

SAS Forum ユーザー会 学術総会 2006

XMLを用いた統計解析結果出力標準化の試案

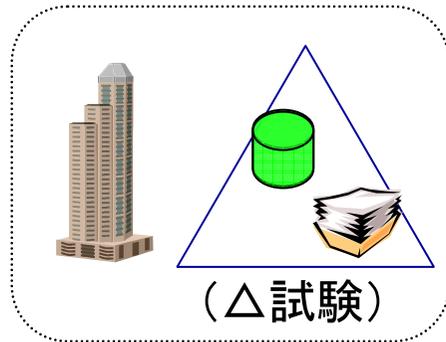
Tentative Plan for Standardization of Output for
Statistical Analysis Outcomes using XML

Masashi Takeyasu
Kiyoshi Saburomaru

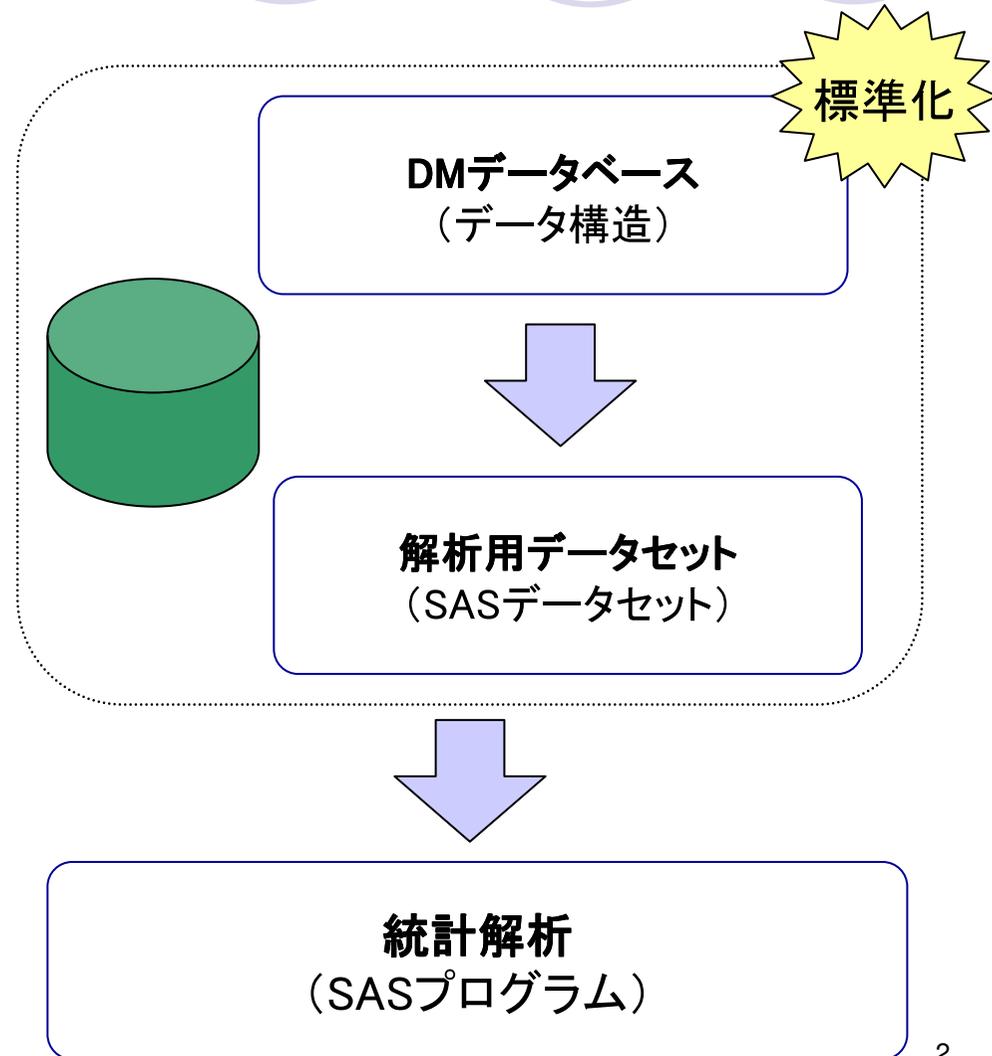
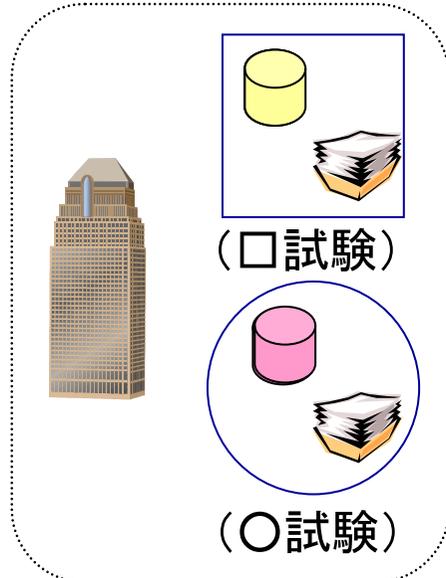
Medical & Pharmaceutical Solution Center II
CAC Corporation

臨床開発業務でのデータ構造

【A社】

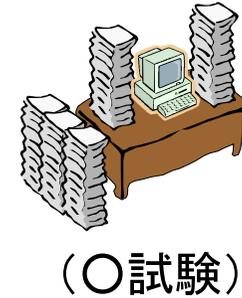
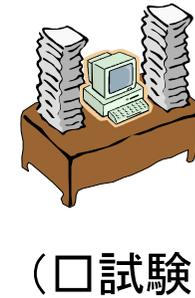
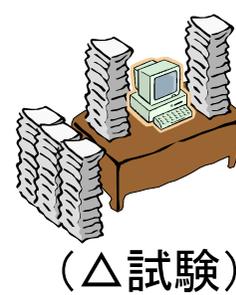
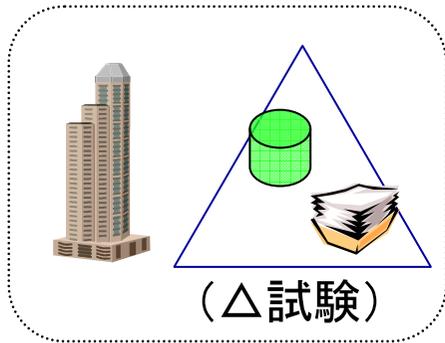


【B社】

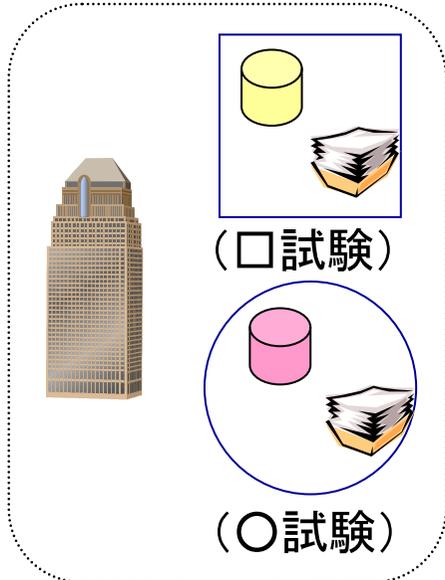


臨床開発業務での出力プロセス

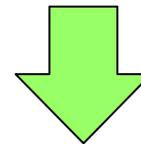
【A社】



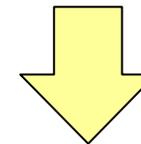
【B社】



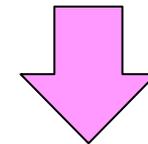
個々の
出力形式に合わせた
SASプログラム



図表作成



図表作成



図表作成

標準化の試案

XML

(Extensible Markup Language)

出力プロセス

SASプログラムと
出力形式の関係

出力プロセスの
標準化

統計解析の
図表作成例

出力形式(1)

【被験者背景一覧A】

A	B	C	D	E	F	G	H
	Usubjid	Group	Sex	Age	Height	Weight	
	A-01	A	M	35	169.2	57.6	
	A-02	B	M	54	167.7	72.6	
	A-03	A	F	28	142.1	47.5	
	A-10	B	F	43	160.5	42.4	

同じ被験者背景情報

【被験者背景一覧B】

A	B	C	D	E	F	G	H
			Study Site	Sex			
	Usubjid	Group	Safety	Age	Height	Weight	
	A-01	A	ABC	M	169.2	57.6	
			Y	35			
	A-02	B	ABC	M	167.7	72.6	
			N	54			
	A-03	A	XYZ	F	142.1	47.5	
			Y	28			
	A-04	A	XYZ	M	170	58.9	
			Y	51			
	A-10	B	ABC	F	160.5	42.4	
			Y	43			

2行

出力形式(2)

	A	B	C	D	E
1					
2			A群		B群
3	投与前		集計値		集計値
4		例数	6		4
5		平均値	3.59		5.10
6		標準偏差	4.47		6.39
7		最大値	12.03		11.68
8		中央値	2.82		5.84
9		最小値	-0.12		-2.98
10	投与中		集計値		集計値
11		例数	6		4
12		平均値	3.03		4.87
13		標準偏差	3.15		6.66
14		最大値	8.83		11.98
15		中央値	2.18		5.30
16		最小値	-0.01		-3.09
17	投与後		集計値		集計値
18		例数	5		4
19		平均値	3.04		3.89
20		標準偏差	4.35		5.94
21		最大値	10.36		9.66
22		中央値	2.21		4.72
23		最小値	-0.36		-3.52

縦

時系列

同じ記述統計量情報

横

	A	B	C	D	E
1					
2			投与前	投与中	投与後
3	A群	例数	6	6	5
4		平均値	3.59	3.03	3.04
5		標準偏差	4.47	3.15	4.35
6		最大値	12.03	8.83	10.36
7		中央値	2.82	2.18	2.21
8		最小値	-0.12	-0.01	-0.36
9	B群	例数	4	4	4
10		平均値	5.10	4.87	3.89
11		標準偏差	6.39	6.66	5.94
12		最大値	11.68	11.98	9.66
13		中央値	5.84	5.30	4.72
14		最小値	-2.98	-3.09	-3.52

SASプログラムと図表の関係

【被験者背景一覧A】

Usubjid	Group	Sex	Age	Height	Weight
A-01	A	M	35	169.2	57.6
A-02	B	M	54	167.7	72.6
A-03	A	F	28	142.1	47.5
A-10	B	F	43	160.5	42.4

DDE (Dynamic Data Exchange)

出力先情報である
シート名・セル範囲

出力先
アプリケーション

出力対象である
変数名・出力項目順

```
*被験者背景一覧A出力;  
filename xls dde "Excel | haikeiA!r3c2:r50c7";  
data _null_;  
  set Baseinfo;  
  Blank=" ";  
  file xls dsd dlm='09'x;  
  put Id Group Sex Height Weight;  
run;
```

SASプログラムと図表の関係

【被験者背景一覧B】

Usubjid	Group	Study Site	Sex	Height	Weight
A-01	A	ABC	M	169.2	57.6
		Y	35		
A-02	B	ABC	M	167.7	72.6
		N	54		
A-03	A	XYZ	F	142.1	47.5
		Y	28		
A-04	A	XYZ	M	170	58.9
		Y	51		
A-10	B	ABC	F	160.5	42.4
		Y	43		

DDE (Dynamic Data Exchange)

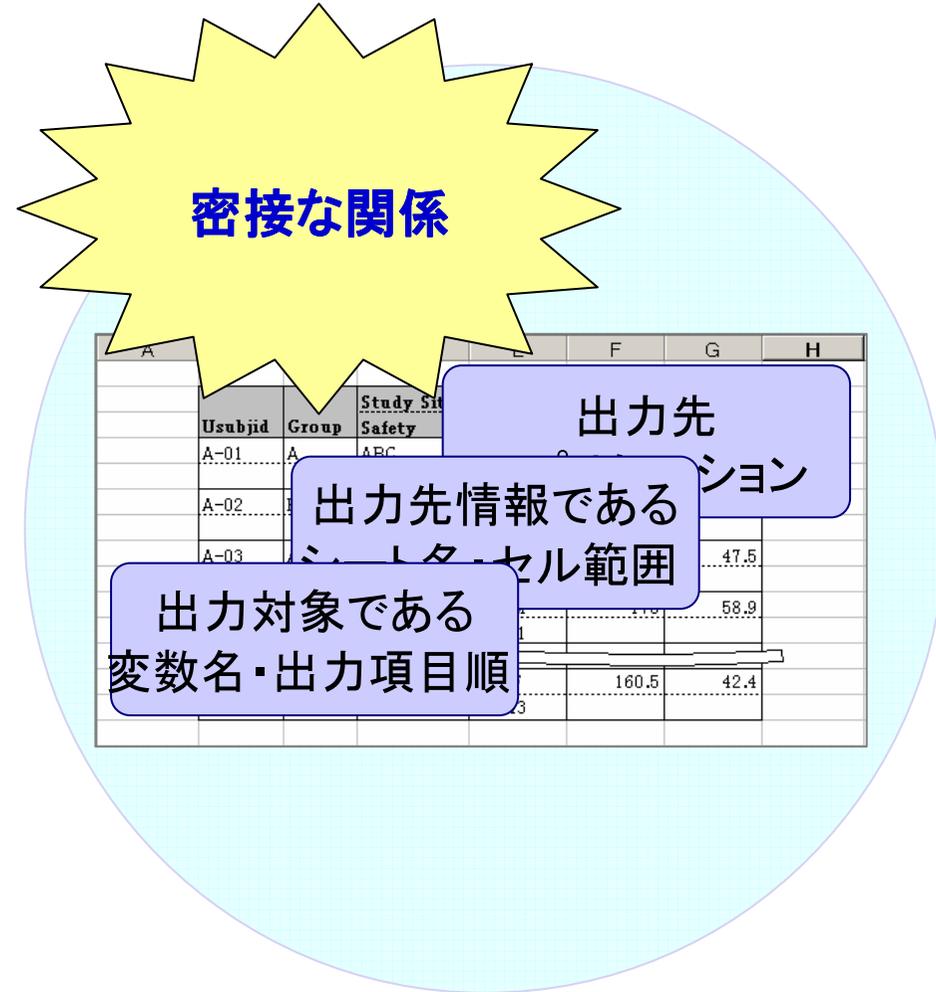
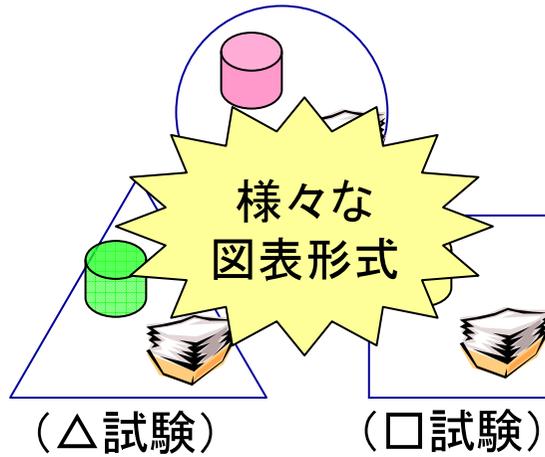
出力先情報である
シート名・セル範囲

出力先
アプリケーション

出力対象である
変数名・出力項目順

```
*被験者背景一覧B出力;  
filename xls dde "Excel | haikeiB!r4c2:r103c7";  
data _null_;  
  set Baseinfo;  
  Blank=" ";  
  file xls dsd dlm='09'x;  
  put Id Group StudySite Sex Height Weight;  
  put Blank Blank Safety Age Blank Blank;  
run;
```

標準化(出カプロセス)の限界



出力プロセス

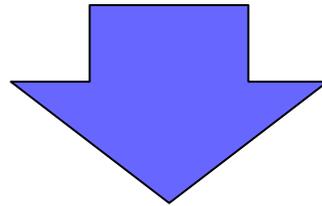
SASプログラムと
出力形式の関係

出力プロセスの
標準化

統計解析の
図表作成例

XMLとは

- 「タグ」とよばれる特殊な文字列を利用してデータや論理構造を記述できるテキスト形式ファイル



- 臨床試験での収集データや統計解析結果データなどのSASデータセットをデータ構造と共に出力が可能

XMLを利用する意図

CSV (Comma Separated Values)

```
A,6,39,9.39503415,51,48,39,31,28  
B,4,49,5.35412613,54,53.5,49.5,44.5,43
```



「,(カンマ)」で区切って並べた形式ファイルであるため何番目のデータが何を意味した値なのか事前に理解しておく必要がある。

「タグ」と呼ばれる特殊な文字列を利用してデータの意味や論理構造を記述できる。



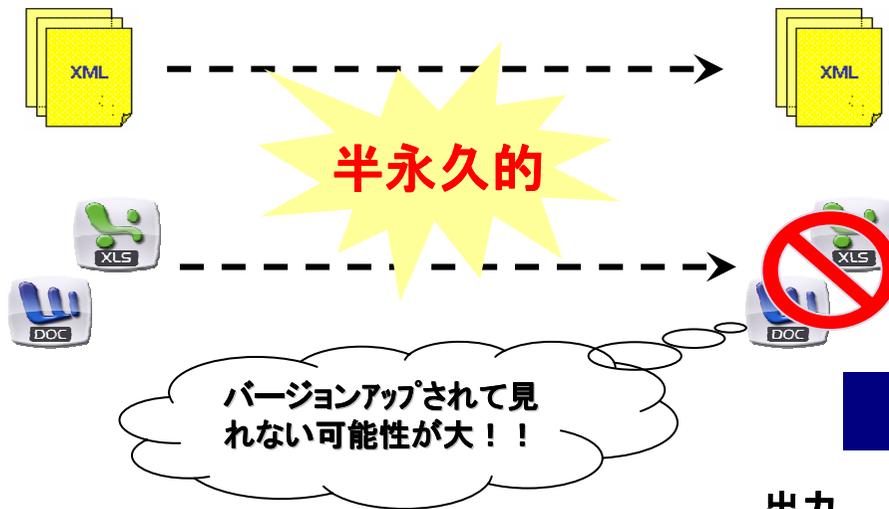
XML (Extensible Markup Language)

```
<?xml version="1.0" encoding="utf-8" ?>  
<TABLE>  
  <AGE>  
    <Group> A </Group>  
    <n> 6 </n>  
    <mean> 39 </mean>  
    <std> 9.39503415 </std>  
    <max> 51 </max>  
    <q3> 48 </q3>  
    <median> 39 </median>  
    <q1> 31 </q1>  
    <min> 28 </min>  
  </AGE>  
  <AGE>  
    <Group> B </Group>  
    <n> 4 </n>  
    <mean> 49 </mean>  
    <std> 5.35412613 </std>  
    <max> 54 </max>  
    <q3> 53.5 </q3>  
    <median> 49.5 </median>  
    <q1> 44.5 </q1>  
    <min> 43 </min>  
  </AGE>  
</TABLE>
```

<タグ>

データの長期保存

XMLによって長期保存が可能



テキストファイルだから
大丈夫！！

バージョンアップされて見
れない可能性が大！！

異なるOS間での共有



出力プロセス

SASプログラムと
出力形式の関係

出力プロセスの
標準化

統計解析の
図表作成例

図表作成例手順

- 統計解析
- 図表作成
- XMLへの出力
- 図表テンプレートの作成
- XMLデータの読み込み
- デモンストレーション

使用するデータ

BaseInfo (被験者背景情報)

データ構造

変数名	変数ラベル	変数型	
Usubjid	被験者ID	Char	-
Group	投与群	Char	-
Study Site	治験実施施設	Char	
Sex	性別	Char	M-男、F-女
Brthdtm	生年月日	Char	
Age	年齢(歳)	Num	
Height	身長(cm)	Num	
Weight	体重(kg)	Num	
Safety	安全性解析対象集団	Char	Y or N

実データ

Usubjid	Group	Study Site	Sex	Brthdtm	Age	Height	Weight	Safety
A-01	A	ABC	M	1970/04/01	35	169.2	57.6	Y
A-02	B	ABC	M	1952/02/17	54	167.7	72.6	N
A-03	A	XYZ	F	1977/06/07	28	142.1	47.5	Y
A-04	A	XYZ	M	1954/11/11	51	170.0	58.9	Y
A-10	B	ABC	F	1962/07/30	43	160.5	42.4	Y

統計解析(1)

/*記述統計量:マクロプログラムの登録*/

```
%macro OutputMeans(dataset, vVar);  
  proc univariate data=&dataset. noprint;  
    class group;  
    var &vVar.;  
    output out=&vVar. N=n MEAN=mean STD=std MIN=min  
      MAX=max MEDIAN=median  
      Q1=q1 Q3=q3;  
  run;  
%mend OutputMeans;
```

統計解析(2)

/*t検定:マクロプログラムの登録*/

```
%macro MacroTtests(dataset, vVar);  
ods output Statistics=_Stats Ttests=_Ttests Equality=_Equality;  
proc ttest data=&dataset.;  
class group;  
var &vVar.;  
run;  
ods output close;
```

/* P値の算出 */

```
data t_&vVar.;  
merge _Ttests _Equality;  
by Variable;  
if ProbF>0.05 and Variances="Unequal" then delete;  
if ProbF=<0.05 and Variances="Equal" then delete;  
keep Variances tValue DF Probt;  
run;
```

```
%mend MacroTtests;
```

統計解析(3)

/*Groupでソート*/

```
proc sort data=Baseinfo out=wk_Base; by group; run;
```

/*記述統計量の算出*/

```
%OutputMeans(wk_Base, Weight); %OutputMeans(wk_Base, Height);  
%OutputMeans(wk_Base, Age);
```

/*t検定*/

```
%MacroTtests(wk_Base, Weight); %MacroTtests(wk_Base, Height);  
%MacroTtests(wk_Base, Age);
```

図表作成

被験者背景の記述統計量

	A	B	C	D	E	F	G	H	I
1									
2	SAS Forum ユーザー会 学術総会 2006 デモ								
3				A群	B群			検定*	
4	年齢			集計値	集計値				
5		記述統計量 (歳)	例数					P=	
6			平均値						
7			標準偏差						
8			最大値						
9			中央値						
10		最小値							
11	身長			集計値	集計値				
12		記述統計量 (cm)	例数					P=	
13			平均値						
14			標準偏差						
15			最大値						
16			中央値						
17		最小値							
18	体重			集計値	集計値				
19		記述統計量 (kg)	例数					P=	
20			平均値						
21			標準偏差						
22			最大値						
23			中央値						
24		最小値							
25	* t 検定(有意水準、両側 5%)								
26									

XML出力(1)

*/*XML出力用マクロプログラムの登録*/*

```
%macro OutputXML (inlib, file, Dataset);  
  libname Trans xml "&file." xmltype=generic encoding='UTF-8';  
  proc copy in=&inlib. out=Trans;  
    select &Dataset.;  
  run;  
%mend OutputXML;
```

たった6行

SASデータセットの
変数名を
タグ名として出力

出力文字コード
「UTF-8」

XML出力(2)

/*XML出力*/

%OutputXML(work, C:¥TEMP¥Weight.xml, Weight);	*体重:記述統計量;
%OutputXML(work, C:¥TEMP¥t_Weight.xml, t_Weight);	*体重:t検定;
%OutputXML(work, C:¥TEMP¥Height.xml, Height);	*身長:記述統計量;
%OutputXML(work, C:¥TEMP¥t_Height.xml, t_Height);	*身長:t検定;
%OutputXML(work, C:¥TEMP¥Age.xml, Age);	*年齢:記述統計量;
%OutputXML(work, C:¥TEMP¥t_Age.xml, t_Age);	*年齢:t検定;

『SAS libname エンジン』を用いることで、出力形式を意識することなく、全てのSASデータセットを簡単に出力することが可能

XML出力(3)

年齢の記述統計量(XMLファイル)

```
<?xml version="1.0" encoding="utf-8" ?>
<TABLE>
  <AGE>
    <Group> A </Group>
    <n> 6 </n>
    <mean> 39 </mean>
    <std> 9.39503415 </std>
    <max> 51 </max>
    <q3> 48 </q3>
    <median> 39 </median>
    <q1> 31 </q1>
    <min> 28 </min>
  </AGE>
  <AGE>
    <Group> B </Group>
    <n> 4 </n>
    <mean> 49 </mean>
    <std> 5.35412613 </std>
    <max> 54 </max>
    <q3> 53.5 </q3>
    <median> 49.5 </median>
    <q1> 44.5 </q1>
    <min> 43 </min>
  </AGE>
</TABLE>
```

図表テンプレート作成

	A	B	C	D	E	F	G	H	I
1									
2	SAS Forum ユーザー会 学術総会 2006 デモ								
3				A群	B群	検定*			
4	年齢			集計値	集計値				
5		記述 統計量 (歳)	例数	age.n			P= t_age.probt		
6			平均値	age.mean					
7			標準偏差	age.std					
8			最大値	age.max					
9			中央値	age.median					
10			最小値	age.min					
11	身長			集計値	集計値				
12		記述 統計量 (cm)	例数	height.n			P= t_height.probt		
13			平均値	height.mean					
14			標準偏差	height.std					
15			最大値	height.max					
16			中央値	height.median					
17			最小値	height.min					
18	体重			集計値	集計値				
19		記述 統計量 (kg)	例数	weight.n			P= t_weight.probt		
20			平均値	weight.mean					
21			標準偏差	weight.std					
22			最大値	weight.max					
23			中央値	weight.median					
24			最小値	weight.min					
25	* t 検定(有意水準、両側 5%)								
26									

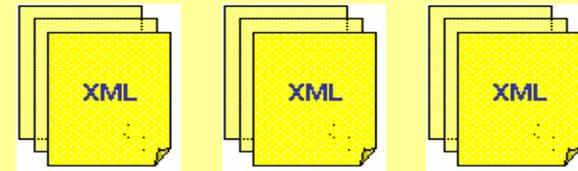
ファイル名.タグ名

XMLデータの読み込み

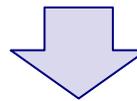
図表テンプレート

A	B	C	D	E	F	G	H	I
SIS Fream ユーザー名 学習結果 2011 アニ								
4	年次	性別	年齢	身長	体重	身長	体重	身長
5	5	男	15	150	40	150	40	150
6	6	女	16	160	45	160	45	160
7	7	男	17	170	50	170	50	170
8	8	女	18	180	55	180	55	180
9	9	男	19	190	60	190	60	190
10	10	女	20	200	65	200	65	200
11	11	男	21	210	70	210	70	210
12	12	女	22	220	75	220	75	220
13	13	男	23	230	80	230	80	230
14	14	女	24	240	85	240	85	240
15	15	男	25	250	90	250	90	250
16	16	女	26	260	95	260	95	260
17	17	男	27	270	100	270	100	270
18	18	女	28	280	105	280	105	280
19	19	男	29	290	110	290	110	290
20	20	女	30	300	115	300	115	300
21	21	男	31	310	120	310	120	310
22	22	女	32	320	125	320	125	320
23	23	男	33	330	130	330	130	330
24	24	女	34	340	135	340	135	340
25	25	男	35	350	140	350	140	350
26	26	女	36	360	145	360	145	360
27	27	男	37	370	150	370	150	370
28	28	女	38	380	155	380	155	380
29	29	男	39	390	160	390	160	390
30	30	女	40	400	165	400	165	400
31	31	男	41	410	170	410	170	410
32	32	女	42	420	175	420	175	420
33	33	男	43	430	180	430	180	430
34	34	女	44	440	185	440	185	440
35	35	男	45	450	190	450	190	450
36	36	女	46	460	195	460	195	460
37	37	男	47	470	200	470	200	470
38	38	女	48	480	205	480	205	480
39	39	男	49	490	210	490	210	490
40	40	女	50	500	215	500	215	500
41	41	男	51	510	220	510	220	510
42	42	女	52	520	225	520	225	520
43	43	男	53	530	230	530	230	530
44	44	女	54	540	235	540	235	540
45	45	男	55	550	240	550	240	550
46	46	女	56	560	245	560	245	560
47	47	男	57	570	250	570	250	570
48	48	女	58	580	255	580	255	580
49	49	男	59	590	260	590	260	590
50	50	女	60	600	265	600	265	600
51	51	男	61	610	270	610	270	610
52	52	女	62	620	275	620	275	620
53	53	男	63	630	280	630	280	630
54	54	女	64	640	285	640	285	640
55	55	男	65	650	290	650	290	650
56	56	女	66	660	295	660	295	660
57	57	男	67	670	300	670	300	670
58	58	女	68	680	305	680	305	680
59	59	男	69	690	310	690	310	690
60	60	女	70	700	315	700	315	700
61	61	男	71	710	320	710	320	710
62	62	女	72	720	325	720	325	720
63	63	男	73	730	330	730	330	730
64	64	女	74	740	335	740	335	740
65	65	男	75	750	340	750	340	750
66	66	女	76	760	345	760	345	760
67	67	男	77	770	350	770	350	770
68	68	女	78	780	355	780	355	780
69	69	男	79	790	360	790	360	790
70	70	女	80	800	365	800	365	800
71	71	男	81	810	370	810	370	810
72	72	女	82	820	375	820	375	820
73	73	男	83	830	380	830	380	830
74	74	女	84	840	385	840	385	840
75	75	男	85	850	390	850	390	850
76	76	女	86	860	395	860	395	860
77	77	男	87	870	400	870	400	870
78	78	女	88	880	405	880	405	880
79	79	男	89	890	410	890	410	890
80	80	女	90	900	415	900	415	900
81	81	男	91	910	420	910	420	910
82	82	女	92	920	425	920	425	920
83	83	男	93	930	430	930	430	930
84	84	女	94	940	435	940	435	940
85	85	男	95	950	440	950	440	950
86	86	女	96	960	445	960	445	960
87	87	男	97	970	450	970	450	970
88	88	女	98	980	455	980	455	980
89	89	男	99	990	460	990	460	990
90	90	女	100	1000	465	1000	465	1000

統計解析結果XML



- age.xml / t_age.xml
- height.xml / t_height.xml
- weight.xml / t_weight.xml
- baseinfo.xml



完成図表

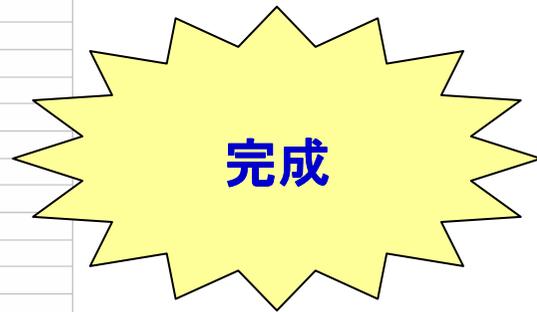
A	B	C	D	E	F	G	H	I
SIS Fream ユーザー名 学習結果 2011 アニ								
4	年次	性別	年齢	身長	体重	身長	体重	身長
5	5	男	15	150	40	150	40	150
6	6	女	16	160	45	160	45	160
7	7	男	17	170	50	170	50	170
8	8	女	18	180	55	180	55	180
9	9	男	19	190	60	190	60	190
10	10	女	20	200	65	200	65	200
11	11	男	21	210	70	210	70	210
12	12	女	22	220	75	220	75	220
13	13	男	23	230	80	230	80	230
14	14	女	24	240	85	240	85	240
15	15	男	25	250	90	250	90	250
16	16	女	26	260	95	260	95	260
17	17	男	27	270	100	270	100	270
18	18	女	28	280	105	280	105	280
19	19	男	29	290	110	290	110	290
20	20	女	30	300	115	300	115	300
21	21	男	31	310	120	310	120	310
22	22	女	32	320	125	320	125	320
23	23	男	33	330	130	330	130	330
24	24	女	34	340	135	340	135	340
25	25	男	35	350	140	350	140	350
26	26	女	36	360	145	360	145	360
27	27	男	37	370	150	370	150	370
28	28	女	38	380	155	380	155	380
29	29	男	39	390	160	390	160	390
30	30	女	40	400	165	400	165	400
31	31	男	41	410	170	410	170	410
32	32	女	42	420	175	420	175	420
33	33	男	43	430	180	430	180	430
34	34	女	44	440	185	440	185	440
35	35	男	45	450	190	450	190	450
36	36	女	46	460	195	460	195	460
37	37	男	47	470	200	470	200	470
38	38	女	48	480	205	480	205	480
39	39	男	49	490	210	490	210	490
40	40	女	50	500	215	500	215	500
41	41	男	51	510	220	510	220	510
42	42	女	52	520	225	520	225	520
43	43	男	53	530	230	530	230	530
44	44	女	54	540	235	540	235	540
45	45	男	55	550	240	550	240	550
46	46	女	56	560	245	560	245	560
47	47	男	57	570	250	570	250	570
48	48	女	58	580	255	580	255	580
49	49	男	59	590	260	590	260	590
50	50	女	60	600	265	600	265	600
51	51	男	61	610	270	610	270	610
52	52	女	62	620	275	620	275	620
53	53	男	63	630	280	630	280	630
54	54	女	64	640	285	640	285	640
55	55	男	65	650	290	650	290	650
56	56	女	66	660	295	660	295	660
57	57	男	67	670	300	670	300	670
58	58	女	68	680	305	680		

XMLデータの読み込み

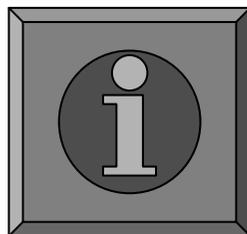


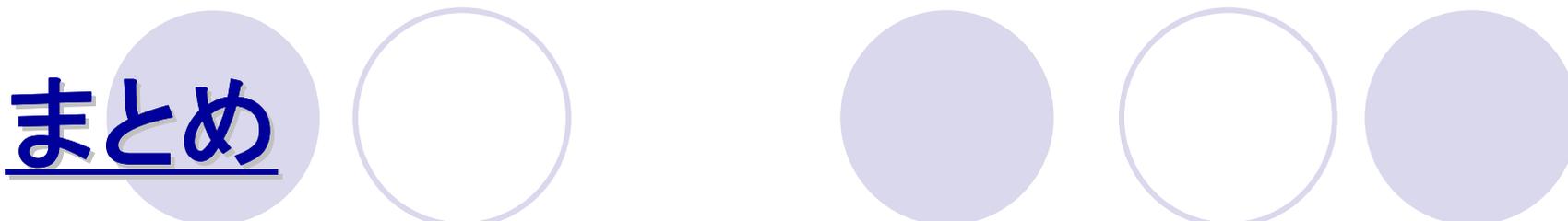
実行

	A	B	C	D	E	F G	H	I
1								
2	SAS Forum ユーザー会 学術総会 2006 デモ							
3				A群	B群	検定*		
4	年齢			集計値	集計値			
5		記述 (統計量)	例数	6	4	P= 0.10233256		
6			平均値	39.3333333	49			
7			標準偏差	9.39503415	5.35412613			
8			最大値	51	54			
9			中央値	39	49.5			
10			最小値	28	43			
11	身長			集計値	集計値			
12		記述 (cm 統計量)	例数	6	4	P= 0.8228622		
13			平均値	162.7666667	164.25			
14			標準偏差	10.9341057	7.99228795			
15			最大値	171.2	173.5			
16			中央値	166.4	164.1			
17			最小値	142.1	155.3			
18	体重			集計値	集計値			
19		記述 (kg 統計量)	例数	6	4	P= 0.82108899		
20			平均値	53.4333333	54.9			
21			標準偏差	6.3804911	13.5730124			
22			最大値	60.4	72.6			
23			中央値	54.25	52.3			
24			最小値	45.3	42.4			
25	*t検定(有意水準、両側 5%)							
26				■ ■ ■ ■ ■ ■ ■ ■ ■ ■	▶	横展開		



デモンストレーション

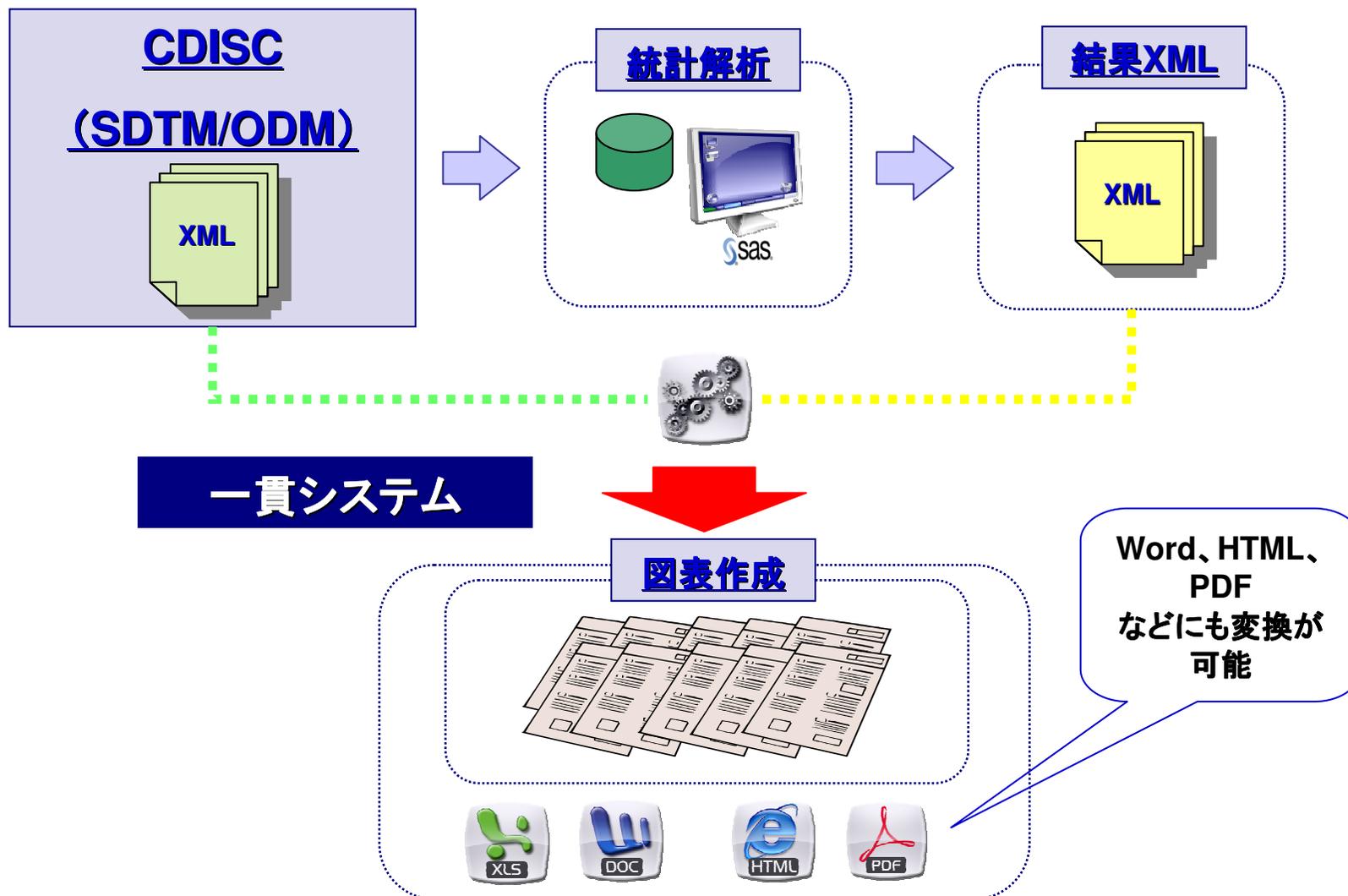




まとめ

- XMLを用いることでSASプログラムと図表との『密接な関係』を切り離し、出力プロセスの標準化の可能性
- 図表テンプレートをそれぞれ作成し、XMLとリンクしておくことで容易に図表作成が可能
- 個々の出力形式に合わせたSASプログラムの作成を不要とするものの可能性

今後の展開: CDISCとの連携



ご清聴ありがとうございました。

株式会社 シーエーシー

システムビジネスユニット 医薬第二センター

武安 雅史

cac15894@cac.co.jp

三郎丸 清

saburoma@cac.co.jp