribution of the investigator j to the interaction sums of squares.

Similar calculations can be made for the case of more than two treatments, but the value can be easily found by squaring the difference between the predicted values from the model without interaction and the model with interaction and summing over the investigator. These values are easily computed using OUTPUT and PREDICTED features of GLM and calculating the differences and sums in a data step.

4. The Macro

The following steps are taken by the macro:

- delete the blocks with empty or small cells
- write the pairwise interaction contrasts to a temporary data set
- perform a GLM for a two way model with interaction on the raw data, and use the above contrasts
- read the output from GLM and save the p-levels for the pairwise interactions
- perform a GLM for a two way model without interaction on the cell means
- display the output.

5. An Example

The example used is from cake data presented by Johnson (1976), who refers to Li (1964) as the source. For purposes of this paper, let the 4 levels of temperature be the treatment and the 5 levels of the recipe be the block. There are no empty cells, so this data set is less complicated than those seen in the analysis of clinical trials. There is highly significant interaction ($p<.0004$), and although there are 4 treatment levels, it is not very difficult to see which levels cause the problem using the plot of the cell means (Figure 2). It is much easier to see the guilty cells in the block chart in Figure 1, and when the cells are unbalanced, this chart is very useful.

Table 1 contains GLM output with the appropriate interaction pairwise contrasts. Table 2 contains a listing of percent contribution and descriptive statistics per cell. Table 3 contains a two way layout with marginal totals for percent contribution. Table 4 contains a summary of block totals, cumulative

totals and the number of times the block was different from the others.

Figure 1 contains a block chart of the per cell contribution to interaction, while Figure 2 contains a plot of the cell means.

5. References

Hwang, D.S., Barry, E.P. and Hamot, H.B., "Treatment by Investigator Interaction: Resolving the Question by Presenting the Evidence". To appear in the ASA Proceedings of the Biopharmaceutical Subsection, 1983

SAS Institute (1982) SAS User's Guide, Cary North Carolina

Snedecor, G.W. and Cochran, W.G. (1980) Statistical Methods, 7th Ed., Iowa State University Press, Ames, Iowa

Johnson, D.E. (1976) "Some New Multiple Comparison Procedures for the Two-Way AOV with Interaction". Biometrics, 32

Li, J.C.R. (1964) Statistical Inference I. Edwards Brothers, Inc., Ann Arbor, Michigan

TABLE 1
SAMPLE RUN
USING JOHNSONS 76 BIOMETRICS EXAMPLE
FROM LI STATISTICAL INFERENCE 64
TWO WAY WITH INTERACTION
AT LEAST 1 PER CELL
VARIABLE=1

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	19	11.58482408	0.60972758	15.23
ERROR	20	0.80054913	0.04002746	PR > F
CORRECTED TOTAL	39	12.38537320		0.0001

R-SQUARE	C.V.	ROOT MSE	Y MEAN
0.935363	4.1558	0.20006863	4.81424952

SOURCE	DF	TYPE I SS	F VALUE	PR > F
TRT	3	7.52952493	62.70	0.0001
BLK	4	1.35641486	8.47	0.0004
TRT*BLK	12	2.69888428	5.62	0.0004

SOURCE	DF	TYPE III SS	F VALUE	PR > F
TRT	3	7.52952493	62.70	0.0001
BLK	4	1.35641486	8.47	0.0004
TRT*BLK	12	2.69888428	5.62	0.0004

CONTRAST	DF	SS	F VALUE	PR > F
INT. BL 1 - BL 2	3	1.56636814	13.04	0.0001
INT. BL 1 - BL 3	3	1.82471820	15.20	0.0001
INT. BL 1 - BL 4	3	1.33846820	11.15	0.0002
INT. BL 1 - BL 5	3	1.51164977	12.59	0.0001
INT. BL 2 - BL 3	3	0.08664999	0.72	0.5508
INT. BL 2 - BL 4	3	0.01514996	0.13	0.9435
INT. BL 2 - BL 5	3	0.13381888	1.11	0.3667
INT. BL 3 - BL 4	3	0.11369992	0.95	0.4368
INT. BL 3 - BL 5	3	0.04881875	0.41	0.7500
INT. BL 4 - BL 5	3	0.10786889	0.90	0.4594

TABLE 2
SAMPLE RUN
USING JOHNSONS 76 BIOMETRICS EXAMPLE
FROM LI STATISTICAL INFERENCE 64
LISTING OF PER CELL
CONTRIBUTION TO INTERACTION

VARIABLE=1 BLOCK=1

TREAT-MENT	CELL SIZE	CELL MEAN	CELL STD	CONT-RIBUTION	% CONT-RIBUTION	% OF PATIENTS
149	2	4.54	0.66	0.9681	35.87	5.00
163	2	4.63	0.00	0.0618	2.29	5.00
190	2	4.52	0.10	0.1070	3.96	5.00
218	2	4.37	0.16	0.8198	30.38	5.00
BLK				1.9567	72.50	20.00

VARIABLE=1 BLOCK=2

TREAT-MENT	CELL SIZE	CELL MEAN	CELL STD	CONT-RIBUTION	% CONT-RIBUTION	% OF PATIENTS
149	2	3.80	0.09	0.1109	4.11	5.00
163	2	4.73	0.25	0.0160	0.59	5.00
190	2	4.92	0.25	0.0006	0.02	5.00
218	2	5.37	0.03	0.0535	1.98	5.00
BLK				0.1810	6.71	20.00

VARIABLE=1 BLOCK=3

TREAT-MENT	CELL SIZE	CELL MEAN	CELL STD	CONT-RIBUTION	% CONT-RIBUTION	% OF PATIENTS
149	2	4.08	0.07	0.1306	4.84	5.00
163	2	4.83	0.04	0.0244	0.90	5.00
190	2	5.43	0.19	0.0780	2.89	5.00
218	2	5.67	0.00	0.0568	2.10	5.00
BLK				0.2898	10.74	20.00

VARIABLE=1 BLOCK=4

TREAT-MENT	CELL SIZE	CELL MEAN	CELL STD	CONT-RIBUTION	% CONT-RIBUTION	% OF PATIENTS
149	2	4.04	0.09	0.0341	1.26	5.00
163	2	4.83	0.06	0.0059	0.22	5.00
190	2	5.01	0.01	0.0091	0.34	5.00
218	2	5.48	0.26	0.0412	1.53	5.00
BLK				0.0903	3.35	20.00

VARIABLE=1 BLOCK=5

TREAT-MENT	CELL SIZE	CELL MEAN	CELL STD	CONT-RIBUTION	% CONT-RIBUTION	% OF PATIENTS
149	2	4.25	0.13	0.0110	0.41	5.00
163	2	4.72	0.11	0.0876	3.24	5.00
190	2	5.35	0.08	0.0282	1.04	5.00
218	2	5.66	0.20	0.0543	2.01	5.00
BLK				0.1811	6.71	20.00
IVAR				2.6989	100.00	100.00
				2.6989	100.00	100.00

TABLE 3
SAMPLE RUN
USING JOHNSONS 76 BIOMETRICS EXAMPLE
FROM LI STATISTICAL INFERENCE 64
CELL PERCENT CONTRIBUTION

	TREATMENT				
	149	163	190	218	TOTAL
	PERCENT	PERCENT	PERCENT	PERCENT	PERCENT
	SUMS OF SQ	SUMS OF SQ	SUMS OF SQ	SUMS OF SQ	SUMS OF SQ
BLOCK					
PLAIN	36	2	4	30	73
3% GMS	4	1	0	2	7
6% GMS	5	1	3	2	11
3% ALDO	1	0	0	2	3
6% ALDO	0	3	1	2	7
TOTAL	46	7	8	38	100

TABLE 4
SAMPLE RUN
USING JOHNSONS 76 BIOMETRICS EXAMPLE
FROM LI STATISTICAL INFERENCE 64
CONTRIBUTION TO INTERACTION

VARIABLE=1

BLOCK	% CONT. TO INT. SS	CUM. % CONT. TO INT. SS	% OF PATIENTS	CUM. % OF PATIENTS	N AND % OF SIG. DIFFERENCES
PLAIN	72.5	72.5	20.0	20.0	4 / 4 = 100.0 %
6% GMS	10.7	83.2	20.0	40.0	1 / 4 = 25.0 %
6% ALDO	6.7	89.9	20.0	60.0	1 / 4 = 25.0 %
3% GMS	6.7	96.7	20.0	80.0	1 / 4 = 25.0 %
3% ALDO	3.3	100.0	20.0	100.0	1 / 4 = 25.0 %

FIGURE 1
SAMPLE RUN
USING JOHNSONS 76 BIOMETRICS EXAMPLE
FROM LI STATISTICAL INFERENCE 64
CELL PERCENT CONTRIBUTION

BLOCK CHART OF PCT_CONT

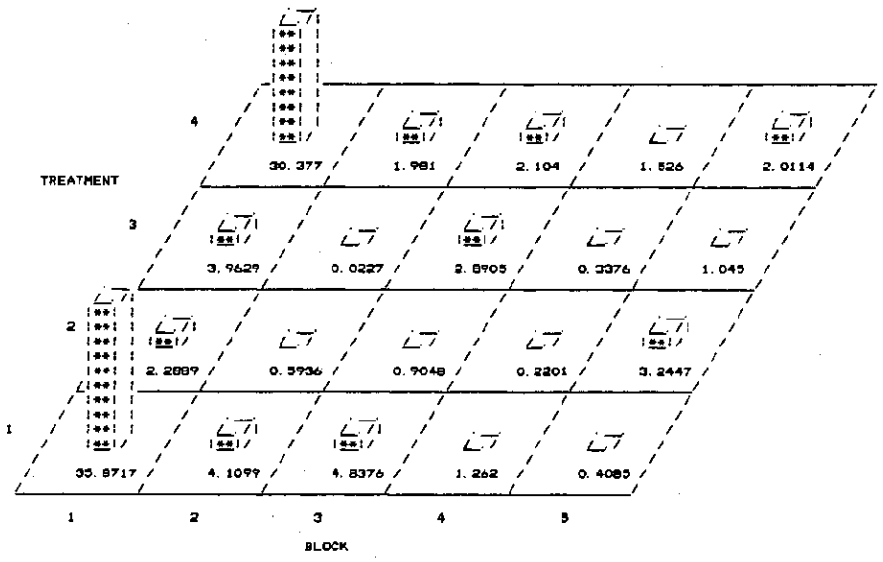
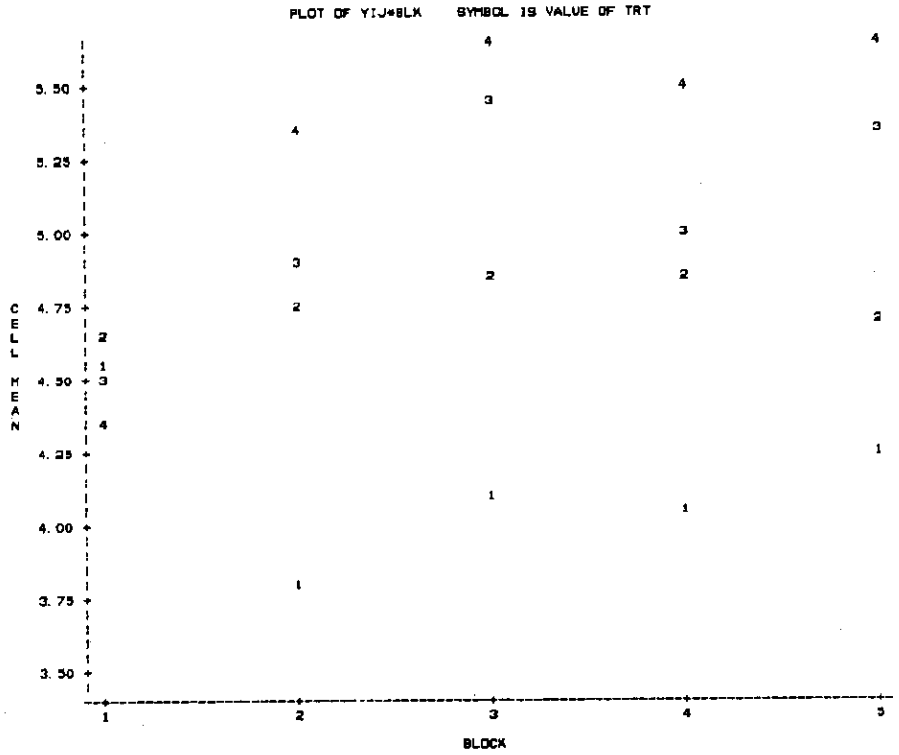


FIGURE 2
SAMPLE RUN
USING JOHNSONS 76 BIOMETRICS EXAMPLE
FROM LI STATISTICAL INFERENCE 64
PLOT OF CELL MEANS



```
// EXEC SAS,SYSOUT='*'
//FT20F001 DD UNIT=DISK,SPACE=(TRK,(5,1))
//SASIN    DD UNIT=DISK,SPACE=(TRK,(5,1))
//SYSIN    DD *

OPTIONS NOSOURCE NODATE NONUMBER LS=110;

 %MACRO CINT(DSNAME=ORIG,      IVAR=1,     VARNAME='VAR 1',
             NTRTS=,           MINCS=1,    PRINTGLM='YES',
             INTPLEV=0.05);

DATA DATA; SET &DSNAME;                    %* select data;
IF IVAR=&IVAR;
PROC FORMAT;                               %* set up varname format;
VALUE VARF  &IVAR = &VARNAME;
PROC SORT; BY IVAR BLK TRT;                %* delete blocks with;
PROC MEANS NOPRINT; BY IVAR BLK TRT;       %* missing/small cells;
VAR Y; OUTPUT OUT=NPC N=NPC;               %* NPC=# per cell;
DATA NPC; SET NPC; BY IVAR;
IF NPC>=&MINCS;
PROC MEANS NOPRINT; BY IVAR BLK;
VAR NPC; OUTPUT OUT=NPB N=NFC;             %* NFC=# filled cells;
DATA NPB; SET NPB; BY IVAR BLK;            %*   per block;
IF NFC<&NTRTS THEN
  PUT 'TOO FEW FOR ' IVAR= BLK=;
IF NFC=&NTRTS; KEEP IVAR BLK;
DATA DATA; MERGE NPB(IN=INNPB) DATA;
BY IVAR BLK; IF INNPB;
KEEP IVAR TRT BLK Y PAT;

DATA BLKS; SET NPB; BY IVAR;               %* make list of blocks;
RETAIN STR; KEEP IVAR NI;                  %*  with commas for DO;
IF FIRST.IVAR THEN DO;                     %*  without for plot;
  STR=PUT(BLK,3.)||REPEAT(' ',4*NI);
  NI=1; END;
ELSE DO; NI+1;
  SUBSTR(STR,NI*3-2,3)=','||PUT(BLK,2.);
  END;
IF LAST.IVAR THEN DO; OUTPUT;              %* get NI on a data set;
  STR=COMPRESS(STR);
  CALL SYMPUT('BLLIST' ,STR);
  STR=TRANSLATE(STR,' ',',');
  CALL SYMPUT('PLLIST' ,STR);
  END;
DATA _NULL_; SET; BY IVAR;                 %* write contrast stmts;
FILE SASIN;                                %*  to file SASIN;
IF &NTRTS>=2 & NI>=2 THEN DO;
 DO I=&BLLIST;
  DO J=&BLLIST;
   IF I < J THEN DO;
    PUT 'CONTRAST ''INT. BL '
        @19 I 2. ' - BL' @27 J 2. '''';
    DO K=2 TO &NTRTS;                      %* K=df (# of lines of;
     PUT 'TRT*BLK' @;                      %*       the contrast);
     DO L=1 TO K; C=1;                     %* L=# rows;
      DO IC=&BLLIST; C+3;
       IF C>69 THEN DO;                    %* new line if >col 69;
        PUT ' '; C=10; END;
       PUT @C ' 0' @;
       IF (L=I-& IC=I)+(L=K-& IC=J) THEN PUT @C ' 1' @;
       IF (L=K & IC=I) | (L=1 & IC=J) THEN PUT @C '-1' @;
       END; %*(IC);
      PUT ' ';
      END; %*(L);
     IF K<&NTRTS THEN PUT ',';
     ELSE PUT ';';
 END; END; END; END; END;

OPTIONS LS=131 PS=499;
PROC PRINTTO UNIT=20 NEW;
PROC GLM DATA=DATA; BY IVAR;
CLASS TRT BLK PAT;
MODEL Y=TRT|BLK/SS1 SS3;
 %INC SASIN;                               %* contrasts;
TITLE TABLE 1; MACTITLE;                   %* users title;
TITLE3 TWO WAY WITH INTERACTION;
TITLE4 AT LEAST  &MINCS  PER CELL;
LABEL IVAR=VARIABLE TRT=TREATMENT;
FORMAT IVAR VARF. TRT TRTF.;
PROC PRINTTO;
OPTIONS LS=132 PS=59;                      %* read contrasts into;
DATA CINTSS; FILE PRINT; TITLE;            %*  CINTSS data set;
INFILE FT20F001 MISSOVER END=END LENGTH=L;
INPUT @1 CC @1.  @; L=L-1;
INPUT @2 LINE @VARYING131. L @;
INPUT @2 FW @CHAR5.  @;
IF FW= 'INT.  ' THEN DO;
  INPUT @10 BLK1 2.  @18 BLK2 2. DF SS F PROB @;
  OUTPUT; END;
INPUT;  DROP CC FW LINE; RETAIN CC LINE;
 %IF &PRINTGLM='YES' %THEN %DO;            %* print GLM output;
  IF CC='1' THEN PUT _PAGE_;               %* carriage control;
  IF CC='0' THEN PUT /;
  PUT @1 LINE @CHAR131.; %END;
OPTIONS PS=499;
PROC SORT DATA=DATA; BY IVAR TRT BLK;
PROC MEANS NOPRINT; BY IVAR TRT BLK; VAR Y;
OUTPUT OUT=DATA N=NIJ MEAN=YIJ STD=SIJ;
PROC PRINTTO UNIT=20 NEW;
PROC GLM; BY IVAR; FREQ NIJ;               %* the second model;
CLASS TRT BLK; MODEL YIJ=TRT BLK;          %* PRINTTO is used to;
OUTPUT OUT=DATA RESIDUAL=DM;               %*  suppress the output;
PROC PRINTTO;

PROC MEANS NOPRINT; BY IVAR; FREQ NIJ;
VAR DM; OUTPUT OUT=SST USS=TS N=TN;
DATA DATA; MERGE DATA SST; BY IVAR;
CS=NIJ*DM**2;
PCT_CONT=ROUND(CS /TS*100,0.0001);         %* round the data;
PCT_PAT =ROUND(NIJ/TN*100,0.01);
SIGDIG=MIN(INT(LOG10(TS)-5),0);
CS=ROUND(CS,10**SIGDIG);

PROC SORT; BY IVAR BLK TRT;                %* print per cell;
OPTIONS PS=59 LS=100;                      %*  contribution;
PROC PRINT SPLIT='/'; BY IVAR BLK;
TITLE TABLE 2; MACTITLE;
TITLE3 LISTING OF PER CELL;
TITLE4 CONTRIBUTION TO INTERACTION;
SUM CS PCT_CONT PCT_PAT; ID TRT;
FORMAT IVAR VARF. TRT TRTF.
       YIJ SIJ PCT_CONT PCT_PAT 6.2;
VAR NIJ YIJ SIJ CS PCT_CONT PCT_PAT;
LABEL IVAR=VARIABLE    BLK=BLOCK
  TRT=TREAT-/ MENT     NIJ=CELL/SIZE
  YIJ=CELL/MEAN        SIJ=CELL/ STD
  CS='  CONT- /RIBUTION'
  PCT_CONT=% CONT-/RIBUTION
  PCT_PAT ='  % OF /PATIENTS';
PROC TABULATE FORMCHAR='              ' F=8.2;
CLASS TRT BLK; VAR PCT_CONT;
LABEL PCT_CONT=PERCENT
   TRT=TREATMENT BLK=BLOCK;
FORMAT TRT TRTF. BLK BLKF.;
TITLE TABLE 3; MACTITLE;
TITLE3 CELL PERCENT CONTRIBUTION;
TABLES (BLK ALL)*F=9,(TRT ALL)*
    (PCT_CONT) / RTS=10;
KEYLABEL SUM='SUMS OF 50' ALL=TOTAL;

PROC CHART;
BLOCK BLK / SUMVAR=PCT_CONT
   DISCRETE TYPE=MEAN GROUP=TRT;
LABEL PCT_CONT=PERCENT
   TRT=TREATMENT BLK=BLOCK;
TITLE FIGURE 1; MACTITLE;
TITLE3 CELL PERCENT CONTRIBUTION;

PROC PLOT;
PLOT YIJ*BLK=TRT / HAXIS=&PLLIST;
LABEL PCT_CONT=PERCENT  YIJ=CELL MEAN
   BLK=BLOCK TRT=TREATMENT;
TITLE FIGURE 2; MACTITLE;
TITLE3 PLOT OF CELL MEANS;

DATA CINTSS; SET CINTSS;                   %* count % sig. diffs.;
SI=(PROB<&INTPLEV);                        %* using INTPLEV;
BLK=BLK1; OUTPUT;
BLK=BLK2; BLK2=BLK1; OUTPUT;
KEEP BLK BLK2 SI;
PROC SORT; BY BLK;
PROC MEANS NOPRINT; BY BLK;
VAR SI; OUTPUT OUT=NSI SUM=NSI;

PROC MEANS NOPRINT DATA=DATA;
BY IVAR BLK; VAR PCT_CONT PCT_PAT;
OUTPUT OUT=RES SUM=PCT_CONT PCT_PAT;
DATA CINTSS; MERGE RES NSI; BY BLK;
PROC SORT; BY DESCENDING PCT_CONT;
DATA CINTSS; MERGE CINTSS BLKS;
BY IVAR; PNSI=NSI/(NI-1)*100;
CNSI=PUT(NSI,2.)||' '||PUT((NI-1),2.)||
    ' = '||PUT(PNSI,5.1)||' %';
CPC_CONT+PCT_CONT; CPC_PAT+PCT_PAT;
DROP NI NSI;

PROC PRINT SPLIT='/' DOUBLE;BY IVAR;%* print contribution;
TITLE TABLE 4; MACTITLE;                   %*   per block;
TITLE3 CONTRIBUTION TO INTERACTION;
ID BLK; FORMAT IVAR VARF. BLK BLKF.
    PCT_CONT CPC_CONT PCT_PAT CPC_PAT 5.1;
VAR PCT_CONT CPC_CONT PCT_PAT CPC_PAT CNSI;
LABEL IVAR=VARIABLE    BLK      =//BLOCK
  CPC_CONT=' CUM. % /CONT. TO/ INT. SS'
  PCT_CONT='    %   /CONT. TO/ INT. SS'
  CPC_PAT =' CUM. % /   OF   /PATIENTS'
  PCT_PAT ='        /  % OF  /PATIENTS'
  CNSI    =' N AND %/ OF SIG./DIFFERENCES';
 %MEND CINT; * -------  end of cint  -------- *;
* Data from Li ('64) Statistical Inference I *;
OPTIONS SOURCE;
DATA ORIG; LENGTH DEFAULT=4;
IVAR=1;  INPUT TRT D1-D5 @@; ARRAY D(BLK) D1-D5; DROP D1-D5;
DO BLK=1 TO 5; Y=D; PAT+1; OUTPUT; END; LIST; CARDS;
4 4.26 5.38 5.67 5.30 5.92 4 4.49 5.39 5.67 5.67 5.80
3 4.59 4.75 5.30 5.00 5.41 3 4.45 5.10 5.57 5.02 5.29
2 4.63 4.56 4.80 4.79 4.45 2 4.63 4.91 4.86 4.98 4.80
1 5.01 3.87 4.13 3.98 4.16 1 4.08 3.74 4.03 4.11 4.35
PROC FORMAT;                               * set up formats;
VALUE TRTF
  1=149 2=163 3=190 4=218;                 * treatment codes;
VALUE TRTPLF
  1='A'    2='B'    3='C'    4='D';         * plot characters;
VALUE BLKF
  1='PLAIN'   2='3% GMB' 3='6% GMB'
  4='3% ALDO' 5='6% ALDO';                 * block = investigator;

 MACRO MACTITLE                            * set up titles;
       TITLE2 SAMPLE RUN;                  * maximum of 3 lines starting at
       TITLE3 USING JOHNSONS 76 BIOMETRICS EXAMPLE;
       TITLE4 FROM LI STATISTICAL INFERENCE 64;
       %
** --  Description of options for the CINT macro  -- **

  Variable  Default          Description
  --------  -------  ---------------------------
  NTRTS       2      # treatments
  IVAR        1      index of the variable to be analyzed
  VARNAME   'VAR 1'  name  of the variable to be analyzed
  MINCS       1      minimum cell size
  PRINTGLM  'YES'      YES/NO print glm output
                       from the 2 way model with interaction
                       including all the interaction contrasts
  INTPLEV   0.05     critical p-level for interaction;

 %CINT(DSNAME=ORIG,     IVAR=1,     VARNAME='VARIABLE 1',
             NTRTS=4,   MINCS=1,    PRINTGLM='YES',
             INTPLEV=0.05);
```