

# SAS による診断・検査データの統計解析：別添資料 データセットおよびサンプルコード

○阿部研自\*・佐伯浩之\*\*

\* 第一製薬株式会社 薬事統計部

\*\* 株式会社第一ラジオアイソトープ研究所 開発業務部

## Statistical Analysis in Diagnostic Medicine via the SAS system

Kenji Abe\* / Hiroyuki Saeki\*\*

\* Regulatory Affairs & Biostatistics Department, Daiichi pharmaceutical Co., Ltd

\*\* Development Administration Department, Daiichi Radioisotope Laboratories, Ltd

本別添資料は、論文中で使用したデータセットおよび SAS サンプルコードの詳細を記述した資料です。内容を「表1. データセットの一覧」、「表 2. SAS サンプルコードの一覧」に示します。本資料に示したコードはあくまでサンプルであり、適用にあたっては必要に応じた加工を想定しております。これらのプログラムに関しては、著作権等を設定しておりませんので、自由な改変・利用を行ってください。しかしながら、このサンプルコードの利用による損害やミスなどに関して著者らは責任を負いかねますので、ご理解の上ご利用ください。

本資料に対するお問い合わせは、下記までお願いいたします。

### 連絡先

阿部 研自： 第一製薬株式会社 薬事統計部  
〒134-8630  
東京都江戸川区北葛西 1-16-13  
e-mail abekei5x@daiichipharm.co.jp

佐伯 浩之： 株式会社第一ラジオアイソトープ研究所 開発業務部  
〒104-0031  
東京都中央区京橋 1-17-10  
e-mail sahiroyuki@drl.co.jp

表 1. データセットの一覧

Daseset name	内容	出典 (Zhou et al.(2002))
data23	2×2 表	Table 2.3
data46	順序カテゴリカルデータ(1 変量)	Table 4.6
data412	連続データ	Table 4.12
data57	順序カテゴリカルデータ(2 変量)	Table 5.7
data4254	クラスターデータ	Table 4.2, Table 5.4

表 2. SAS サンプルコードの一覧

Sample code	内容	主要プロシジャ	Dataset
1	2×2 表における各種統計量とその信頼区間	FREQ	data23
2	Wilson's score 法による信頼区間	FREQ	data23
3	Youden's index とその信頼区間	FREQ	data23
4	Bayes の定理に基づく事後確率の推定と Bayes 曲線の描画	LOGISTIC, GPLOT	data23
5	Empirical ROC の推定	GPLOT, LOGISTIC, NPARIWAY	data412
6	bootstrap 法による AUC の分散推定	LOGISTIC, MULTTEST	data412, data46
7	ROCKIT 用データ変換	—	data57
8	Binormal ROC の推定(直接推定)	—	data412
9	平均損失を最小にする cutoff 値の選択	LOGISTIC, GPLOT	data412
10	Wald, Newcombe-Wilson 法による信頼区間 (対応のあるデータ)	—	data23
11	クラスターデータの解析	MEANS	data4254

## 1. データセット

```
*****
**** Sample code      :   Dataset
****
**** Title            :   データセット準備
****
**** SAS version      :   8.2
****
**** Dataset          :   Xiao-Hua Zhou et al. (2002) Statistical Methods in
****                      Diagnostic Medicine
****
*****;
```

```
TITLE1 "Table2.1 dataset(Zhou et al. (2002) Table 2.3)";
TITLE2 "Mammogram Results of 30 Patients with and 30 Patients without Breast Cancer";
```

```
DATA data23;
  INPUT std var freq;
  DO i=1 TO freq;
    OUTPUT;
  END;
  DROP freq i;
CARDS;
0 0 11
0 1 19
1 0 1
1 1 29
;
RUN;
```

```
TITLE1 "Table2.2 dataset(Zhou et al. (2002) Table 4.6)";
TITLE2 "Results of Workstation-Displayed Digitized-Film Mammogram";
TITLE3 "(BIRAD-System Reader Responses)";
```

```
DATA data46;
  INPUT std var freq;
  DO i=1 TO freq;
    OUTPUT;
  END;
  DROP freq i;
CARDS;
0 1 22
0 2 8
0 3 7
0 4 8
0 5 0
1 1 1
1 2 0
1 3 1
1 4 11
1 5 0
;
RUN;
```

```
TITLE1 "Table 2.2 dataset(Zhou et al. (2002) Table 5.7)";
TITLE2 "Paired Test Results for Mammography";
TITLE3 "(Plain and Digitized Film)";
```

```
DATA data57;
  INPUT std var1 var2 freq;
  DO i=1 TO freq;
    OUTPUT;
  END;
  DROP freq i;
CARDS;
```

0	1	1	17
0	1	2	1
0	1	3	1
0	1	4	2
0	2	1	2
0	2	2	4
0	2	3	1
0	3	1	2
0	3	2	2
0	3	3	4
0	3	4	3
0	4	1	1
0	4	2	1
0	4	3	1
0	4	4	3
1	1	1	1
1	2	4	1
1	3	3	1
1	4	4	10

;

RUN;

TITLE1 "Table2.3 dataset(Zhou et al. (2002) Table 4.12)";

TITLE2 "Data for 60 Patients with Severe Head Trauma";

DATA data412;

INPUT age var std;

ln\_var=log(var);

IF age<=20 THEN group=0;

ELSE group=1;

CARDS;

4	140	1
7	1087	1
8	230	1
11	183	1
15	1256	1
16	700	1
16	16	1
16	800	1
17	253	1
18	740	1
18	126	1
18	153	1
19	283	1
19	90	1
19	303	1
19	193	1
20	76	1
20	1370	1
20	543	1
20	913	1
20	230	1
21	463	1
22	60	1
23	509	1
23	576	1
24	671	1
29	80	1
29	490	1
29	156	1
30	356	1
40	350	1
41	323	1
45	1560	1
45	120	1
50	216	1
51	443	1
56	523	1

59	76	1
61	303	1
61	353	1
62	206	1
6	136	0
6	286	0
7	281	0
8	23	0
8	200	0
10	146	0
11	220	0
12	96	0
12	100	0
16	60	0
17	17	0
18	27	0
18	126	0
19	100	0
24	253	0
28	70	0
35	40	0
38	6	0
46	46	0

```
;
RUN;
```

```
TITLE1 "Table 2.4 dataset(Zhou et al. (2002) Table 4.2 & 5.4)";
TITLE2 "Cluster data Paired Test Results";
TITLE3 "(PET and SPECT)";
```

```
DATA data4254;
  INPUT id mod unit var @@;
CARDS;
```

1	1	1	0
1	1	2	0
1	1	3	0
1	2	1	1
1	2	2	1
1	2	3	1
2	1	1	0
2	1	2	1
2	1	3	1
2	2	1	1
2	2	2	1
2	2	3	1
3	1	1	1
3	1	2	1
3	1	3	1
3	2	1	1
3	2	2	1
3	2	3	1
4	1	1	1
4	2	1	1
5	1	1	0
5	1	2	1
5	1	3	1
5	2	1	1
5	2	2	1
5	2	3	1
6	1	1	1
6	1	2	1
6	1	3	1
6	1	4	1
6	2	1	1
6	2	2	1
6	2	3	1
6	2	4	1
7	1	1	1

7	1	2	1
7	1	3	1
7	2	1	1
7	2	2	1
7	2	3	1
8	1	1	1
8	1	2	1
8	2	1	1
8	2	2	1
9	1	1	1
9	1	2	1
9	2	1	0
9	2	2	1
10	1	1	1
10	2	1	1
11	1	1	0
11	1	2	1
11	1	3	1
11	2	1	0
11	2	2	1
11	2	3	1
12	1	1	1
12	1	2	1
12	2	1	1
12	2	2	1
13	1	1	1
13	1	2	1
13	1	3	1
13	2	1	1
13	2	2	1
13	2	3	1
14	1	1	1
14	1	2	1
14	2	1	1
14	2	2	1
15	1	1	0
15	1	2	0
15	2	1	1
15	2	2	1
16	1	1	0
16	1	2	1
16	1	3	1
16	2	1	0
16	2	2	1
16	2	3	1
17	1	1	0
17	1	2	1
17	1	3	1
17	2	1	0
17	2	2	1
17	2	3	1
18	1	1	0
18	1	2	1
18	1	3	1
18	2	1	1
18	2	2	1
18	2	3	1
19	1	1	1
19	1	2	1
19	2	1	1
19	2	2	1
20	1	1	1
20	2	1	1
21	1	1	1
21	1	2	1
21	2	1	1
21	2	2	1

;

RUN;

TITLE1;  
TITLE2;  
TITLE3;

\*\*\*\*\* PG END \*\*\*\*\*;

## 2. SAS サンプルコード

```

*****
**** Sample code      :   No.1
****
**** Title            :   2×2 表における各種統計量とその信頼区間
****                  (感度, 特異度, 偽陽性, 偽陰性, 陽性尤度比, 陰性尤度比, オッズ比)
****
**** Input    dataset :   データセット
****              std  :   基準変数
****              var  :   検査変数
****              true :   基準変数真
****              fals :   基準変数偽
****              pos  :   検査変数陽性
****              neg  :   検査変数陰性
****
**** SAS version      :   8.2
****
**** Reference        :   Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                  Medicine
****
**** Dataset          :   Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                  Medicine p17 Table2.3
****
*****;

%MACRO EST(dataset, std, var, true, fals, pos, neg);

TITLE1 "SAS Sample Code 1 : Sammary Statistics";

PROC FORMAT;
  VALUE stdc 1="Present" 2="Absent";
  VALUE varc 1="Positive" 2="Negative";
RUN;

DATA caldata;
  SET &dataset;
  IF &std=&true THEN std=1;
  IF &std=&fals THEN std=2;
  IF &var=&pos THEN var=1;
  IF &var=&neg THEN var=2;
RUN;

PROC FREQ DATA=caldata ;
  TABLES std * var / NOCOL NOPERCENT RISKDIFF RELRISK;
  FORMAT std stdc. var varc.;
RUN;

TITLE1;

%MEND EST;

%EST(
dataset = data23, /* データセット */
std      = std,   /* 基準変数 */
var      = var,   /* 検査変数 */
true     = 1,     /* 基準変数真 */
fals     = 0,     /* 基準変数偽 */
pos      = 1,     /* 検査変数陽性 */
neg      = 0      /* 検査変数陰性 */
);

***** PG END *****;

```



```

*****
**** Sample code      :   No.2
****
**** Title           :   Wilson' s score 法による信頼区間
****
**** Input    dataset :   データセット
****              std  :   基準変数
****              var  :   検査変数
****              true :   基準変数真
****              fals :   基準変数偽
****              pos  :   検査変数陽性
****              neg  :   検査変数陰性
****
**** Output          :   Sensitivity, Lower 95%CI_Sensitivity, Upper 95%CI_Sensitivity,
****                      Specificity, Lower 95%CI_Specificity, Upper 95%CI_Specificity
****
**** SAS version     :   8.2
****
**** Reference       :   Xiao-Hua Zhou et al.(2002) Statistical Methods in Diagnostic
****                      Medicine
****
**** Dataset         :   Xiao-Hua Zhou et al.(2002) Statistical Methods in Diagnostic
****                      Medicine p17 Table2.3
****
*****;

%MACRO SCORE(dataset, std, var, true, fals, pos, neg);

TITLE1 "SAS Sample Code 2 : Wilson's Score CI";

DATA caldata;
  SET &dataset;
  IF &std=&true THEN std=1;
  IF &std=&fals THEN std=2;
  IF &var=&pos THEN var=1;
  IF &var=&neg THEN var=2;
RUN;

PROC FREQ DATA=caldata NOPRINT;
  TABLES std * var / OUT=outpct;
RUN;

DATA dmy;
  DO std=1 to 2;
    DO var=1 to 2;
      COUNT=0;
      OUTPUT;
    END;
  END;
RUN;

DATA outpct;
  MERGE dmy outpct;
  BY std var;
RUN;

PROC TRANSPOSE DATA=outpct OUT=outpct_2;
RUN;

DATA outdata (KEEP = sens ci_l_sens ci_u_sens spec ci_l_spec ci_u_spec);
  SET outpct_2;
  WHERE _name_ = 'COUNT';
  sens = col1/(col1+col2);
  spec = col4/(col3+col4);
  ci_l_sens=((sens+PROBIT(0.975)**2/(2*(col1+col2)))
-
(PROBIT(0.975)*SQRT((sens*(1-sens)+PROBIT(0.975)**2/(4*(col1+col2)))/(col1+col2))))
/ (1+PROBIT(0.975)**2/(col1+col2));

```

```

ci_u_sens=((sens+PROBIT(0.975)**2/(2*(col1+col2)))
+
(PROBIT(0.975)*SQRT((sens*(1-sens)+PROBIT(0.975)**2/(4*(col1+col2)))/(col1+col2))))
/ (1+PROBIT(0.975)**2/(col1+col2));
ci_l_spec=((spec+PROBIT(0.975)**2/(2*(col3+col4)))
-
(PROBIT(0.975)*SQRT((spec*(1-spec)+PROBIT(0.975)**2/(4*(col3+col4)))/(col3+col4))))
/ (1+PROBIT(0.975)**2/(col3+col4));
ci_u_spec=((spec+PROBIT(0.975)**2/(2*(col3+col4)))
+
(PROBIT(0.975)*SQRT((spec*(1-spec)+PROBIT(0.975)**2/(4*(col3+col4)))/(col3+col4))))
/ (1+PROBIT(0.975)**2/(col3+col4));
RUN;

PROC PRINT DATA=outdata NOOBS L;
LABEL sens = 'Sensitivity'
      ci_l_sens = 'Lower 95%CI_Sensitivity'
      ci_u_sens = 'Upper 95%CI_Sensitivity'
      spec = 'Specificity'
      ci_l_spec = 'Lower 95%CI_Specificity'
      ci_u_spec = 'Upper 95%CI_Specificity'
      ;
VAR sens ci_l_sens ci_u_sens spec ci_l_spec ci_u_spec;
RUN;

TITLE1;

%MEND;

%SCORE(
dataset = data23, /* データセット */
std      = std,   /* 基準変数 */
var      = var,   /* 検査変数 */
true     = 1,     /* 基準変数真 */
fals     = 0,     /* 基準変数偽 */
pos      = 1,     /* 検査変数陽性 */
neg      = 0      /* 検査変数陰性 */
);

***** PG END *****;

```

```

*****
**** Sample code      :   No.3
****
**** Title           :   Youden' s index とその信頼区間
****
**** Input    dataset :   データセット
****              std  :   基準変数
****              var  :   検査変数
****              true :   基準変数真
****              fals :   基準変数偽
****              pos  :   検査変数陽性
****              neg  :   検査変数陰性
****
**** Output          :   Youden Index, SE, Lower 95%CI, Upper 95%CI
****
**** SAS version     :   8.2
****
**** Reference       :   Youden WJ. Index for rating diagnostic tests. Cancer.
****                   2001:3:32-35
****
**** Dataset         :   Xiao-Hua Zhou et al.(2002) Statistical Methods in Diagnostic
****                   Medicine p17 Table2.3
****
*****;

%MACRO YODEN(dataset, std, var, true, fals, pos, neg);

TITLE1 "SAS Sample Code 3 : Youden's Index";

DATA caldata;
  SET &dataset;
  IF &std=&true THEN std=1;
  IF &std=&fals THEN std=2;
  IF &var=&pos THEN var=1;
  IF &var=&neg THEN var=2;
RUN;

PROC FREQ DATA=caldata NOPRINT;
  TABLES std * var / OUT=outpct;
RUN;

DATA dmy;
  DO std=1 to 2;
    DO var=1 to 2;
      COUNT=0;
      OUTPUT;
    END;
  END;
RUN;

DATA outpct;
  MERGE dmy outpct;
  BY std var;
RUN;

PROC TRANSPOSE DATA=outpct OUT=outpct_2;
RUN;

DATA outdata (KEEP = j v se ci_l ci_u );
  SET outpct_2;
  WHERE _name_ = 'COUNT';
  j = col1/(col1+col2)+col4/(col3+col4)-1;
  v = col1*col2/((col1+col2)**3)+col3*col4/((col3+col4)**3);
  se = SQRT(v);
  ci_l = j-PROBIT(0.975)*se;
  ci_u = j+PROBIT(0.975)*se;
RUN;

```

```

PROC PRINT DATA=outdata NOOBS L;
  LABEL j="Youden's Index"
        se='SE'
        ci_l='Lower 95%CI'
        ci_u='Upper 95%CI'
  ;
  VAR j se ci_l ci_u;
RUN;

TITLE1;

%MEND;

%YOUDEN(
dataset = data23, /* データセット */
std      = std,   /* 基準変数 */
var      = var,   /* 検査変数 */
true     = 1,     /* 基準変数真 */
fals     = 0,     /* 基準変数偽 */
pos      = 1,     /* 検査変数陽性 */
neg      = 0      /* 検査変数陰性 */
);

```

```

***** PG END *****;

```

```

*****
**** Sample code      :   N0.4
****
**** Title            :   Bayes の定理に基づく事後確率の推定と Bayes 曲線の描画
****
**** Input      dataset :   データセット
****                std  :   基準変数
****                var  :   検査変数
****                true  :   基準変数真
****                fals  :   基準変数偽
****                pos   :   検査変数陽性
****                neg   :   検査変数陰性
****
**** Output          :   Sensitivity, Specificity, Pre_Probability, Positive
****                    Post_Probability, Negative Post_Probability
****
**** SAS version     :   8.2
****
**** Reference       :   Xiao-Hua Zhou et al.(2002) Statistical Methods in Diagnostic
****                    Medicine
****
**** Dataset         :   Xiao-Hua Zhou et al.(2002) Statistical Methods in Diagnostic
****                    Medicine p17 Table2.3
****
*****;

%MACRO BAYES(dataset, std, var, true, fals, pos, neg);

TITLE1 "SAS Sample Code 4 : Bayes Curve";

DATA caldata;
  SET &dataset;
  IF &std=&true THEN std=1;
  IF &std=&fals THEN std=2;
  IF &var=&pos THEN var=1;
  IF &var=&neg THEN var=2;
RUN;

ODS LISTING SELECT "Classification Table";
PROC LOGISTIC DATA=caldata;
  MODEL std=var / CTABLE PEVENT=(0.01 TO 0.99 BY 0.01) PPROB=0.5 ;
  ODS OUTPUT Classification=ctable;
RUN;
ODS LISTING;

DATA post;
  SET ctable;
  post_p=100-falsepositive;
  post_n=falsenegative;
  pre=100*probevent;
RUN;

GOPTIONS CBACK=WHITE
          COLORS=(BLACK)
          BORDER;
PROC GPLOT DATA=POST;
  SYMBOL1 I=JOIN V=NONE C=BLACK L=1;
  SYMBOL2 I=JOIN V=NONE C=BLACK L=1;
  PLOT (post_p post_n) * pre / HAXIS=AXIS1 VAXIS=AXIS2 OVERLAY;
  AXIS1 ORDER=(0 TO 100 BY 10) LENGTH=2.5IN LABEL = ('Pretest Probability');
  AXIS2 ORDER=(0 TO 100 BY 10) LENGTH=2.5IN LABEL = ('Posttest Probability');
RUN;
QUIT;

TITLE1;

%MEND BAYES;

```

```
%BAYES (
dataset = data23, /* データセット */
std     = std,    /* 基準変数 */
var     = var,    /* 検査変数 */
true    = 1,      /* 基準変数真 */
fals    = 0,      /* 基準変数偽 */
pos     = 1,      /* 検査変数陽性 */
neg     = 0,      /* 検査変数陰性 */
);
```

```
***** PG END *****;
```

```

*****
**** Sample code      :   N0.5
****
**** Title            :   Empirical ROC の推定
****                  :   (ROC の描画, AUC の推定 C 統計量と Wilcoxon-Mann-Whitney 統計量)
****
**** Input      dataset :   データセット
****              std   :   基準変数
****              var   :   検査変数
****              true  :   基準変数真
****              fals  :   基準変数偽
****
**** Output          :   LOGISTIC : デフォルト
****                  :   Mann-Whitney : Class, N, Sum of Score, U, AUC
****
**** SAS version      :   8.2
****
**** Reference        :   Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                  :   Medicine
****
**** Dataset          :   Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                  :   Medicine
****                  :   p183 Table4.12
****
*****;

%MACRO ROC_empirical(dataset, std, var, true, fals);

TITLE1 "SAS Sample Code 5 : ROC Analysis (Empirical)";

DATA caldata;
  SET &dataset;
  IF &std=&>true THEN std=1;
  IF &std=&fals THEN std=2;
  var=&var;
RUN;

ODS LISTING SELECT "Association Statistics";
PROC LOGISTIC DATA=caldata;
  MODEL std=var/ OUTROC=roc ROCEPS=0;
RUN;
ODS LISTING;

ODS LISTING CLOSE;
PROC NPAR1WAY DATA=caldata WILCOXON;
  CLASS std;
  VAR var;
  ODS OUTPUT WILCOXONSCORES=wilcoxonscores;
RUN;

DATA mannwhitney;
  SET wilcoxonscores;
  u=sumofscores-n*(n+1)/2;
  m=lag(n);
  u_=lag(u);
  u=max(u, u_);
  mn=m*n;
  auc=u/mn;
  KEEP class n sumofscores u auc;
RUN;

ODS LISTING;
PROC PRINT DATA=mannwhitney NOOBS L;
  LABEL class = 'Class'
         n = 'N'
         sumofscores = 'Sum of Scores'
         u = 'U'
         auc = 'AUC'

```

```

;
VAR class n sumofscores u auc;
RUN;

GOPTIONS CBACK=WHITE
         COLORS=(BLACK)
         BORDER;

PROC Gplot DATA=roc;
  SYMBOL1 I=JOIN V=NONE ;
  PLOT _sensit_ *_1mspec_ / HAXIS=AXIS1 VAXIS=AXIS2;
  AXIS1 LENGTH=2.5IN;
  AXIS2 ORDER=(0 TO 1 BY 0.2) LENGTH=2.5IN;
RUN;

QUIT;

TITLE1;

%MEND ROC_empirical;

%ROC_empirical(
dataset = data412, /* データセット */
std      = std,    /* 基準変数 */
var      = var,    /* 検査変数 */
true     = 1,      /* 基準変数真 */
fals     = 0       /* 基準変数偽 */
);

***** PG END *****;

```



```

*****
**** Sample code      : No. 6
****
**** Title            : Empirical ROC の推定
****                   (bootstrap 法による AUC の分散推定)
****
**** Input    dataset : データセット
****              std  : 基準変数
****              var  : 検査変数
****              true : 基準変数真
****              fals : 基準変数偽
****              nsam : シミュレーション回数
****              seed : シミュレーションシード
****
**** SAS version      : 8.2
****
**** Reference        : 古川敏仁, 牧野建一. ROC 分析による検査診断能の比較 -AUC の
****                   bootstrap, 最尤推定を中心に-.
****                   第 15 回日本 SAS ユーザー会総会および研究発表会論文集.
****                   1996: 355-366.
****
**** Dataset          : Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                   Medicine
****                   p111 Table 4.6, p138 Table 4.12
****
*****;

%MACRO ROCAUC_bootstrap(dataset, std, var, true, fals, nsam, seed);

TITLE1 "SAS Sample Code 6 : ROC Analysis (Bootstrap Estimator for Variance)";

DATA caldata;
  SET &dataset;
  IF &std=&true THEN std=1;
  IF &std=&fals THEN std=2;
RUN;

PROC SORT DATA=caldata;
  BY std;
RUN;

ODS LISTING SELECT "Association Statistics";
PROC LOGISTIC DATA=caldata;
  MODEL std=var;
RUN;

PROC MULTTEST DATA=caldata NSAMPLE=&nsam OUTSAMP=out SEED=&seed NOCENTER NOPRINT BOOTSTRAP;
  BY std;
  TEST MEAN(var);
  CLASS std;
RUN;

PROC SORT DATA=out;
  BY _sample_ std;
RUN;

ODS LISTING CLOSE;
ODS OUTPUT "Association Statistics"=aucout;
PROC LOGISTIC DATA=out;
  MODEL std=var /ROCEPS=0 OUTROC=roc;
  BY _sample_;
RUN;

ODS LISTING;

PROC MEANS DATA=aucout(WHERE=(label2="c")) n mean std median Q1 Q3 min max;
  VAR nvalue2;
RUN;

```

```

TITLE1;

%MEND ROCAUC_bootstrap;

%ROCAUC_bootstrap(
dataset = data412, /* データセット */
std     = std,     /* 基準変数 */
var     = var,     /* 検査変数 */
true    = 1,       /* 基準変数真 */
fals    = 0,       /* 基準変数偽 */
nsam    = 1000,    /* シミュレーション回数 */
seed    = 5963     /* シミュレーションシード */
);

%ROCAUC_bootstrap(
dataset = data46, /* データセット */
std     = std,     /* 基準変数 */
var     = var,     /* 検査変数 */
true    = 1,       /* 基準変数真 */
fals    = 0,       /* 基準変数偽 */
nsam    = 1000,    /* シミュレーション回数 */
seed    = 5963     /* シミュレーションシード */
);

***** PG END *****;

```

```

*****
**** Sample code       :   No.7
****
**** Title             :   ROCKIT 用データ変換
****
**** Input    dataset  :   データセット
****                std   :   基準変数
****                var1  :   検査変数 1
****                var2  :   検査変数 2
****                var3  :   検査変数 3
****                var4  :   検査変数 4
****                var5  :   検査変数 5
****                true  :   基準変数真
****                fals  :   基準変数偽
****                nam1  :   検査名ラベル 1
****                nam2  :   検査名ラベル 2
****                nam3  :   検査名ラベル 3
****                nam4  :   検査名ラベル 4
****                nam5  :   検査名ラベル 5
****                typ1  :   属性 1
****                typ2  :   属性 2
****                typ3  :   属性 3
****                typ4  :   属性 4
****                typ5  :   属性 5
****                file  :   ROCKIT 用データセット出力ファイル名
****
**** 「検査名ラベルは任意の文字をダブルコーテーション(“)で括って指定」
**** 「属性は 3 文字で表記」
**** 1 文字目⇒データの尺度 連続データ : C, カテゴリカルデータ:D
**** 2 文字目⇒基準変数の真値の方向 連続データ : 大 L, 小 S
****                                     カテゴリカルデータ:小 1, 大 最大値
**** 3 文字目⇒L (ライン入力形式)
**** 例 : 連続データ, 基準変数の真値の方向が大の場合 ⇒CLL
****       5 区分のカテゴリカルデータ, 基準変数の真値の方向が大の場合⇒D5L
****
**** SAS version       :   8.2
****
**** Reference         :   Metz C.E. Basic principles of ROC analysis. Semin. Nucl. Med.
****                       1978: 8: 283-298.
****                       Metz C.E., Herman B.A., and Shen J. Maximum likelihood estimation
****                       of receiver operating characteristic (ROC) curves from
****                       continuously distributed data. Stat. Med. 1998: 17: 1033-1053.
****                       URL http://xray.bsd.uchicago.edu/krl/KRL\_ROC/software\_index.htm
****
**** Dataset           :   Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                       Medicine p175 Table5.7
****
*****

```

```

DATA header;
  LENGTH c1 $15. c2 $15. c3 $15. c4 $15. c5 $15.;
  INPUT  c1 $ c2 $ c3 $ c4 $ c5 $ @@;
  CARDS;
  ROCKIT_DATA . . . .
  KIT . . . .
  name name name name name
  type type type type type
;

```

RUN;

```

%MACRO RKDATA(dataset,std, var1, var2, var3, var4, var5, true, fals, nam1, nam2,
  nam3, nam4, nam5, typ1, typ2, typ3, typ4, typ5, file);

```

```

DATA data;
  SET &dataset END=EOF ;
  %IF &var1^= %THEN c1=COMPRESS(PUT (&var1,5.));;
  %IF &var2^= %THEN c2=COMPRESS(PUT (&var2,5.));;
  %IF &var3^= %THEN c3=COMPRESS(PUT (&var3,5.));;

```

```

%IF &var4^= %THEN c4=COMPRESS(PUT (&var4, 5.));;
%IF &var5^= %THEN c5=COMPRESS(PUT (&var5, 5.));;
IF c1="." THEN c1="#";
IF c2="." THEN c2="#";
IF c3="." THEN c3="#";
IF c4="." THEN c4="#";
IF c5="." THEN c5="#";
OUTPUT;
IF EOF THEN DO;
  c1='*'; c2=''; c3=''; c4=''; c5='';
  DO &std=&fals, &>true;
    OUTPUT;
  END;
END;
%IF &var1^= %THEN KEEP c1;;
%IF &var2^= %THEN KEEP c2;;
%IF &var3^= %THEN KEEP c3;;
%IF &var4^= %THEN KEEP c4;;
%IF &var5^= %THEN KEEP c5;;
KEEP &std;
RUN;

/* データセットの統合 */
DATA rkdata;
  SET header data(WHERE=(&std=&fals)) data(WHERE=(&std=&>true));
  IF c1='name' THEN c1=&nam1;
  IF c1='type' THEN c1=&typ1;
  IF c2='name' THEN c2=&nam2;
  IF c2='type' THEN c2=&typ2;
  IF c3='name' THEN c3=&nam3;
  IF c3='type' THEN c3=&typ3;
  IF c4='name' THEN c4=&nam4;
  IF c4='type' THEN c4=&typ4;
  IF c5='name' THEN c5=&nam5;
  IF c5='type' THEN c5=&typ5;
  KEEP c1;
  %IF &var2^= %THEN KEEP c2;;
  %IF &var3^= %THEN KEEP c3;;
  %IF &var4^= %THEN KEEP c4;;
  %IF &var5^= %THEN KEEP c5;;
  DROP &std;
RUN;

DATA _NULL_;
  SET rkdata;
  FILE &FILE;
  %IF &var1^= AND &var2 = AND &var3 = AND &var4 = AND &var5 = %THEN PUT c1;;
  %IF &var1^= AND &var2^= AND &var3 = AND &var4 = AND &var5 = %THEN PUT c1 c2;;
  %IF &var1^= AND &var2^= AND &var3^= AND &var4 = AND &var5 = %THEN PUT c1 c2 c3;;
  %IF &var1^= AND &var2^= AND &var3^= AND &var4^= AND &var5 = %THEN PUT c1 c2 c3 c4;;
  %IF &var1^= AND &var2^= AND &var3^= AND &var4^= AND &var5^= %THEN PUT c1 c2 c3 c4 c5;;
RUN;

%MEND RKDATA;

%RKDATA(
dataset =data57,          /* データセット */
std      =std,            /* 基準変数 */
var1     =var1,          /* 検査変数 1 */
var2     =var2,          /* 検査変数 2 */
var3     = ,              /* 検査変数 3 */
var4     = ,              /* 検査変数 4 */
var5     = ,              /* 検査変数 5 */
true     =1,              /* 基準変数陽性 */
fals     =0,              /* 基準変数陰性 */
nam1     ='Plain',        /* 検査名ラベル 1 */
nam2     ='Digitized',    /* 検査名ラベル 2 */
nam3     =' ',            /* 検査名ラベル 3 */

```

```

nam4    =' ',          /* 検査名ラベル 4 */
nam5    =' ',          /* 検査名ラベル 5 */
typ1    =' D5L',       /* 属性 1 */
typ2    =' D5L',       /* 属性 2 */
typ3    =' ',          /* 属性 3 */
typ4    =' ',          /* 属性 4 */
typ5    =' ',          /* 属性 5 */
file    =' 任意のパスを設定¥output.dat' /*データセット出力ファイル名*/
);

```

```

***** PG END *****;

```

```

*****
**** Sample code      : No.8
****
**** Title           : Binormal ROC の推定 (直接推定)
****                  (ROC の描画, AUC の推定)
****
**** Input    dataset : データセット
****                  std  : 基準変数
****                  var1 : 検査変数 1
****                  var2 : 検査変数 2
****                  true  : 基準変数真
****                  fals  : 基準変数偽
****
**** Output          : a, b, AUC, Std Error, Lower 95% CI, Upper 95% CI, AUC1, AUC2,
****                  Difference Value, Difference Variance, Difference Std Error,
****                  Lower 95% CI, Upper 95% CI, Z value, P value
****
**** SAS version     : 8.2
****
**** Reference       : Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                  Medicine
****
**** Dataset         : Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                  Medicine
****                  p138 Table 4.12, p175 Table5.7
****
*****

```

```
%MACRO ROCAUC(var);
```

```
DATA auc_&var;
```

```
  MERGE simplestats(WHERE=(std_=1 and variable="&var") RENAME=(mean=mu1 stddev=sd1 nobs=n1))
    simplestats(WHERE=(std_=2 and variable="&var") RENAME=(mean=mu2 stddev=sd2
nobs=n2));
```

```
  IF mu1>mu2 THEN DO;
```

```
    a=(mu1-mu2)/sd1;
```

```
    b=sd2/sd1;
```

```
    auc=PROBNORM(a/SQRT(1+b**2));
```

```
    e=EXP(-(a**2)/(2*(1+b**2)));
```

```
    rard=e/SQRT(2*3.14159265358979*(1+b**2)*sd1**2);
```

```
    rarsx=-a*b*e/(2*sd1*sd2*SQRT(2*3.14159265358979*(1+b**2)**3));
```

```
    rarsy=-a/(2*sd1)*rard-(b**2)*rarsx;
```

```
    vd=(sd2**2)/n2+(sd1**2)/n1;
```

```
    vsx=2*(sd2**4)/(n2-1);
```

```
    vsy=2*(sd1**4)/(n1-1);
```

```
    var=(rard**2)*vd+(rarsx**2)*vsx+(rarsy**2)*vsy;
```

```
  END;
```

```
ELSE DO;
```

```
  a=(mu2-mu1)/sd2;
```

```
  b=sd1/sd2;
```

```
  auc=PROBNORM(a/SQRT(1+b**2));
```

```
  e=EXP(-(a**2)/(2*(1+b**2)));
```

```
  rard=e/SQRT(2*3.14159265358979*(1+b**2)*sd2**2);
```

```
  rarsx=-a*b*e/(2*sd1*sd2*SQRT(2*3.14159265358979*(1+b**2)**3));
```

```
  rarsy=-a/(2*sd2)*rard-(b**2)*rarsx;
```

```
  vd=(sd1**2)/n1+(sd2**2)/n2;
```

```
  vsx=2*(sd1**4)/(n1-1);
```

```
  vsy=2*(sd2**4)/(n2-1);
```

```
  var=(rard**2)*vd+(rarsx**2)*vsx+(rarsy**2)*vsy;
```

```
END;
```

```
se=SQRT(var);
```

```
psi=LOG((1+auc)/(1-auc));
```

```
varpsi=4*var/((1-auc**2)**2);
```

```
lowerpsi=psi-PROBIT(0.975)*SQRT(varpsi);
```

```
upperpsi=psi+PROBIT(0.975)*SQRT(varpsi);
```

```

lowercl=(1-EXP(-lowerpsi))/(1+EXP(-lowerpsi));
uppercl=(1-EXP(-upperpsi))/(1+EXP(-upperpsi));
KEEP variable mu1 mu2 sd1 sd2 n1 n2 a b auc var se lowercl uppercl rard rarsx rarsy;

RUN;

%MEND ROCAUC;

%MACRO ROC_binormal (dataset, std, var1, var2, true, fals);

DATA caldata;
  SET &dataset;
  IF &std=&true THEN std_=1;
  IF &std=&fals THEN std_=2;
RUN;

PROC SORT DATA=caldata OUT=caldata2;
  BY std_;
RUN;

ODS LISTING CLOSE;
PROC CORR DATA=caldata2;
  VAR &var1 &var2;
  BY std_;
  ODS OUTPUT SIMPLESTATS=simplestats PEARSONCORR=pearsoncorr;
RUN;
ODS LISTING;

%IF &var2^= %THEN %DO;
TITLE1 "SAS Sample Code 8 : ROC Analysis (Binormal)";
TITLE2 "Binormal Parameters & AUC";
%ROCAUC(&var1);
%ROCAUC(&var2);

DATA out1;
  SET auc_&var1 auc_&var2;
RUN;

PROC PRINT DATA=out1 NOOBS L;
  VAR variable a b auc se lowercl uppercl;
  LABEL auc="AUC"
         se="Std Error"
         lowercl="Lower 95% CI"
         uppercl="Upper 95% CI"
         ;
RUN;

TITLE1 "SAS Sample Code 8 : ROC Analysis (Binormal)";
TITLE2 "Test & CI for AUC";
DATA out2;
  MERGE auc_&var1 (RENAME=(auc=auc1 var=variance1 sd1=sd11 sd2=sd12 n1=n11 n2=n12 rard=rard1
rarsx=rarsx1 rarsy=rarsy1))
        auc_&var2 (RENAME=(auc=auc2 var=variance2 sd1=sd21 sd2=sd22 n1=n21 n2=n22 rard=rard2
rarsx=rarsx2 rarsy=rarsy2))
        pearsoncorr (WHERE=(std_=1 and variable="&var2") RENAME=(&var1=r1))
        pearsoncorr (WHERE=(std_=2 and variable="&var2") RENAME=(&var1=r2));
  aucdiff=auc1-auc2;
  IF mu1>mu2 THEN DO;
    covdiff=r1*sd11*sd21/n11+r2*sd12*sd22/n12;
    covsx=2*r1*sd11**2*sd21**2/(n11-1);
    covsy=2*r2*sd12**2*sd22**2/(n12-1);
  END;
  ELSE DO;
    covdiff=r2*sd12*sd22/n12+r1*sd11*sd21/n11;
    covsx=2*r2*sd12**2*sd22**2/(n12-1);
    covsy=2*r1*sd11**2*sd21**2/(n11-1);
  END;
  covauc=rard1*rard2*covdiff+rarsx1*rarsx2*covsx+rarsy1*rarsy2*covsy;

```

```

var=variance1+variance2-2*covauc;
se=SQRT(var);

lowercl=aucdiff-PROBIT(0.975)*SQRT(var);
uppercl=aucdiff+PROBIT(0.975)*SQRT(var);

sita1=LOG((1+auc1)/(1-auc1));
varsita1=(2/(1-auc1**2))*2*variance1;
sita2=LOG((1+auc2)/(1-auc2));
varsita2=(2/(1-auc2**2))*2*variance2;
covsita=4*covauc/((1-auc1**2)*(1-auc2**2));

z=(sita1-sita2)/SQRT(varsita1+varsita2-2*covsita);
p=2*(1-PROBNORM(ABS(z)));
KEEP auc1 auc2 aucdiff variance1 variance2 covauc var se
      lowercl uppercl z p;
RUN;

PROC PRINT DATA=out2 NOOBS L;
VAR auc1 auc2 aucdiff
    var se lowercl uppercl z p;
LABEL auc1="AUC1"
      auc2="AUC2"
      aucdiff="Difference Value"
      var="Difference Variance"
      se="Difference Std Error"
      lowercl="Lower 95% CI"
      uppercl="Upper 95% CI"
      z="Z value"
      p="P value"
      ;
RUN;

TITLE2 "ROC curve for &var1 & &var2";
DATA plotdata;
merge auc_&var1 (RENAME=(a=&var1._a b=&var1._b))
      auc_&var2 (RENAME=(a=&var2._a b=&var2._b));
DO _1MSPEC_=0 TO 1 BY 0.01;
    IF FUZZ(_1mspec_)=0 THEN _sensit1_=0;
    ELSE IF FUZZ(_1mspec_)=1 THEN _sensit1_=1;
    ELSE _sensit1_=1-PROBNORM(&var1._b*PROBIT(1-_1mspec_)-&var1._a);
    IF FUZZ(_1mspec_)=0 THEN _sensit2_=0;
    ELSE IF FUZZ(_1mspec_)=1 THEN _sensit2_=1;
    ELSE _sensit2_=1-PROBNORM(&var2._b*PROBIT(1-_1mspec_)-&var2._a);
    OUTPUT;
END;
RUN;
GOPTIONS CBACK=WHITE
          COLORS=(BLACK)
          BORDER;

PROC GPLOT DATA=plotdata;
SYMBOL1 I=JOIN V=NONE c = BLACK L=1 W=1;
SYMBOL2 I=JOIN V=NONE c = RED L=1 W=1;
PLOT (_sensit1_ _sensit2_) * _1mspec_ / HAXIS=AXIS1 VAXIS=AXIS2 OVERLAY;
AXIS1 ORDER=(0 TO 1.0 BY 0.2) LENGTH=2.5in label = ('1-Specificity');
AXIS2 ORDER=(0 TO 1.0 BY 0.2) LENGTH=2.5in label = ('Sensitivity');
RUN;
QUIT;

%END;

%IF &var2= %THEN %DO;
TITLE1 "SAS Sample Code 8 : ROC Analysis (Binormal)";
TITLE2 "Binormal Parameters & AUC";
%ROCAUC(&var1);
PROC PRINT DATA=auc_&var1 NOOBS L;
VAR variable a b auc se lowercl uppercl;

```



```

    LABEL auc="AUC"
           se="Std Error"
           lowercl="Lower 95% CI"
           uppercl="Upper 95% CI"
           ;
RUN;
TITLE2 "ROC curve for &var1";
DATA plotdata;
    SET auc_&var1(keep=a b);
    DO _1MSPEC_=0 TO 1 BY 0.01;
        IF FUZZ(_1mspec_)=0 THEN _sensit_=0;
        ELSE IF FUZZ(_1mspec_)=1 THEN _sensit_=1;
        ELSE _sensit_=1-PROBNORM(b*PROBIT(1-_1mspec_)-a);
        OUTPUT;
    END;
RUN;
GOPTIONS CBACK=WHITE
          COLORS=(BLACK)
          BORDER;

PROC GPLOT DATA=plotdata;
    SYMBOL1 I=JOIN V=NONE ;
    PLOT _sensit_ *_1mspec_ / HAXIS=AXIS1 VAXIS=AXIS2;
    AXIS1 ORDER=(0 TO 1.0 BY 0.2) LENGTH=2.5in label = ('1-Specificity');
    AXIS2 ORDER=(0 TO 1.0 BY 0.2) LENGTH=2.5in label = ('Sensitivity');
RUN;

QUIT;

%END;

TITLE1;
TITLE2;

%MEND ROC_binormal;

%ROC_binormal(
dataset = data57, /* データセット */
std      = std,   /* 基準変数 */
var1     = var1,  /* 検査変数 1 */
var2     = var2,  /* 検査変数 2 */
true     = 1,     /* 基準変数真 */
fals     = 0      /* 基準変数偽 */
);

%ROC_binormal(
dataset = data412, /* データセット */
std      = std,   /* 基準変数 */
var1     = var,   /* 検査変数 1 */
var2     = ,      /* 検査変数 2 */
true     = 1,     /* 基準変数真 */
fals     = 0      /* 基準変数偽 */
);

*****PG END *****;

```

```

*****
**** Sample code      :   No.9
****
**** Title           :   平均損失を最小にする cutoff 値の選択
****
**** Input    dataset :   データセット
****          std      :   基準変数
****          var      :   検査変数
****          true     :   基準変数真
****          fals     :   基準変数偽
****          prevalence :   有病率
****          c_fp     :   偽陽性のコスト
****          c_tn     :   真陰性のベネフィット
****          c_fn     :   偽陰性のコスト
****          c_tp     :   真陰性のベネフィット
****
**** Output          :   Cutoff Value , Sensitivity , Specificity , Se-m(1-Sp)
****
**** SAS version     :   8.2
****
**** Reference       :   Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                     Medicine
****
**** Dataset         :   Xiao-Hua Zhou et al. (2002) Statistical Methods in Diagnostic
****                     Medicine
****                     p138 Table 4.12
****
*****

```

```
%MACRO OPTIMAL (dataset, std, var, true, fals, prevalence, c_fp, c_tn, c_fn, c_tp);
```

```
TITLE1 "SAS Sample Code 9 : Optimal Cutoff Value on ROC curve";
```

```
TITLE2 "Prevalence=&Prevalence";
```

```
TITLE3 "C_FP=&c_fp , C_TN=&c_tn , C_FN=&c_fn , C_TP=&c_tp";
```

```
DATA caldata;
```

```
SET &dataset;
```

```
IF &std=&>true THEN std=1;
```

```
IF &std=&fals THEN std=2;
```

```
var=&var;
```

```
RUN;
```

```
ODS LISTING CLOSE;
```

```
PROC LOGISTIC DATA=caldata;
```

```
MODEL std=var/ OUTROC=roc ROCEPS=0;
```

```
OUTPUT OUT=out P=_prob_;
```

```
RUN;
```

```
DATA temp;
```

```
SET roc;
```

```
_prob_=ROUND(_prob_, .000000001);
```

```
RUN;
```

```
DATA out;
```

```
SET out;
```

```
_prob_=ROUND(_prob_, .000000001);
```

```
RUN;
```

```
PROC SORT DATA=temp NODUPKEY;
```

```
BY _prob_;
```

```
RUN;
```

```
PROC SORT DATA=out NODUPKEY;
```

```
BY _prob_;
```

```
RUN;
```

```
DATA prob;
```

```

MERGE out temp;
BY _prob_;
_spec = 1 - 1mspec_;
m = (1 - &prevalence) / (&prevalence * ((&c_fp - &c_tn) / (&c_fn - &c_tp));
cost = _sensit_ - m * 1mspec_;
KEEP var _sensit_ _spec_ cost;
RUN;

ODS LISTING;

PROC PRINT DATA=prob NOOBS L;
  LABEL var="Cutoff Value"
         _sensit_="Sensitivity"
         _spec_="Specificity"
         cost="Se-m(1-Sp)"
         ;
RUN;

GOPTIONS CBACK=WHITE
         COLORS=(BLACK)
         BORDER;

PROC GLOT DATA=prob;
  SYMBOL1 I=JOIN V=NONE C=BLACK;
  PLOT cost * var / HAXIS=AXIS1 VAXIS=AXIS2;
  AXIS1 LENGTH=4IN LABEL=('Cutoff Value');
  AXIS2 LENGTH=4IN LABEL=('Se-m(1-Sp)');
RUN;
QUIT;

TITLE1;
TITLE2;
TITLE3;

%MEND OPTIMAL;

%OPTIMAL(
dataset      = data412 , /*データセット*/
std           = std     , /*基準変数*/
var           = var     , /*検査値*/
true          = 1       , /*基準変数真*/
fals          = 0       , /*基準変数偽*/
prevalence    = 0.3     , /*有病率*/
c_fp          = 0.7     , /*偽陽性のコスト*/
c_tn          = 0.9     , /*真陰性のベネフィット*/
c_fn          = 0.0     , /*偽陰性のコスト*/
c_tp          = 1.0     , /*真陰性のベネフィット*/
);

***** PG END *****;

```

```

*****
**** Sample code      : No.10
****
**** Title           : 対応のあるデータの解析 (2×2 分割表)
****                  (McNemar 検定,
****                  Wald 法, Newcombe-Wilson score 法による信頼区間)
****
**** Input    dataset : データセット
****              var1 : 検査変数 1 (基準変数と一致, 不一致が入力されている変数)
****              var2 : 検査変数 2 (基準変数と一致, 不一致が入力されている変数)
****              pos  : 検査変数一致
****              neg  : 検査変数不一致
****
**** Output         : FREQ : デフォルト
****                  その他 : Point Estimate of Difference, Lower 95% CI (Wald),
****                  Upper 95% CI (Wald), Lower 95% CI (Newcombe-Wilson GC),
****                  Upper 95% CI (Newcombe-Wilson GC)
****
**** SAS version    : 8.2
****
**** Reference      : Newcombe, R. G. Two-sided confidence intervals for the single
****                  proportion:
****                  Comparison of seven methods. Stat. Med, 1998: 17: 873-890.
****                  Newcombe, R. G. Improved confidence intervals for the difference
****                  between binormal proportion based of paired data. Stat. Med,
****                  1998: 17: 2635-2650.
****
**** Dataset        : Xiao-Hua Zhou et al.(2002) Statistical Methods in Diagnostic
****                  Medicine p17 Table2.3
****
*****

```

```
%MACRO MCNEM(dataset, var1, var2, pos, neg);
```

```

TITLE1 "SAS Sample Code 10 : Comparison of Paired Data";
TITLE2 "Contingency Table & McNemar Test";

```

```

DATA caldata;
  SET &dataset;
  IF &var1=&pos THEN var1=1;
  IF &var1=&neg THEN var1=2;
  IF &var2=&pos THEN var2=1;
  IF &var2=&neg THEN var2=2;
RUN;

```

```

PROC FORMAT;
  VALUE pnc 1="Positive" 2="Negative";
RUN;

```

```

ODS LISTING SELECT "クロス集計表" "MCNEMAR の検定";
PROC FREQ DATA=caldata;
  TABLES var1 * var2 / NOCOL NOROW OUT=outpct;
  EXACT MCNEM;
  FORMAT var1 var2 pnc.;
RUN;
ODS LISTING;

```

```

DATA dmy;
  DO var1 = 1 TO 2;
    DO var2 = 1 TO 2;
      COUNT = 0;
      OUTPUT;
    END;
  END;
RUN;

```

```

DATA outpct;
  MERGE dmy outpct;

```

```

    BY var1 var2;
    RUN;
PROC TRANSPOSE DATA=output OUT=output2;
RUN;

/*****

TITLE2 "Wald CI";

DATA waldci;
    SET output2(WHERE=(name="COUNT") RENAME=(col1=a col2=b col3=c col4=d));
    diff=(b-c)/SUM(a,b,c,d);
    var=((a+d)*(b+c)+4*b*c)/SUM(a,b,c,d)**3;
    lower=diff-PROBIT(0.975)*SQRT(var);
    upper=diff+PROBIT(0.975)*SQRT(var);
RUN;

PROC PRINT DATA=waldci NOOBS L;
    VAR diff lower upper;
    LABEL diff="Point Estimate of Difference"
           lower="Lower 95% CI (Wald)"
           upper="Upper 95% CI (Wald)"
           ;
RUN;

*****/

TITLE2 "Newcombe-Wilson Score CI (Continuity Correction)";

DATA wilsonci;
    SET output2(WHERE=(name="COUNT") RENAME=(col1=a col2=b col3=c col4=d));

    diff=(b-c)/SUM(a,b,c,d);

    i22=2*(a+b)+PROBIT(0.975)**2;
    i23=PROBIT(0.975)*SQRT(PROBIT(0.975)**2 + 4*(a+b)*(c+d)/SUM(a,b,c,d));
    i24=2*(SUM(a,b,c,d)+PROBIT(0.975)**2);
    i25=(i22-i23)/i24;
    i26=(i22+i23)/i24;
    i27=(a+b)/SUM(a,b,c,d)-i25;
    i28=i26-(a+b)/SUM(a,b,c,d);
    i29=2*(a+c)+PROBIT(0.975)**2;
    i30=PROBIT(0.975)*SQRT(PROBIT(0.975)**2+4*(a+c)*(b+d)/SUM(a,b,c,d));
    i31=(i29-i30)/i24;
    i32=(i29+i30)/i24;
    i33=(a+c)/SUM(a,b,c,d)-i31;
    i34=i32-(a+c)/SUM(a,b,c,d);
    i35=(a+b)*(c+d)*(a+c)*(b+d);
    i36=SIGN(i35);
    i37=a*d-b*c;
    i38=SIGN(MAX(i37,0));
    i39=MAX(i37-SUM(a,b,c,d)/2,0);
    i40=i38*i39+(1-i38)*i37;
    i41=i36*i40;
    i42=i36*SQRT(i35)+(1-i36);
    i43=i36*i40/(i36*SQRT(i35)+(1-i36));
    i44=i27**2-2*i43*i27*i34+i34**2;
    i45=i28**2-2*i43*i28*i33+i33**2;

    lower=diff-SQRT(i44);
    upper=diff+SQRT(i45);

RUN;

PROC PRINT DATA=wilsonci NOOBS L;
    LABEL diff="Point Estimate of Difference"
           lower="Lower 95% CI (Newcombe-Wilson CC)"
           upper="Upper 95% CI (Newcombe-Wilson CC)"
           ;

```

```

;
VAR diff lower upper;
RUN;

TITLE1;
TITLE2;

%MEND MCNEM;

%MCNEM(
dataset=data23, /* データセット */
var1=std,       /* 検査変数 1 */
var2=var,       /* 検査変数 2 */
pos=1,          /* 検査変数一致 */
neg=0           /* 検査変数不一致 */
);

***** PG END *****;

```

```

*****
**** Sample code      :   No.11
****
**** Title            :   クラスターデータの解析
****                  :   (1 標本問題における推定値と分散の出力, 2 標本問題における検定)
****
**** Input    dataset :   データセット名
****           id      :   症例番号
****           mod     :   モダリティ
****           unit    :   ユニット
****           var     :   判定の真偽
****           num     :   モダリティの数 (1 or 2)
****           side    :   片側(1)、両側(2)検定の指定
****
**** Output          :   Estimator_1, Variance_1, Lower 95%CI_Est1, Upper 95%CI_Est1,
****                  :   Estimator_2, Variance_2, Lower 95%CI_Est2, Upper 95%CI_Est2,
****                  :   Covariance, Variance_diff, Lower 95%CI_Diff, Upper 95%CI_Diff
****                  :   z value, P value
****
**** SAS version      :   8.2
****
**** Reference        :   Xiao-Hua Zhou et al.(2002) Statistical Methods in Diagnostic
****                  :   Medicine
****
**** Dataset          :   Xiao-Hua Zhou et al.(2002) Statistical Methods in Diagnostic
****                  :   Medicine p104, Table4.2, p169, Tab5.4
****
*****;

%MACRO CLCOMP(dataset,id, mod, unit, var, num, side);

TITLE1 "SAS Sample Code 11 : Analysis for Clustered Binary Data";

PROC SORT DATA=&dataset;
  BY &id &unit &var;
RUN;

PROC TRANSPOSE DATA=&dataset OUT=data1;
  BY &id &unit;
  ID &mod;
RUN;

PROC MEANS DATA=data1 NOPRINT;
  VAR _1 _2;
  BY &id;
  OUTPUT OUT=data2 SUM=sum1 sum2;
RUN;

%IF &num=2 %THEN %DO;
PROC MEANS DATA=data2 NOPRINT;
  VAR sum1 sum2 _freq_;
  OUTPUT OUT=OUT N=n SUM=sum1 sum2 sum3;
RUN;

DATA _NULL_;
  SET OUT;
  CALL SYMPUT ("est1", sum1/sum3);
  CALL SYMPUT ("est2", sum2/sum3);
  CALL SYMPUT ("nbar", sum3/n);
  CALL SYMPUT ("n", n);
RUN;

DATA vdata1 vdata2;
  SET data2;
  RETAIN sumcore1 0 sumcore2 0 sumcore3 0;
  index1=sum1/_freq_;
  core1=(_freq_&nbar*(index1-&est1))*2;
  sumcore1=sum(sumcore1, core1);

```

```

index2=sum2/_freq_;
core2=(_freq_/&nbar*(index2-&est2))**2;
sumcore2=sum(sumcore2, core2);
core3=(_freq_/&nbar)**2*(index1-(&est1+&est2)/2)*(index2-(&est1+&est2)/2);
sumcore3=sum(sumcore3, core3);
v1=1/(&n*(&n-1))*sumcore1;
sd1=SQRT(v1);
v2=1/(&n*(&n-1))*sumcore2;
sd2=SQRT(v2);
cov=1/(&n*(&n-1))*sumcore3;
v_diff=v1+v2-2*cov;
sd_diff=SQRT(v_diff);
z=(&est1-&est2)/sd_diff;
p=(1-PROBNORM(ABS(Z)))*&side;
IF _n_=. THEN OUTPUT vdata1;
IF _n_=&n THEN DO;
    index1=&est1;
    index2=&est2;
    diff=index1-index2;
    ci_l_ind1=index1-PROBIT(0.975)*SQRT(v1);
    ci_u_ind1=index1+PROBIT(0.975)*SQRT(v1);
    ci_l_ind2=index2-PROBIT(0.975)*SQRT(v2);
    ci_u_ind2=index2+PROBIT(0.975)*SQRT(v2);
    ci_l_diff=(index1-index2)-PROBIT(0.975)*SQRT(v_diff);
    ci_u_diff=(index1-index2)+PROBIT(0.975)*SQRT(v_diff);
    OUTPUT vdata2;
END;
RUN;

PROC PRINT DATA=vdata2 NOOBS L;
    LABEL index1 = 'Estimator_1'
           sd1 = 'SD_1'
           ci_l_ind1 = 'Lower 95%CI_Est1'
           ci_u_ind1 = 'Upper 95%CI_Est1'
           index2 = 'Estimator_2'
           sd2 = 'SD_2'
           ci_l_ind2 = 'Lower 95%CI_Est2'
           ci_u_ind2 = 'Upper 95%CI_Est2'
           cov = 'Covariance'
           diff = 'Difference'
           sd_diff = 'SD_diff'
           ci_l_diff = 'Lower 95%CI_Diff'
           ci_u_diff = 'Upper 95%CI_Diff'
           z = 'Z value'
           p = 'P value'
           ;
    VAR index1 sd1 ci_l_ind1 ci_u_ind1 index2 sd2 ci_l_ind2 ci_u_ind2 cov diff sd_diff
    ci_l_diff
    ci_u_diff z p;
RUN;

%END;

%IF &num=1 %THEN %DO;
PROC MEANS DATA=data2 NOPRINT;
    VAR sum1 _freq_;
    OUTPUT OUT=OUT N=n SUM=sum1 sum2;
RUN;

DATA _NULL_;
    SET OUT;
    CALL SYMPUT ("est", sum1/sum2);
    CALL SYMPUT ("nbar", sum2/n);
    CALL SYMPUT ("n", n);
RUN;

DATA VDATA1 VDATA2;
    SET data2;

```



```

retain sumcore 0;
index=sum1/_freq_;
core=( _freq_/&nbar*(index-&est))**2;
sumcore=sum(sumcore, core);
v=1/(&n*(&n-1))*sumcore;
sd=SQRT(v);
IF _n_=. THEN OUTPUT vdata1;
IF _n_=&n THEN DO;
    index=&est;
    ci_l_ind=index-PROBIT(0.975)*SQRT(v);
    ci_u_ind=index+PROBIT(0.975)*SQRT(v);
    OUTPUT VDATA2;
END;
RUN;

PROC PRINT DATA=vdata2 NOOBS L;
    LABEL index = 'Estimator'
           sd= 'SD'
           ci_l_ind = 'Lower 95%CI'
           ci_u_ind = 'Upper 95%CI'
           ;
    VAR index sd ci_l_ind ci_u_ind;
RUN;

%END;

TITLE1;

%MEND CLCOMP;

/* 2 変数の比較の場合 */
%CLCOMP (
dataset = data4254,
id = id ,
mod = mod,
unit = unit,
var = var,
num = 2,
side = 2
);
/* データセット名 */
/* 症例番号 */
/* モダリティ */
/* ユニット */
/* 判定の真偽 */
/* モダリティの数 (1 or 2) */
/* 片側(1)、両側(2)検定の指定 */

/* 1 変数の場合 */
%CLCOMP (
dataset = data4254,
id = id ,
mod = mod,
unit = unit,
var = var,
num = 1,
side = 2
);
/* データセット名 */
/* 症例番号 */
/* モダリティ */
/* ユニット */
/* 判定の真偽 */
/* モダリティの数 (1 or 2) */
/* 片側(1)、両側(2)検定の指定 */

***** PG END *****;

```