

Friday, November 17, 2023



Teaching Methods for SAS Programmers. How to Make Learning Fun

EPS Corporation
Yutaka Morioka

Menu :


1. **Love** the Data Handling.
2. Presentations outside the company increase engagement (**Love**)
3. Visualization makes you **Love** Stat and Data.
4. **Love** programming with games
5. You will **Love** the results of your own manual analysis. No Procedure! No Library!

1. Love the Data Handling.

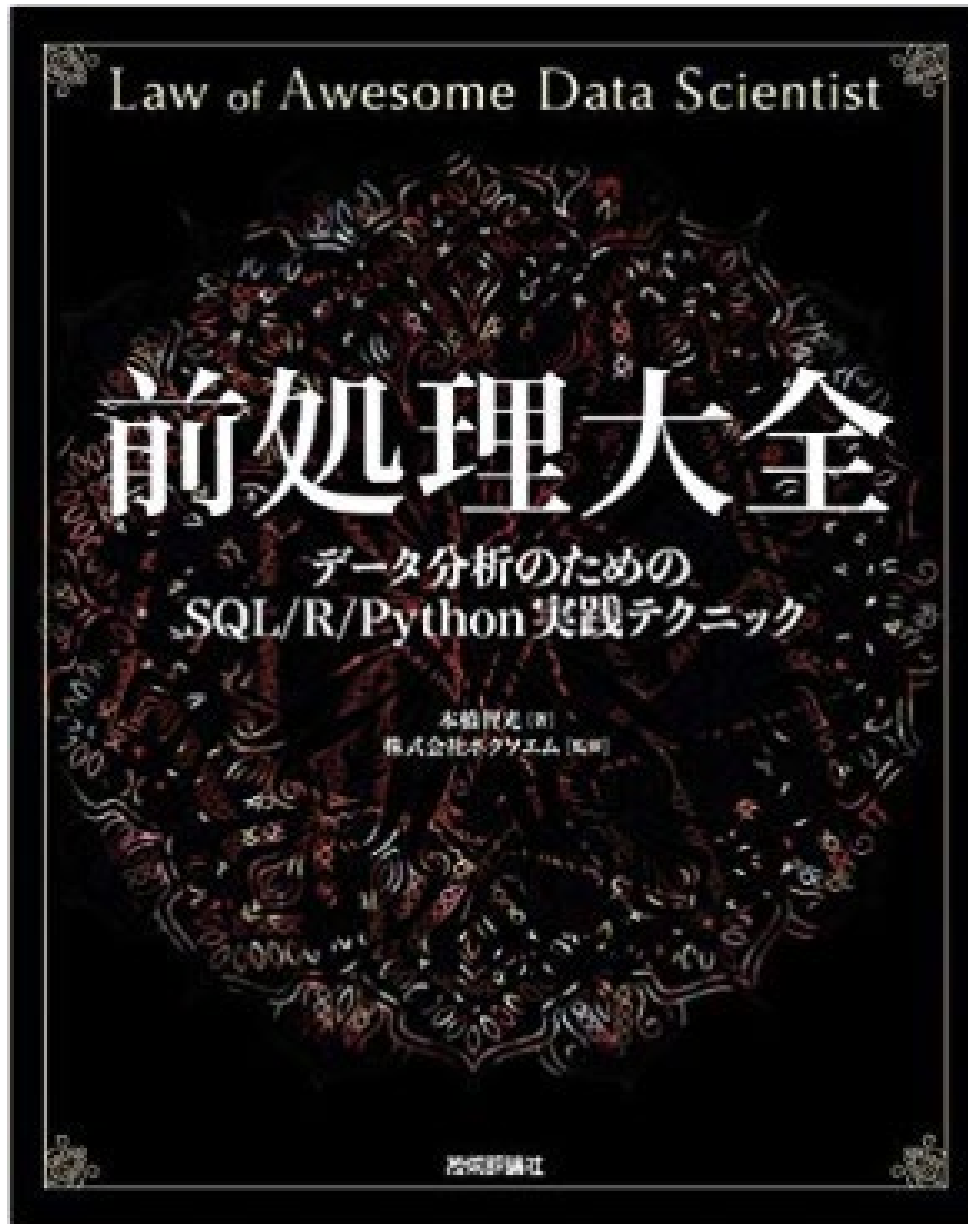
Data Scientist: The Sexiest Job of the 21st Century

Do you enjoy statistical programming?

Data Handling Programming Programming Programming
Data Handling Programming Programming
Data Handling Programming
Data Handling OJT OJT OJT
OJT OJT



Law of Awesome Data Scientist



2018年4月13日発売
本橋智光 著, 株式会社ホクソエム 監修
技術評論社
B5変形判 / 336ページ
定価 (本体3,000円+税)
ISBN 978-4-7741-9647-3

データサイエンスの現場において、その業務は「前処理」と呼ばれるデータの整形に90%以上の時間を費やすと言われていています。

本書はデータサイエンスに取り組む上で欠かせない

「前処理」を効率よくこなすことで、予測モデルの構築やデータモデリングといった

本来のデータサイエンス業務に時間を割くことができるわけです。

「前処理スキル」の効率的な処理方法を網羅的に習得できる構成となっています。ほとんどの問題についてR、Python、SQLを用いた実装方法を紹介しますので、複数のプロジェクトに関わるようなデータサイエンスの現場で重宝するでしょう。

Law of Awesome Data Scientist



A great book . But only R, Python and SQL

I'll show you SAS version.



I hope you download and read it!

SASユーザー総会2023

 圧縮版PDFダウンロード(12.1 MB)

Day 1 2023/09/13(水)

開始	終了	トラック1 福武ラーニングシアター	ト 福
9:30		受付開始	
9:50	10:00	プレオープニング イーピーエス株式会社 森岡 裕(世話人)	
10:00	10:30	前処理大全 SASバージョン イーピーエス株式会社 森岡 裕 	SA SA 絹
10:30	11:00	解析プロシジャで作成されるODS統計解析Plotに ついて イーピーエス株式会社 坂村 奏羊 	M イ 小

<https://sas-user2023.ywstat.jp/program.html>

第2章 抽出

2-1 データ列指定による抽出

SASではkeep/dropステートメント・データセットオプションで列(変数)抽出が可能です

```
data A;  
set sashelp.FISH;  
keep Species Weight;  
run;
```

```
data A;  
set sashelp.class(keep=Species Weight);  
run;
```

○ 先に変数を絞って処理に入るため負荷が軽い(速い)

○ ノーマルで見やすい
データステップ中で新規作成
した変数も指定可能

```
data A(keep=Species Weight);  
set sashelp.class;  
run;
```

△ 左のコードと同じ意味。特にメリットがないので普通に書いてと思う

	Species	Weight	Length1	Length2	Length3	Height	Width
1	Bream	242	23.2	25.4	30	11.52	4.02
2	Bream	290	24	26.3	31.2	12.48	4.3056
3	Bream	340	23.9	26.5	31.1	12.3778	4.6961
4	Bream	363	26.3	29	33.5	12.73	4.4555
5	Bream	430	26.5	29	34	12.444	5.134
6	Bream	450	26.8	29.7	34.7	13.6024	4.9274
7	Bream	500	26.8	29.7	34.5	14.1795	5.2785
8	Bream	390	27.6	30	35	12.67	4.69
9	Bream	450	27.6	30	35.1	14.0049	4.8438



	Species	Weight
1	Bream	242
2	Bream	290
3	Bream	340
4	Bream	363
5	Bream	430
6	Bream	450
7	Bream	500
8	Bream	390
9	Bream	450

第2章 抽出

2-1 データ列指定による抽出

```
data A;  
set sashelp.fish;  
keep W: ;  
run;
```

Weight Width
接頭語による列抽出

```
data A;  
set sashelp.fish;  
keep Length1-Length3 ;  
run;
```

Length1 Length2 Length3
連番による列抽出

```
data A;  
set sashelp.fish;  
keep Length2--Height ;  
run;
```

Length2 Length3 Height
変数位置による列抽出

```
data A;  
set sashelp.fish;  
keep _numeric_ ;  
run;
```

Weight Length1 Length2 Length3 Height Width
数値型指定の列抽出

```
data A;  
set sashelp.fish;  
keep _character_ ;  
run;
```

Species
文字型指定の列抽出

	Species	Weight	Length1	Length2	Length3	Height	Width
1	Bream	242	23.2	25.4	30	11.52	4.02
2	Bream	290	24	26.3	31.2	12.48	4.3056
3	Bream	240	22.9	26.5	21.1	12.2772	4.2921

第2章 抽出

2-2 条件指定による抽出

```
data A;  
set sashelp.class;  
where age=11;  
run;
```

ノーマル

```
data A;  
set sashelp.class;  
if age=11;  
run;
```



サブセットIFのメリットはデータステップ中で生成した変数で絞れること。この場合は元から存在している変数なので、時間の無駄でしかない。

*BMI = 体重 (ポンド) / (身長 (インチ) x 身長 (インチ)) x 703

```
data A;  
set sashelp.class;  
BMI=divide(Weight , Height**2)*703;  
if BMI >= 18 ;  
run;
```

余談：

私は割り算をスラッシュ「/」で記述することを好まない。Divide関数だと0除算のWARNINGもせずに、無限大の特殊欠損(.I)で返してくれるので異常検知しやすい。演算子の実行順も関数の方がわかりやすいと感じる。賛同者があまりいないが…

作成後のデータセットにBMIを残したいのであれば、サブセットIFが正解

```
data A;  
set sashelp.class;  
if _N_ in (2,3,5) ;  
run;
```



_N_などの自動変数や,first. last. ln.eofオプションなどを利用する場合もサブセットIFでなければできないので、正解

**Awesome
or
Not Awesome**

2. Get Out of the office ! !

The graphic features a dark purple background with a grid of rounded squares in various shades of purple and blue. A white rectangular border frames the central text. The text is white and reads: SAS[®] GLOBAL FORUM 2019. Below the border, the text 'USERS PROGRAM' and 'APRIL 28 - MAY 1, 2019 | DALLAS, TX' is displayed in a smaller white font. A small globe icon is visible in the bottom right corner of the graphic.

SAS[®]
GLOBAL
FORUM
2019

USERS PROGRAM

APRIL 28 - MAY 1, 2019 | DALLAS, TX

Approach 4: DOSUBL Function + SQL View + Hash Object

The solution

```
data output4;
  set sashelp_class;
  if _N_ = 1 then do;
    rc=dosubl("proc sql noprint; create view v1 as
              select sex ,mean(age) as age_mean
              from sashelp_class group by sex; quit;");
    drop rc;
    declare hash h1(dataset:"v1");
    h1.definekey("sex");
    h1.definedata("age_mean");
    h1.definedone();
  end;
  if h1.find() ne 0 then call missing(age_mean);
run;
```

	Sex	age_mean
1	F	13.222222222
2	M	13.4

SQL view: v1

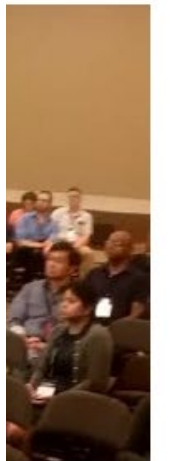
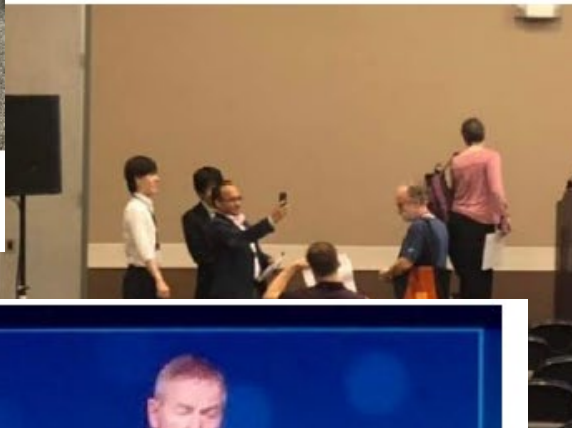
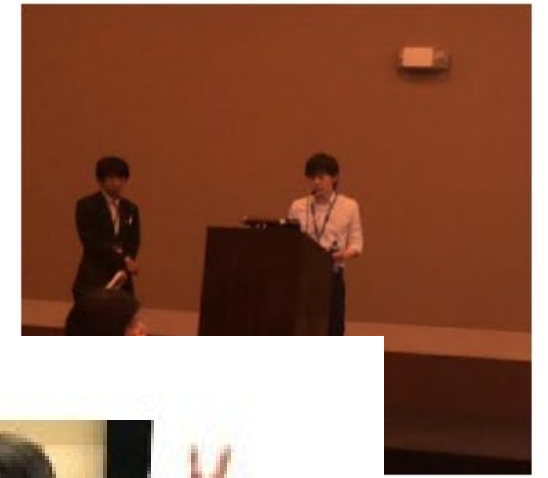


新手一生

Kōzō Masuda (升田 幸三)
Japanese Professional Shogi Player

At first I was afraid putting a SQL view created through the DOSUBL function into a hash object could be an error, but when I tried, it actually worked.

The above kanji is a motto of a professional Shogi (Japanese Chess) player, meaning “always try a new move”. I always wish to be like such an leading SAS programmer.



The slide features a white background with a prominent horizontal blue band across the middle. On the right side, there are several thin, light blue curved lines that sweep across the top and right edges of the slide. The text '3. Visualization makes you Love Stat and Data.' is centered within the blue band. The word 'Love' is highlighted in a pink color, while the rest of the text is black.

3. Visualization makes you Love Stat and Data.

I hope you download and read it!

SASユーザー総会2023

開始	終了	トラック1 福武ラーニングシアター
9:30		受付開始
9:50	10:00	プレオープニング イーピーエス株式会社 森岡 裕(世話人)
10:00	10:30	前処理大全 SASバージョン イーピーエス株式会社 森岡 裕 
10:30	11:00	解析プロシジャで作成されるODS統計解析Plotに ついて イーピーエス株式会社 折村 奈美 

<https://sas-user2023.ywstat.jp/program.html>

Output Delivery SystemとODS Graphics

Output Delivery System(ODS)

プロシジャによって作成された表やグラフの出力を管理し、HTMLやPDFなど様々な形式で出力する機能

ODS Graphics

解析プロシジャにおけるグラフを描く機能

結果データセットの処理に労力を割いて、グラフを有効活用できていない…？

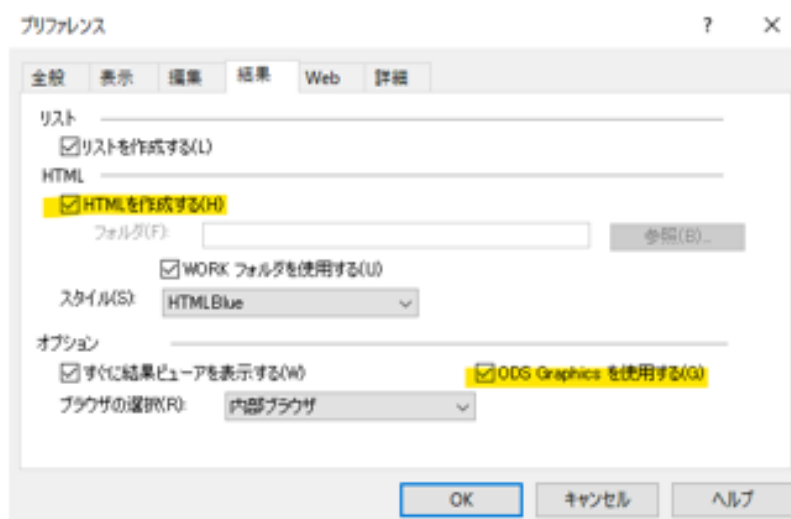
**ODS Graphicsの機能とグラフを紹介し
その有用性を示す**

ODS Graphicsを使用する前に…

ツール→オプション→プリファレンス

HTMLを作成する

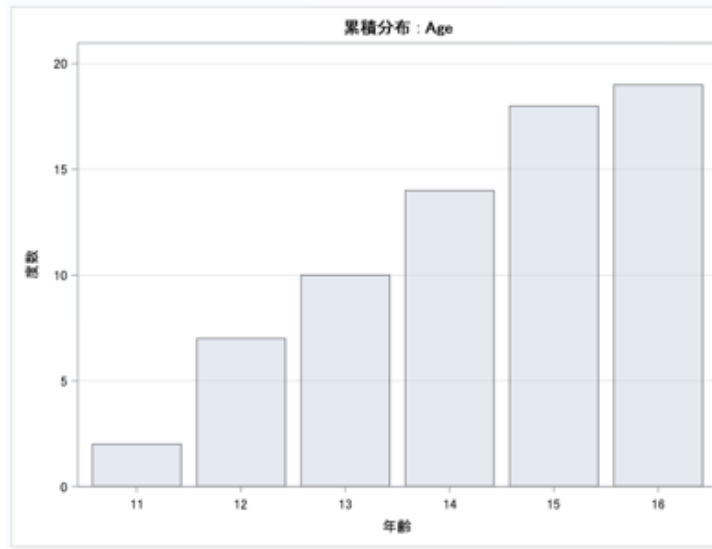
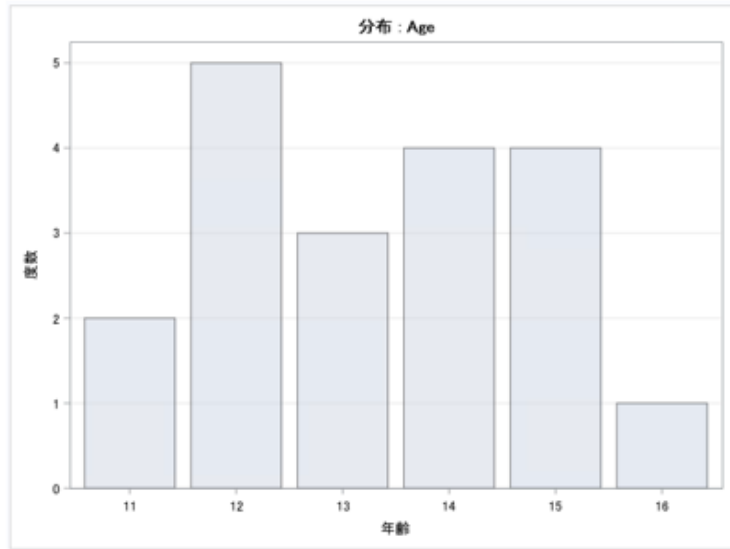
ODS Graphicsを使用するにチェック



FREQプロシジャ - FreqPlot, CumfreqPlot

```
proc freq data=sashelp.class;
  tables age/ plots=(freqplot cumfreqplot);
run;
```

年齢				
Age	度数	パーセント	累積度数	累積パーセント
11	2	10.53	2	10.53
12	5	26.32	7	36.84
13	3	15.79	10	52.63
14	4	21.05	14	73.68
15	4	21.05	18	94.74
16	1	5.26	19	100.00



度数プロットや累積度数プロットを作成し、データの特徴をとらえることができる

FREQプロシジャ - MosaicPlot

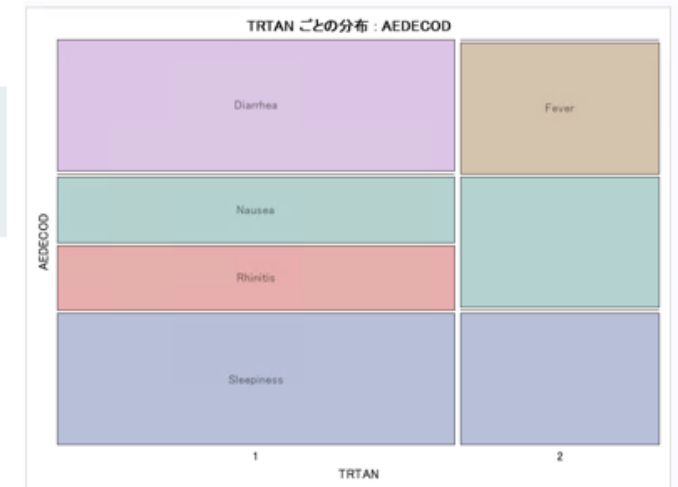
```
data test;
length AEDECOD $200;
  TRTAN=1;
  AEDECOD="Rhinitis"; output;
  AEDECOD="Diarrhea"; output;
  AEDECOD="Nausea"; output;
  AEDECOD="Sleepiness"; output;
  AEDECOD="Diarrhea"; output;
  AEDECOD="Sleepiness"; output;

  TRTAN=2;
  AEDECOD="Sleepiness"; output;
  AEDECOD="Nausea"; output;
  AEDECOD="Fever"; output;
run;
```

```
proc freq data=test;
  tables AEDECOD*TRTAN/ plots=mosaicplot;
run;
```

カテゴリカルデータの群ごとの偏りと群内での比率をビジュアルで把握できる

AEDECOD	表: AEDECOD * TRTAN		
	TRTAN		合計
	1	2	
Diarrhea	2	0	2
	22.22	0.00	22.22
	100.00	0.00	
	33.33	0.00	
Fever	0	1	1
	0.00	11.11	11.11
	0.00	100.00	
	0.00	33.33	
Nausea	1	1	2
	11.11	11.11	22.22
	50.00	50.00	
	16.67	33.33	
Rhinitis	1	0	1
	11.11	0.00	11.11
	100.00	0.00	
	16.67	0.00	
Sleepiness	2	1	3
	22.22	11.11	33.33
	66.67	33.33	
	33.33	33.33	
合計	6	3	9
	66.67	33.33	100.00



Simple and easy option to visualization of the analysis results

FREQプロシジャ - AgreePlot

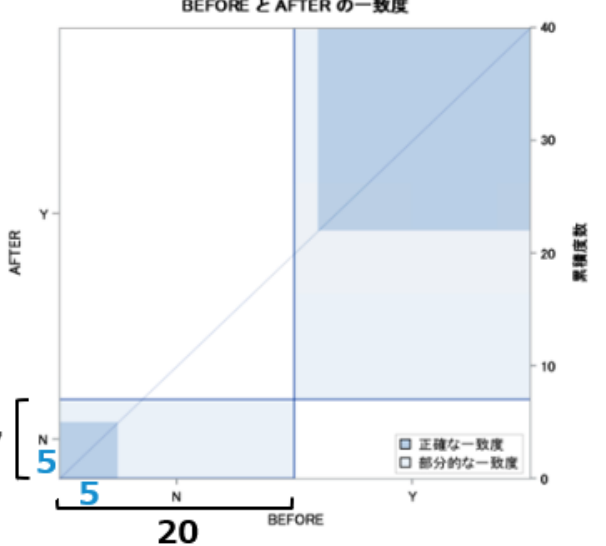
```
data test;
  BEFORE="Y"; AFTER="Y"; _FREQ_=18; output;
  BEFORE="Y"; AFTER="N"; _FREQ_=2; output;
  BEFORE="N"; AFTER="Y"; _FREQ_=15; output;
  BEFORE="N"; AFTER="N"; _FREQ_=5; output;
run;
```

```
proc freq data=test;
  tables BEFORE * AFTER / agree;
  weight _FREQ_ / zeros;
run;
```

BEFORE	AFTER		合計
	N	Y	
N	5	15	20
	12.50	37.50	50.00
	25.00	75.00	
	71.43	45.45	
Y	2	18	20
	5.00	45.00	50.00
	10.00	90.00	
	28.57	54.55	
合計	7	33	40
	17.50	82.50	100.00

McNemar の検定		
カイ 2 乗値	自由度	Pr > ChiSq
9.9412	1	0.0016

推定値	標準誤差	95% 信頼限界
0.1500	0.1188	-0.0628 0.3628



- 薄青と濃青の面積の差が大きいほど BeforeとAfterの一致度が低い
- 45度線と薄青の四角形の交点のずれ = 全体に対する比の偏り具合

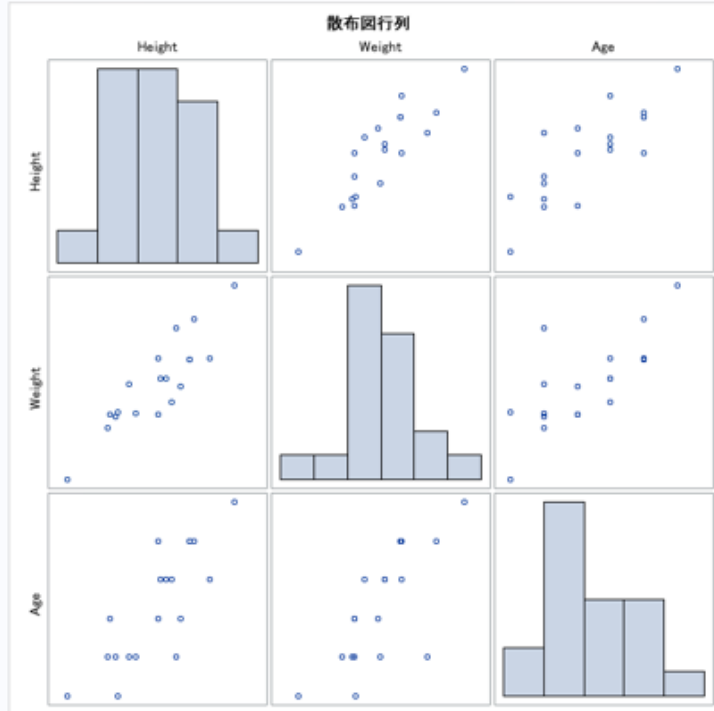
一致度、全体に対する比の偏り具合を可視化でき有用

CORRプロシジャ - 散布図行列

```
proc corr data=sashelp.class nomiss plots=matrix(histogram);
  var Height Weight Age;
run;
```

単純統計量						
変数	N	平均	標準偏差	合計	最小値	最大値
Height	19	62.33684	5.12708	1184	51.30000	72.00000
Weight	19	100.02632	22.77393	1901	50.50000	150.00000
Age	19	13.31579	1.49267	253.00000	11.00000	16.00000

	Height	Weight	Age
Height 身長(インチ)	1.00000	0.87779 <.0001	0.81143 <.0001
Weight 体重(ポンド)	0.87779 <.0001	1.00000	0.74089 0.0003
Age 年齢	0.81143 <.0001	0.74089 0.0003	1.00000

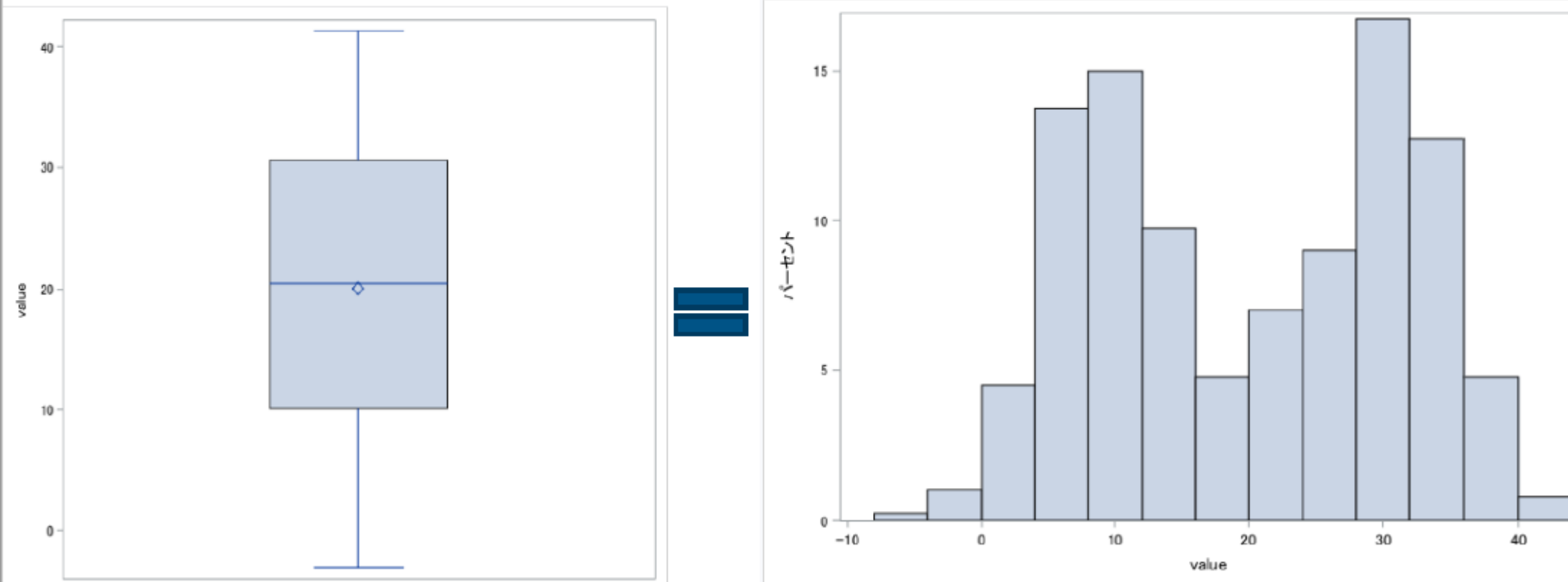


データ間の相関関係を実感的にとらえることができるので 相関のある変数の組み合わせを見つける際に便利である

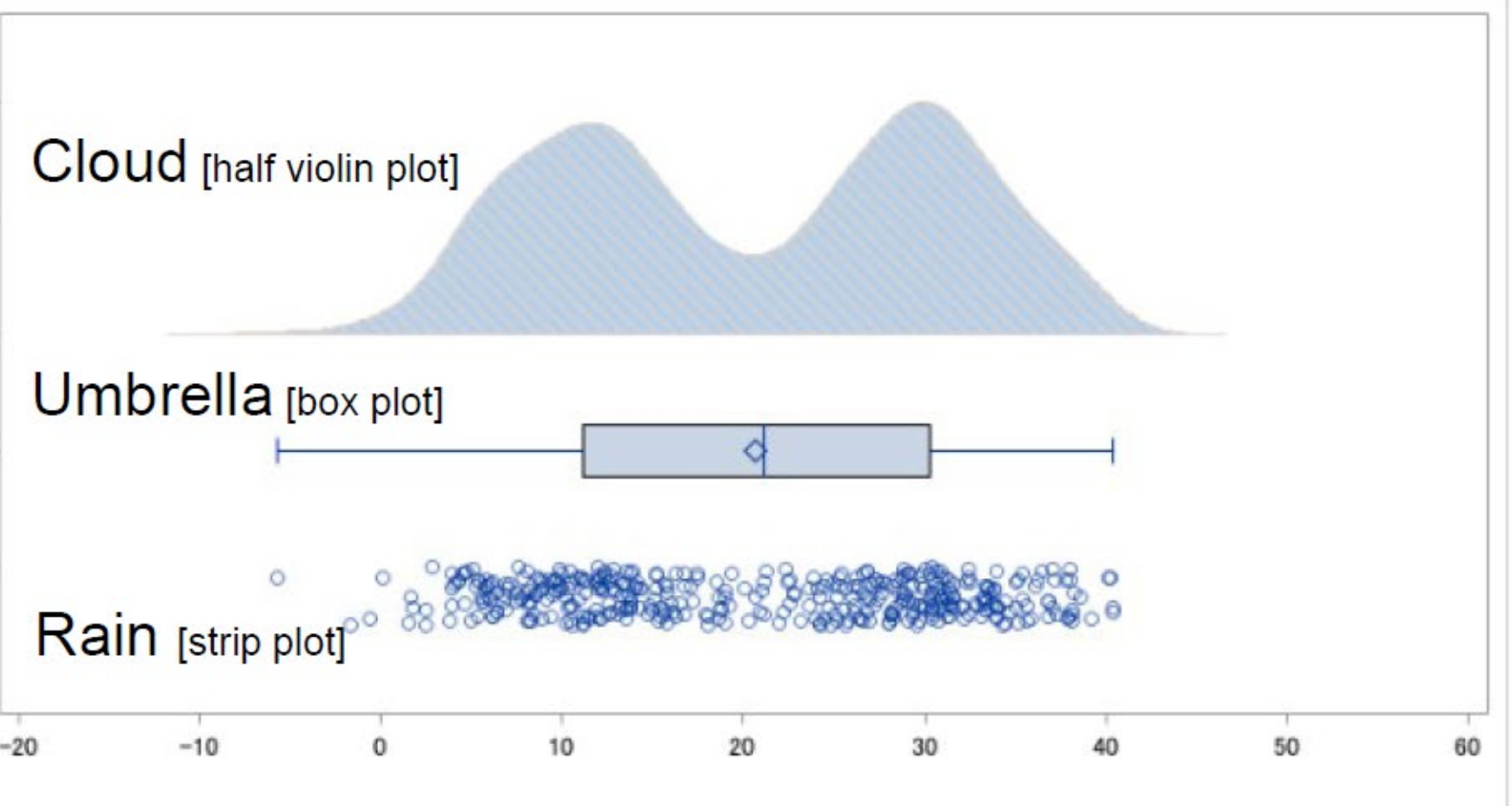
It is possible to get an image of the statistical method from the graphs.

Observe the data carefully.

We cannot always grasp the data distribution from Boxplot .



Raincloud plot



Graph Template

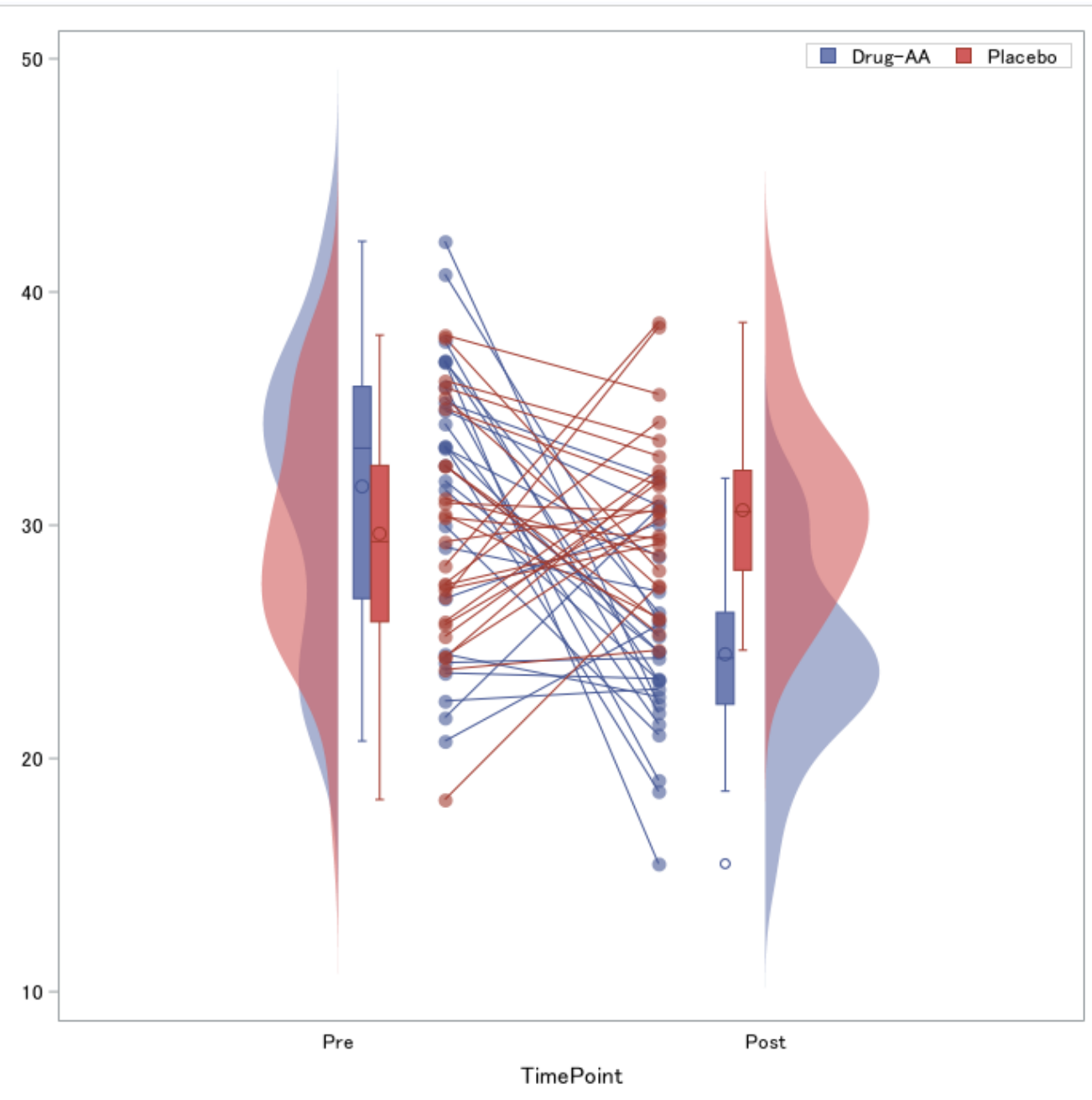
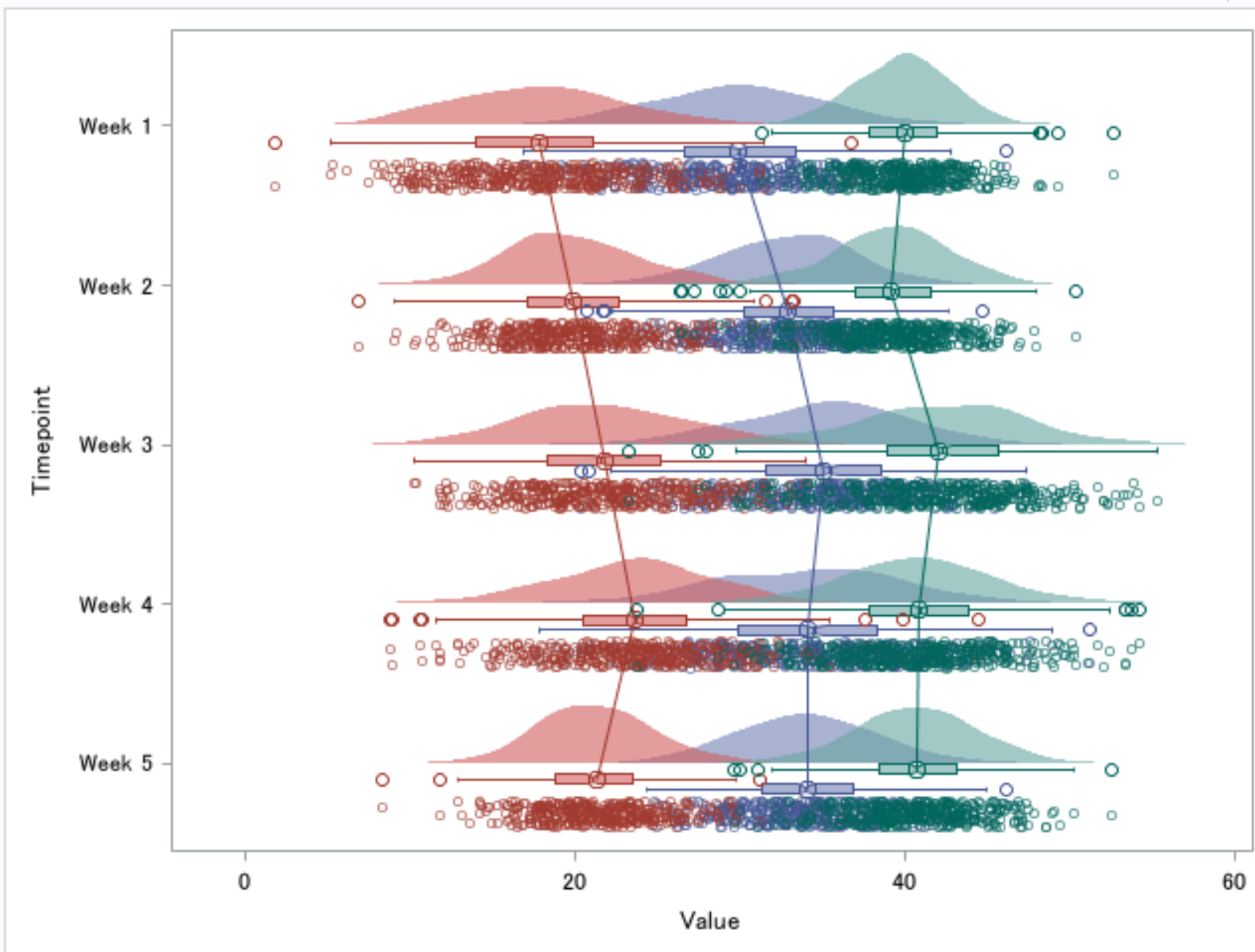
```
ods graphics on / height = 3in width = 12in;
proc template;
  define statgraph RCP;
    begingraph;
      layout overlay
        / xaxisopts = (label = ' ' type= linear
                      linearopts = (viewmin = -20 viewmax = 60
                                     tickvaluesequence = (start = -20 end = 60 increment = 10)))
          yaxisopts = (label = ' ');
        bandplot x = value limitlower = 0 limitupper = density/ display = all;
        boxplot y = val x = eval(-0.02 + coalesce(0, val))
          / orient = horizontal
            boxwidth = 0.3;
        scatterplot x = val
          y = eval(-0.05 + 0.01*cdf('NORMAL', rannor(1234))+coalesce(0, val))
          / markerattrs = (symbol = circle size = 8 transparency = 0.4);
      endlayout;
    endgraph;
  end;
run;
```

Axis definition

Violin plot

Box plot

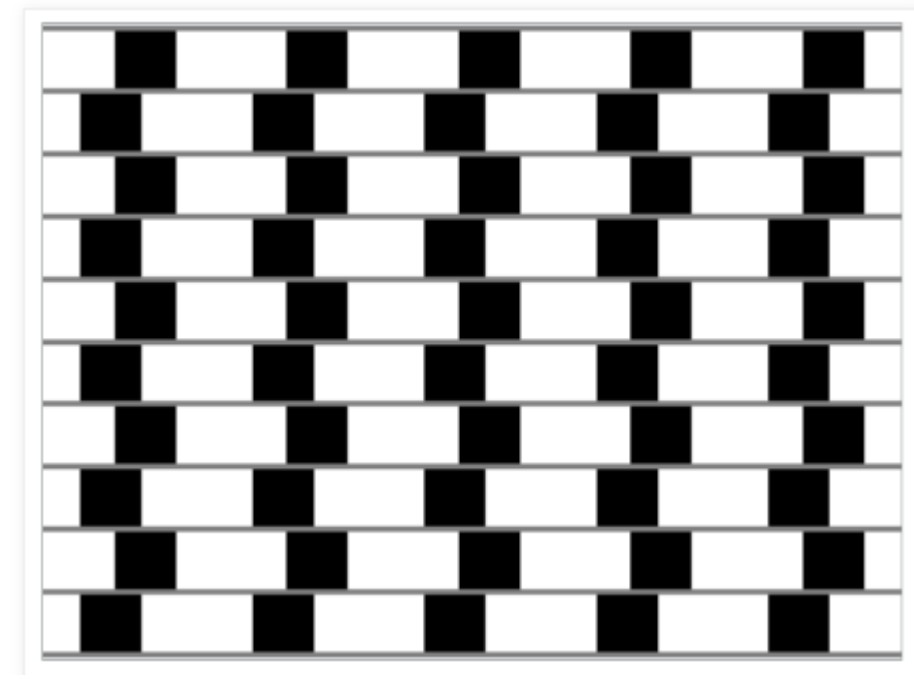
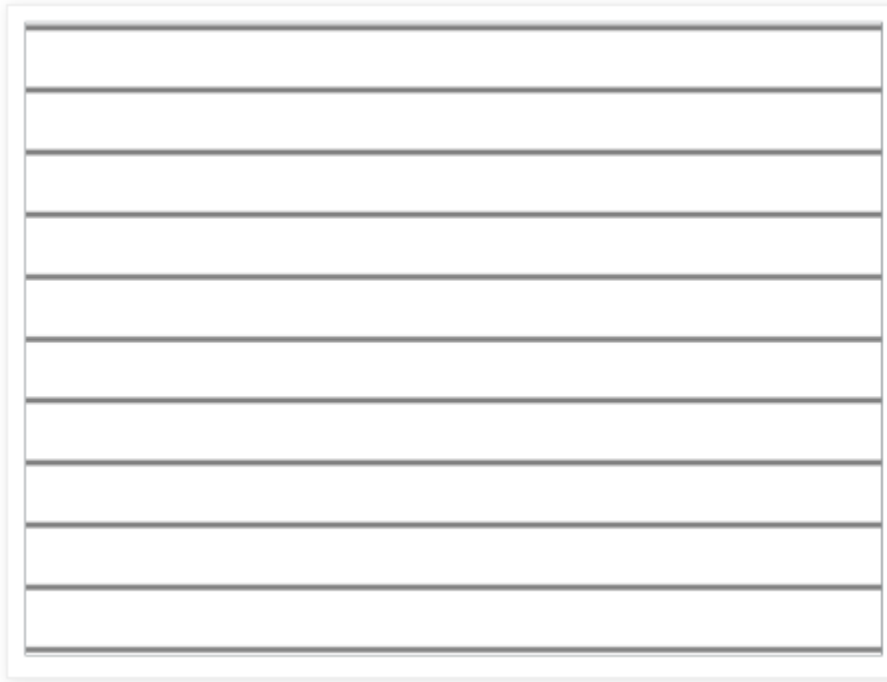
Stripplot (with random jittering)



4. Love programming with games.

optical illusion

```
data wk2;
do x=1.2 to 10.2 by 2;
do y=1.5 to 10.5 by 2;
output;
end;
end;
do x=0.8 to 8.8 by 2;
do y=0.5 to 8.5 by 2;
output;
end;
end;
run;
proc sgplot data=wk2;
/*scatter x=x y=y /markerattrs=(symbol="squarefilled" size=43 color=black);*/
refline 0 1 2 3 4 5 6 7 8 9 10/lineattrs=(thickness=4 color=gray);
axis values=(0 to 10) offsetmin=0 offsetmax=0 display=none;
yaxis values=(0 to 10) offsetmin=0.009 offsetmax=0.009 display=none;
run;
```





新卒1年目作品

First year with the company

```
data pi:
  pi = constant('pi');
  do tau = -pi to pi by pi/100:
    x = cos(tau);
    y = 0.95*sin(tau)-1;
    id=1;
    output:
  end:

  do tau = 0.62*pi to 2.38*pi by pi/100:
    x = cos(tau)+0.8;
    y = 0.95*sin(tau)+0.8-0.8;
    id=2;
    output:
  end:

  do tau = -pi to pi by pi/100:
    x = cos(tau)+0.1;
    y = sin(tau)+0.1+0.9;
    id=3;
    output:
  end:

  do tau = -pi to pi by pi/100:
    x = 0.5*cos(tau)+0.3-0.15;
    y = sin(tau)+0.3+1.5;
    id=4;
    output:
  end:

  do tau = -pi to pi by pi/100:
    x = 0.5*cos(tau)+0.3-0.15;
    y = sin(tau)+0.3+1.5;
    id=5;
    output:
  end:

  do tau = -pi to pi by pi/100:
    x = 0.3*cos(tau)+0.2-0.15;
    y = 0.5*sin(tau)+0.2+1.35;
    id=6;
    output:
  end:

  do tau = -pi to pi by pi/100:
    x = 0.3*cos(tau)+0.2-0.15;
    y = 0.5*sin(tau)+0.2+1.35;
    id=7;
    output:
  end:

  do tau = pi to 2*pi by pi/100:
    x = cos(tau)+0.3;
    y = 0.8*sin(tau)+0.3+0.5;
    id=8;
    output:
  end:




x=-0.3; y=0.8; id=9; output: x=-0.85; y=0.8; id=9; output:
x=0.3; y=0.8; id=10; output: x=-0.85; y=0.8; id=10; output:
x=0.3; y=0.7; id=11; output: x=-0.85; y=0.8; id=11; output:
x=-0.3; y=0.7; id=12; output: x=-0.85; y=0.8; id=12; output:
x=-0.3; y=0.9; id=13; output: x=-0.85; y=1.0; id=13; output:
x=0.3; y=0.9; id=14; output: x=-0.85; y=1.0; id=14; output:
x=0.3; y=0.5; id=15; output: x=-0.3; y=0.5; id=15; output:
x=0; y=0.6; id=16; output: x=0; y=0.5; id=16; output:

run:

proc eplot data=pi noborder noautolegend aspect=1;
  polygon x=x y=y id=id / colorresponse=id fill colormodel=(skyblue white red white white black black red black black black black black black);
  series x=x y=y / group=id colorresponse=id colormodel=(black black black black black black black black black black black black black black);
  xaxis display=none;
  yaxis display=none;
run;
```

Competitive Programming Competition in EPS



第二回SASコーディング杯  チャット  ファイル 

ちょっとシミュレーションの組み方の勉強をかねた練習問題ですが、グリコじゃんけんをします

グーで勝利すると「グリコ」で3点
パーで勝利すると「パイナップル」で6点
チョキで勝利すると「チョコレート」で5点
アイコは互いに0点で次の勝負へ

5回ワンセットで合計得点の高い方が勝ちという勝負を100セット試行します
今、完全に1/3で手をきめるマクロを作って自己対戦させると

```
/*選手マクロ*/  
%macro morioka(type);  
data _null_;  
  if rand("uniform") < 1/3 then te="G";  
  else if rand("uniform") < 2/3 then te="C";  
  else te="P";  
  call symputx("t&type.",te,"G");  
run;  
%mend;
```

Champion Algorithm

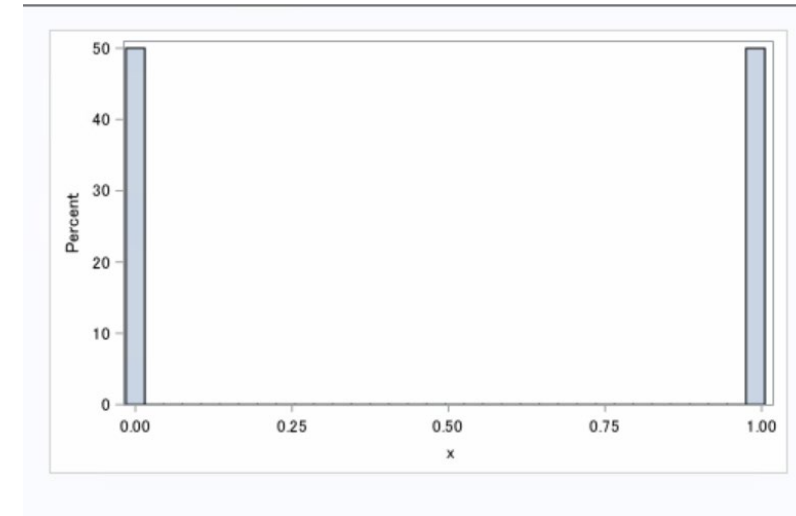
【仕様】

- ・デフォルトの手の割合はナッシュ均衡となる6:5:3
- ・負け続けても最終的に勝てる場合チョキだけを出す
- ・パーで勝ち続けるしか勝ち筋がない場合パーだけを出す
- ・グーで勝っても意味がない場合パーとチョキを等確率で出す
- ・5手目で手の優劣がない場合(点差-2~+2)G,C,Pを等確率で出す
- ・平場（1-3手目）での直近10回の相手の手に偏りがある場合、または直近10回の1or2or3手目での相手の手に偏りがある場合、こちらも手を偏らせる
8割以上→(80%,10%,10%),10割→(90%,5%,5%)

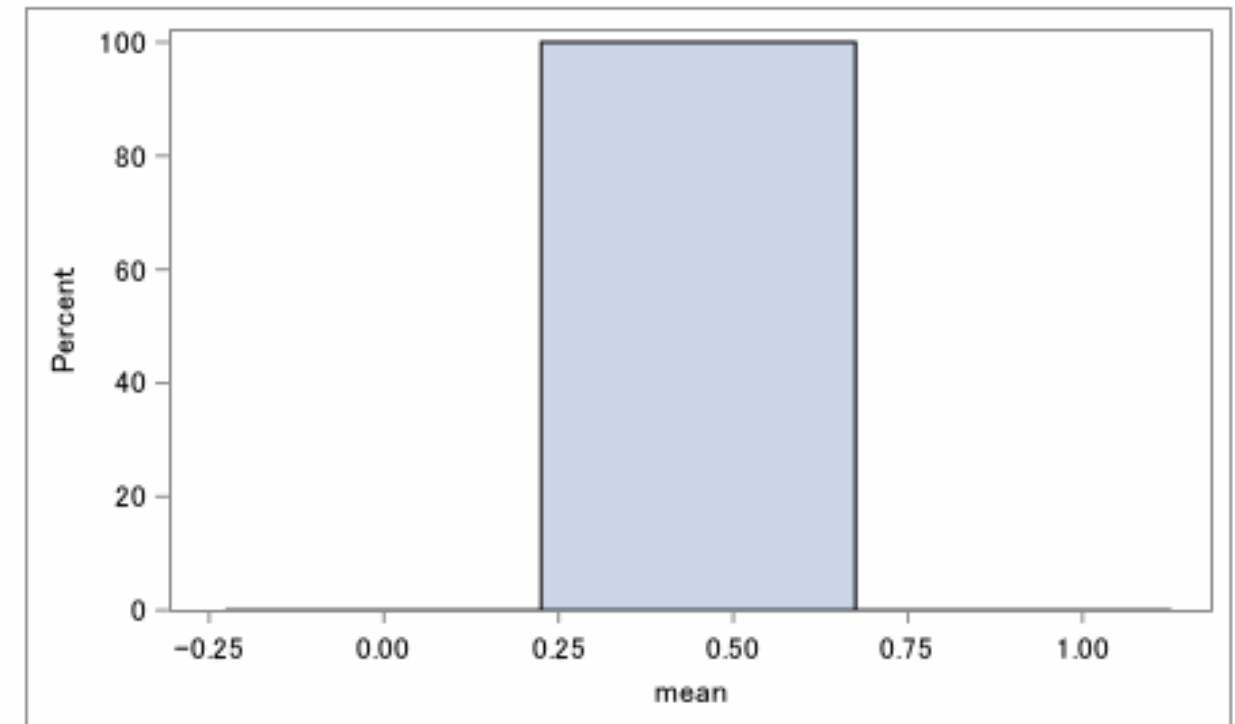
チャンピオンデータですが勝ち-負け=20ぐらいは何回か回したら出ました。

central limit theorem GIF Animation

```
data a;  
do i=1 to 10000;  
    call streaminit(100);  
    x=rand("BINOMIAL",0.5,1);  
    output;  
end;  
run;  
proc sgplot data=a;  
    histogram x ;  
run;  
options papersize=('5 in', '3 in') printerpath=gif animation=start  
    animduration=0.5 animloop=yes noanimoverlay;  
ods printer file='XXXXXXXXXXXXXXXXXXtest%aa.gif';  
ods graphics / width=5in height=3in imagefmt=GIF;  
%macro loop;  
%do i=1 %to 1001 %by 10;  
    proc surveyselect data=a method=srs rep=&i  
        n=20 seed=12345 out=out noprint;  
run;  
proc summary data=out nway;  
    class Replicate;  
    var x;  
    output out=ds_mean mean=mean;  
run;  
proc sgplot data=ds_mean;  
    histogram mean ;  
run;  
%end;  
%mend;  
%loop  
options printerpath=gif animation=stop;  
ods printer close;
```

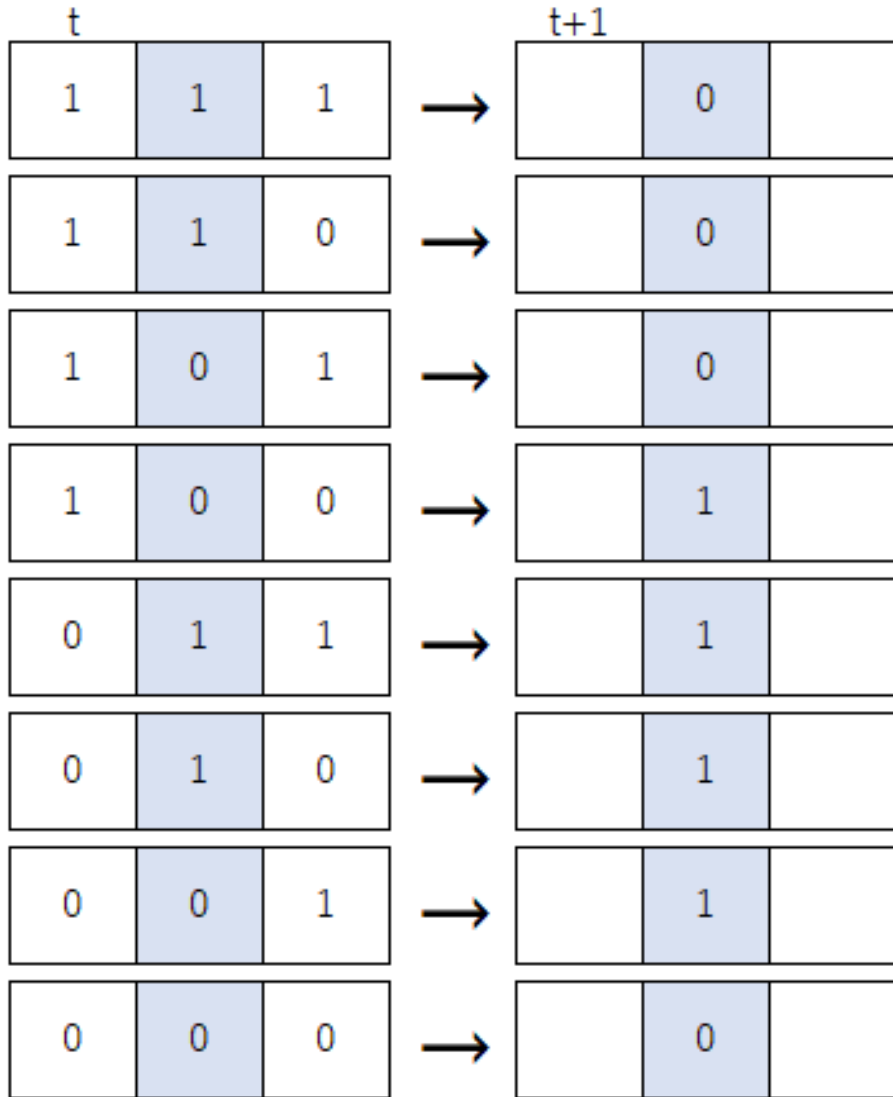


2023年 5月16日 火曜日 14時31分50秒 1



Elementary cellular automata

Rule30



```

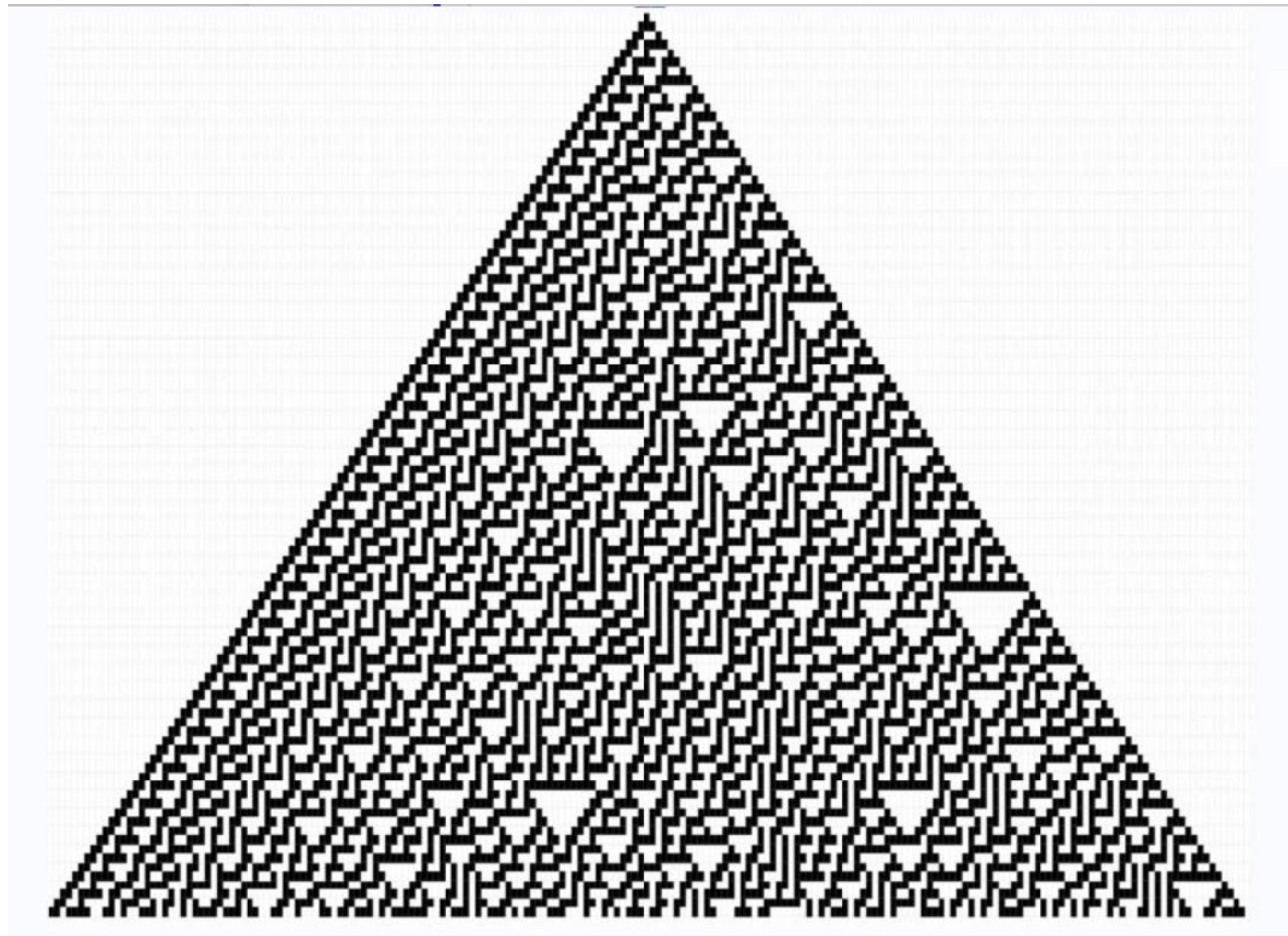
data wk1;
array col{200} (200*0);
array ncol{200} (200*0);
col{100}=1;
do obs = 1 to 100;
  do i = 1 to 200;
    if obs = 1 then do;
      ncol{i}=col{i} ;
    end;
  else do;
    if i=1 then pattern=cats(0, col{i},col{i+1});
    if (1+1)<=i<=(200-1) then pattern=cats(col{i-1},col{i},col{i+1});
    if i=200 then pattern=cats(col{i-1}, col{i},0);
    /*rule*/
    select (pattern);
      when("000") ncol{i} =0;
      when("001") ncol{i} =1;
      when("010") ncol{i} =1;
      when("011") ncol{i} =1;
      when("100") ncol{i} =1;
      when("101") ncol{i} =0;
      when("110") ncol{i} =0;
      when("111") ncol{i} =0;
    end;
  end;
end;
output;
do i = 1 to 200;
  col{i}=ncol{i} ;
end;
end;
keep ncol;;
run;

```

```

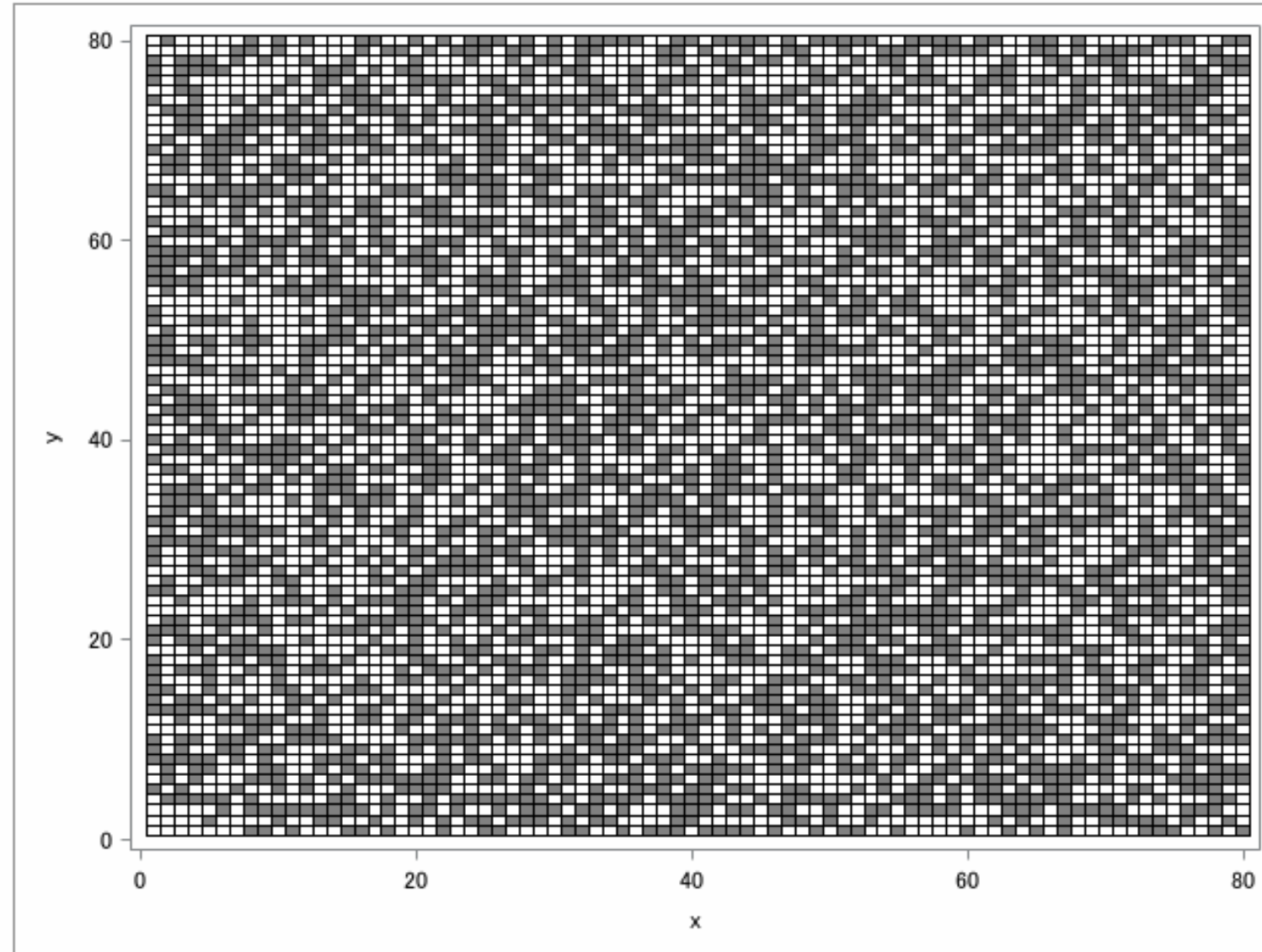
data _null_;
set wk1 end=eof;
array AR ncol;;
if _N_=1 then do;
  dcl odsout ob();
  ob.table_start();
end;
ob.row_start();
do over AR;
  if AR=1 then do; background="black";color="black";end;
  else if AR=0 then do; background="white";color="white";end;
  text=catx(" ", " color=",color, " height=0.01 width=0.01 vjust=center
background=",background);
  ob.format_cell(data:AR,style_attr:text);
  call missing(of background color);
end;
ob.row_end();
if eof then do;
  ob.table_end();
end;
run;

```

Conway's Game of Life

1



Develop skills in managing coordinates like Arrays and Hash objects.

I want people to learn algorithms and visualization while playing with them, like cellular automata, etc.

5. You will **Love** the results of your own manual analysis. No Procedure! No Library!.

t-test

```
proc ttest data=sashelp.class;  
class sex;  
var weight;  
run;
```

Sex	手法	平均	平均の 95% 信頼限界		標準偏差	標準偏差の 95% 信頼限界	
女子		90.1111	75.2113	105.0	19.3839	13.0930	37.1351
男子		109.0	92.6920	125.2	22.7272	15.6326	41.4910
Diff (1-2)	Pooled	-18.8389	-39.4090	1.7313	21.2196	15.9229	31.8112
Diff (1-2)	Satterthwaite	-18.8389	-39.2325	1.5547			

手法	分散	自由度	t 値	Pr > t
Pooled	Equal	17	-1.93	0.0702
Satterthwaite	Unequal	16.962	-1.95	0.0680

等分散性				
手法	分子の自由度	分母の自由度	F 値	Pr > F
Folded F	9	8	1.37	0.6645

```

proc summary data=sashelp.class nway ;
class sex;
var weight;
output out=wk1 n= mean= var= /autoname;
run;
proc transpose data=wk1 out=mean(drop=_) prefix=mean;
var weight_mean;
run;
proc transpose data=wk1 out=var(drop=_) prefix=var;
var weight_var;
run;
proc transpose data=wk1 out=n(drop=_) prefix=n;
var weight_n;
run;
data ans;
set n;
set var;
set mean;
df = n1 + n2 - 2;
sp =sqrt( ( (n1 - 1) * var1 + (n2 - 1) * var2 ) / df );
t = ( mean1 - mean2 ) / ( sp * sqrt( (1 / n1 + 1 / n2 ) ));
if t < 0 then p = 2 * probt(t , df) ;
else p = 2 * (1 - probt(t , df));
put n1= mean1= ;
put n2= mean2= / ;
put df= t= p=;
run;

```

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

```

598 data ans;
599 set n;
600 set var;
601 set mean;
602
603 df = n1 + n2 - 2;
604 sp =sqrt( ( (n1 - 1) * var1 + (n2 - 1) * var2 ) / df );
605 t = ( mean1 - mean2 ) / ( sp * sqrt( (1 / n1 + 1 / n2 ) ));
606
607 if t < 0 then p = 2 * probt(t , df) ;
608 else p = 2 * (1 - probt(t , df));
609
610 put n1= mean1= ;
611 put n2= mean2= / ;
612 put df= t= p=;
613
614 run;

n1=9 mean1=90.111111111
n2=10 mean2=108.95
df=17 t=-1.932246615 p=0.0701687815

```


Pearson correlation coefficient

```
proc corr data=sashelp.class;  
var weight height;  
run;
```

SAS システム
CORR プロシジャ

2 変数 : Weight Height

単純統計量							
変数	N	平均	標準偏差	合計	最小値	最大値	ラベル
Weight	19	100.02632	22.77393	1901	50.50000	150.00000	体重(ポンド)
Height	19	62.33684	5.12708	1184	51.30000	72.00000	身長(インチ)

Pearson の相関係数, N = 19
H0: Rho=0 に対する Prob > |r|

	Weight	Height
Weight 体重(ポンド)	1.00000	0.87779 <.0001
Height 身長(インチ)	0.87779 <.0001	1.00000

```

data wk2;
set sashelp.class;
if _N_=1 then set wk1;
weight_deviation = weight -weight_mean;
height_deviation = height -height_mean;

multi_deviation = weight_deviation * height_deviation;

```

```

weight_deviation_sq = weight_deviation**2;
height_deviation_sq = height_deviation**2;
run;

```

それぞれの平均ともとめて、共分散と身長・体重の分散が求まる

```

proc summary data=wk2;
var multi_deviation weight_deviation_sq height_deviation_sq;
output out = wk3 mean=covariance variance_weight variance_height;;
run;

```

最後は分子にルートかけてから、掛け算して、分子には共分散をおく

```

data output;
set wk1;
set wk3;
corr =covariance / ( sqrt(variance_weight) * sqrt(variance_height));
run;

```

確認してみる

```

proc print data=output noobs;
var corr;
run;

```

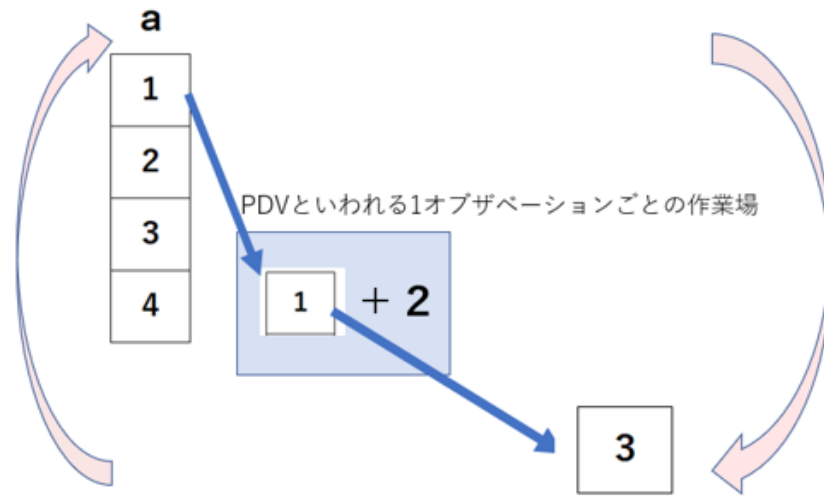
$$r_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

SAS システム	
corr	0.87779



an extra
オマケ

SASとRの違いって



a	1	2	3	4
	+	+	+	+
	2	2	2	2
	3	4	5	6

私見ですが

SASとRの違いの根本はここだと思ってます

SASは、観測に対して、処理を暗黙にループすることで捌いており

Rはベクトルの処理に帰着させることで捌いている

Python & SAS

Python パンダステップ100万回 - pythonicに天衣無縫 投稿 ファイル Wiki NIさん資料③ NIさん資料② NIさん資料① pip install

返信

森岡 裕(038) 2020/06/02 12:45 編集済み

6/3 DataFrame①

DataFrameは2次元データです

Python

```
1 import pandas as pd
2 df=pd.DataFrame([[10,"a",True],
3                  [20,"b",True],
4                  [30,"c",False],
5                  [40,"c",False].
```

展開 (9 行)

とすると
以下のように出力されます。

	0	1	2
0	10	a	True
1	20	b	True
2	30	c	False
3	40	c	False
4	50	c	False
5	60	c	False

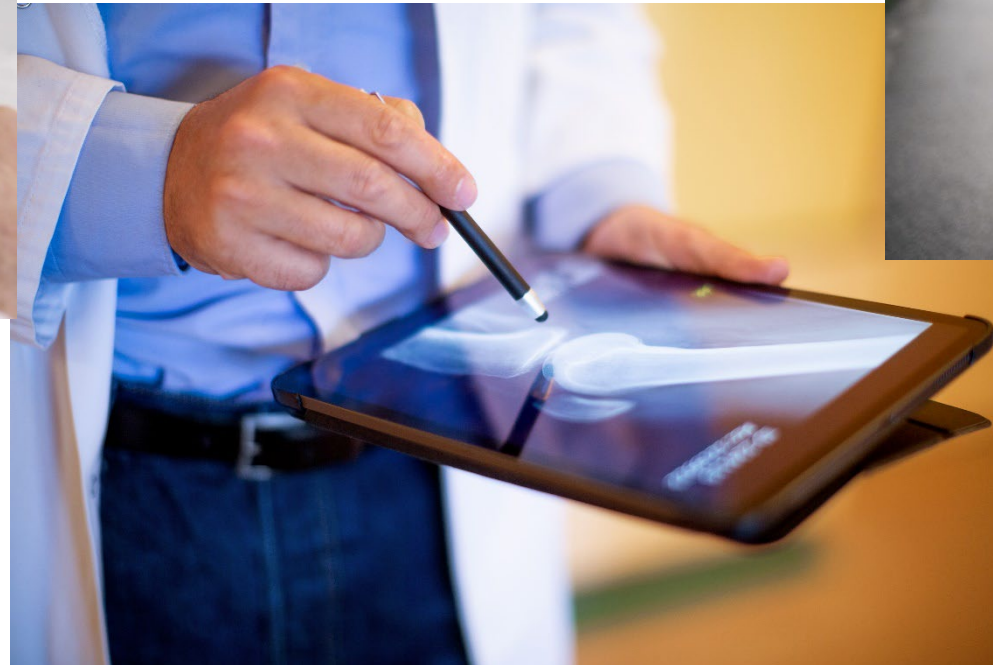
ここで、注意なのが、今

Out[19]:

	0	1	2
0	10	a	True
1	20	b	True
2	30	c	False
3	40	c	False

EPUB

EPUB is an e-book file format



Remote work is increasing, and the days of handing out paper training materials are gone.

EPUB can be viewed on mobile devices and can be created with embedded images, videos, and program code.

カイ二乗検定について

カイ二乗検定統計量:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

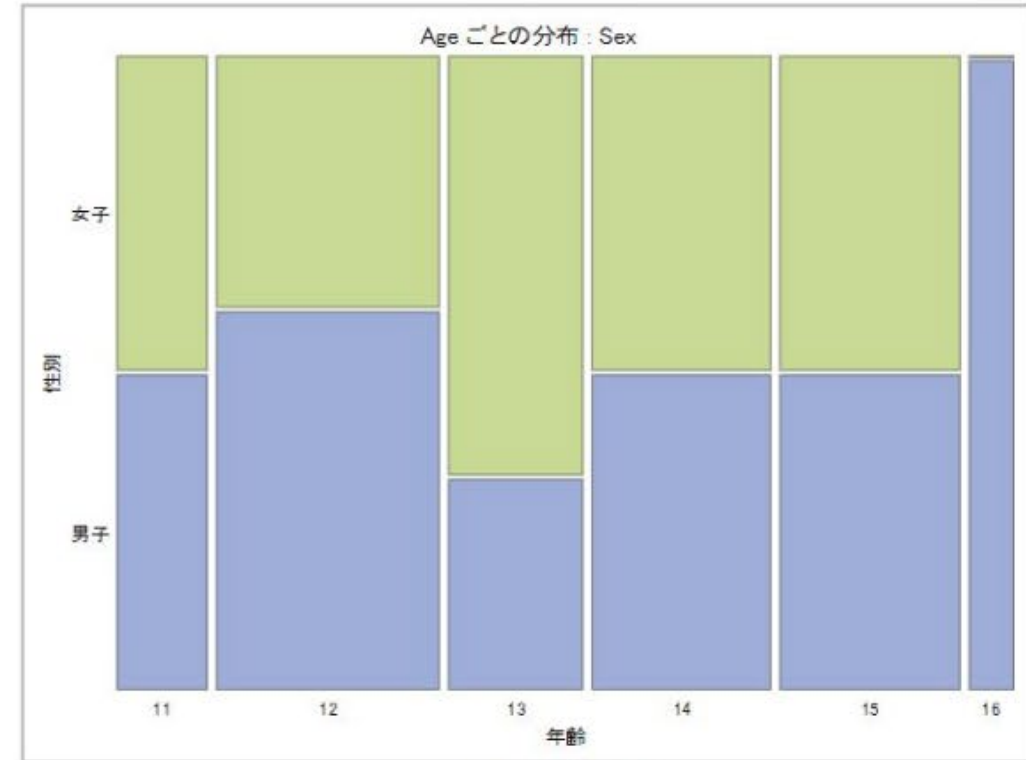
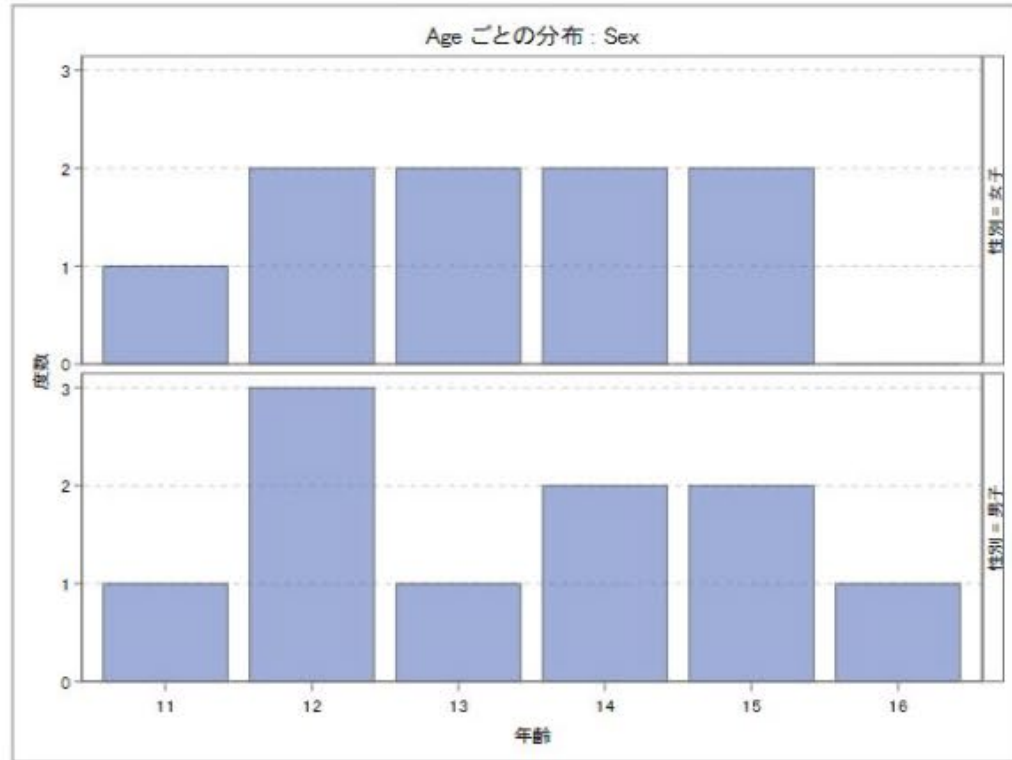
ここでO = 頻度の観測値, E = 帰無仮説の下における頻度の期待値(理論値)である:

SASコード:

```
PROC FREQ DATA=SASHELP.CLASS ;
TABLES SEX*AGE / CHISQ PLOT=ALL;
RUN;
```

度数
パーセント
行のパーセント
列のパーセント

表: Sex * Age							
Sex(性別)	Age(年齢)						合計
	11	12	13	14	15	16	
女子	1	2	2	2	2	0	9
	5.26	10.53	10.53	10.53	10.53	0.00	47.37
	11.11	22.22	22.22	22.22	22.22	0.00	
	50.00	40.00	66.67	50.00	50.00	0.00	
男子	1	3	1	2	2	1	10
	5.26	15.79	5.26	10.53	10.53	5.26	52.63
	10.00	30.00	10.00	20.00	20.00	10.00	
	50.00	60.00	33.33	50.00	50.00	100.00	
合計	2	5	3	4	4	1	19
	10.53	26.32	15.79	21.05	21.05	5.26	100.00



Sex * Age の統計量

統計量	自由度	値	p 値
カイ 2 乗値	5	1.4848	0.9148
尤度比カイ 2 乗値	5	1.8748	0.8662
Mantel-Haenszel のカイ 2 乗値	1	0.0672	0.7955
ファイ係数		0.2795	
一致係数		0.2692	
Cramer の V 統計量		0.2795	

ODS EPUB and MathML

```
1 title "カイ二乗検定について";
2 ods epub file="D:\Users\10089669\Desktop\a\test.epub" title="MathML Test";
3
4 ods escapechar='^';
5
6 proc odstext contents="";
7 p 'カイ二乗検定統計量:';
8 p '^{\mathml <math xmlns="http://www.w3.org/1998/Math/MathML">
9   <msup>
10    <mi>\chi</mi>
11    <mn>2</mn>
12  </msup>
13 <mo>=</mo>
14 <mo>&#x02211;</mo>
15 <mfrac>
16   <msup>
17   <mrow>
18     <mn>O</mn>
19     <mo>-</mo>
20     <mn>E</mn>
21   </mrow>
22   <mn>2</mn>
23 </msup>
24 <mn>E</mn>
25 </mfrac>
26 </math>}';
27 p 'ここでO = 頻度の観測値, E = 帰無仮説の下における頻度の期待値 (理論値) である:';
28
29 run;
30
```

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

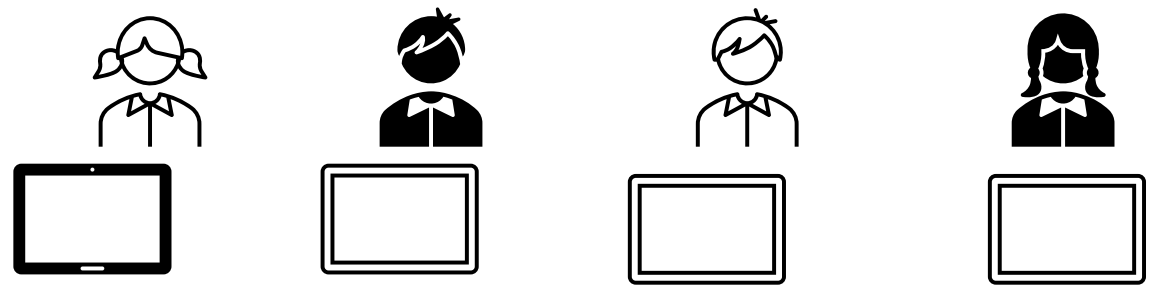
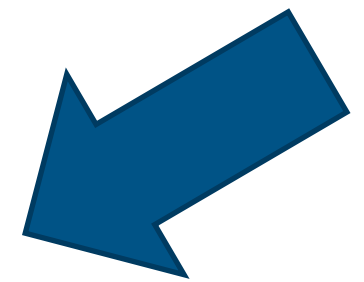
Keeps beginner techniques and training fast and up-to-date

実行コード

```
proc freq data=sashelp.class;  
tables age*sex;  
run;
```

FREQ フォント

Age	F	M	合計
11	1	1	2
12	2	2	4
13	3	3	6
14	4	4	8
15	5	5	10
合計	15	15	30



Thank you!!

2015年 8月13日 木曜日 03時18分14秒 1

