

Making a Data Driven Macro to Enhance Programming Efficiency

Ben Wang, MPH

Biomarin

BASAS 2019

About me

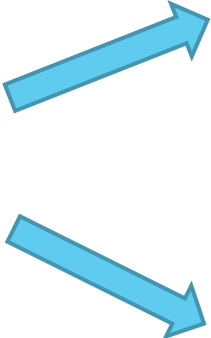
My name is Ben Wang. I am working at Biomarin as the lead Medical Affairs statistical programmer. I graduated from Carnegie Mellon with a BS in statistics and Rutgers Medical School with a Masters of Public Health in Epidemiology. Prior to working at Biomarin, I worked as a statistician at Roche Molecular Diagnostics. I have eight plus years of pharmaceutical experience.

Background

Parameter	Statistics
Continous Variable	
N	XXX
Mean (SD)	XX.XXX (XX.XXX)
Median	XX
q1, q3	XX, XX
min, max	XX, XX
Categorical N(%)	
Cat. A	XX (XX.XX)
Cat. B, etc.	XX (XX.XX)

Ideology

Metadata

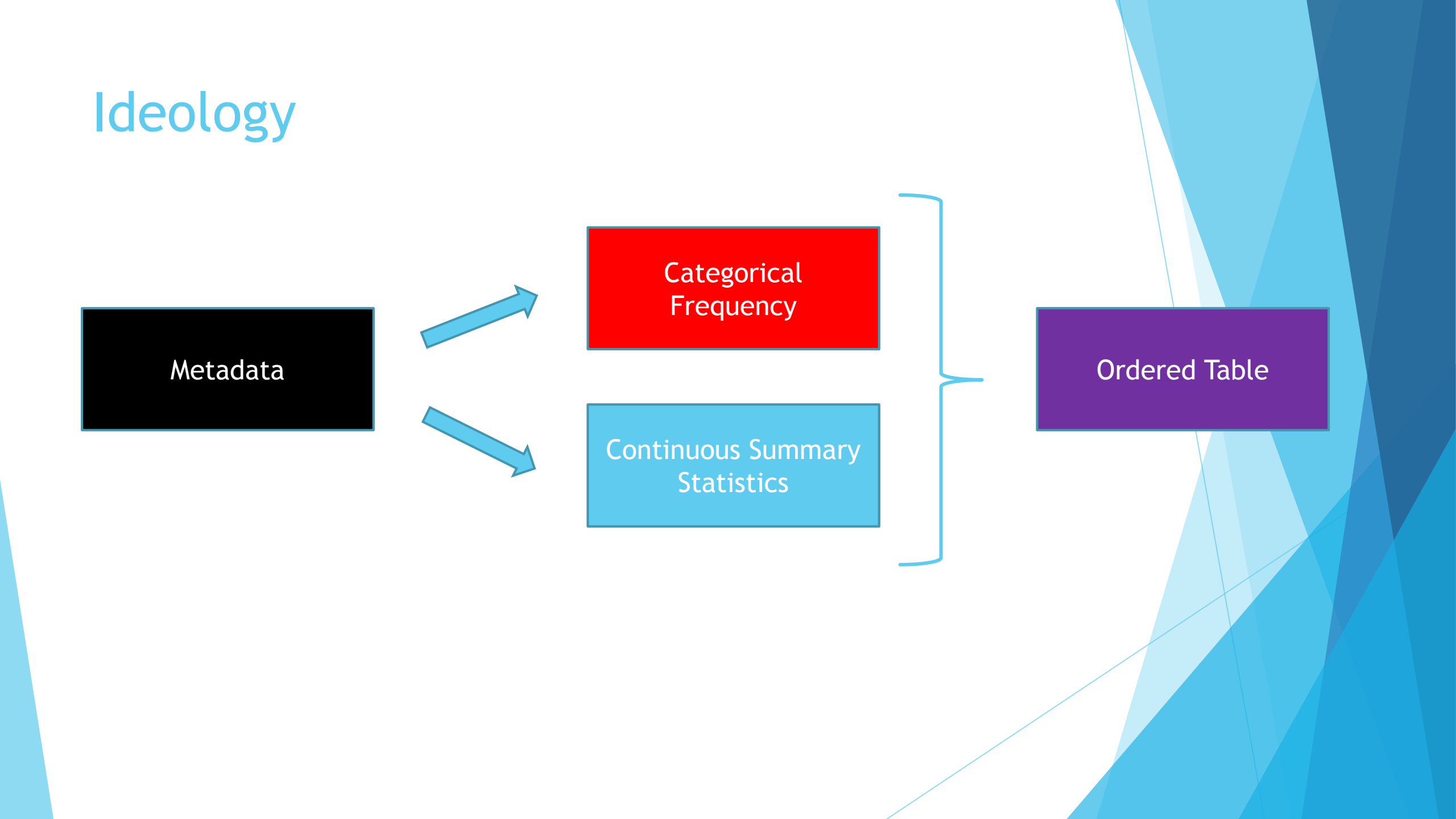


Categorical Frequency

Continuous Summary Statistics



Ordered Table



Methodology

- ▶ Examine Mock Table
- ▶ Obtain Metadata
- ▶ Handling categorical type variables
- ▶ Handling continuous type variables
- ▶ Let the metadata drive the analyses
- ▶ Output table

Example Mock Table

Parameter	Statistics
Age	
N	XXX
Mean (SD)	XX.XXX (XX.XXX)
Median	XX
q1, q3	XX, XX
min, max	XX, XX
Sex N(%)	
F	XX (XX.XX)
M	XX (XX.XX)

Obtain Metadata

Metadata

- ▶ Using sashelp.class

```
PROC CONTENTS Data=sashelp.class out=sum_class;  
run;
```

LIBNAME	MEMNAME	NAME	TYPE	LENGTH	VARNUM	LABEL
SASHELP	CLASS	AGE	1	8	3	
SASHELP	CLASS	HEIGHT	1	8	4	
SASHELP	CLASS	NAME	2	8	1	
SASHELP	CLASS	SEX	2	1	2	
SASHELP	CLASS	WEIGHT	1	8	5	

Obtain Metadata (p2)

Metadata

```
PROC SORT Data=sashelp.class out=class0; by name; run;
DATA class;
  set class0;
  label NAME="First Name" SEX="Sex" AGE="Age"
  Height="Height (cm)" WEIGHT="Weight (kg)";
RUN;
PROC CONTENTS Data=class out=drive0(keep=MEMNAME
NAME TYPE LABEL) NOPRINT; RUN;
```

MEMNAME	NAME	TYPE	LABEL
CLASS	AGE	1	Age
CLASS	HEIGHT	1	Height (cm)
CLASS	NAME	2	First Name
CLASS	SEX	2	Sex
CLASS	WEIGHT	1	Weight (kg)

Handling Categorical Variables

```
PROC FREQ data=class NLEVELS;
  tables sex/ out=freq;
RUN;

DATA freq2;
  set freq;
  where cmiss(sex)=0;
  pct=round(percent, 0.01);
  statc=catt( count, '(',pct,')');
  ord=_n_;
  keep sex statc ord;
RUN;

DATA name;
  set drive0;
  where name=upcase("sex");
  ord=0;
RUN;

DATA freq_sex;
  length param $200. var $20.;
  merge name freq2;
  by ord;
  if ord=0 then param=cats(label)||'
N(%)';
  else param=' ||cats(sex);
  var='SEX';
  keep param statc;
RUN;
```

Categorical
Frequency

Sex N(%)	
F	XX (XX.XX)
M	XX (XX.XX)

PARAM	STATC
Sex N(%)	
F	9 (47.37)
M	10 (52.63)

Handling Categorical Variables (p2)

```
%MACRO freq_ds(dsin=, var=, driver=);  
PROC FREQ data=&dsin. NLEVELS NOPRINT;  
    tables &var./ out=freq;  
RUN;  
DATA freq2;  
    set freq;  
    where cmiss(sex)=0;  
    pct=round(percent, 0.01);  
    statc=catt( count, ' (,pct,')');  
    ord=_n_;  
    keep &var stat ord;  
RUN;  
DATA name;  
    set &driver.;  
    where name=upcase("&var.");  
    ord=0;  
RUN;  
DATA freq_&var.;  
    length param $200. Var $20.;  
    merge name freq2;  
    by ord;  
    if ord=0 then param=cats(label)||' N(%)';  
    else param=' ||cats(&var.);  
    var="&var.";   
RUN;  
%MEND freq_ds;
```

Categorical
Frequency

Handling Continuous Variables

Continuous Summary Statistics

```
PROC FORMAT;
  value mnfmt
  1=' N'
  2=' Mean (SD)'
  3=' Median'
  4=' q1, q3'
  5=' min, max';
```

Age	
N	XXX
Mean (SD)	XX.XXX (XX.XXX)
Median	XX
q1, q3	XX, XX
min, max	XX, XX

```
RUN;
PROC SUMMARY data=class;
  var age;
  OUTPUT OUT=outstat
  N=n MEAN=mean STD=std MEDIAN=median Q1=q1
  Q3=q3 MIN=min MAX=max;
RUN;
```

```
DATA mean1;
  set outstat;
  r_mean=round(mean, 0.001);
  r_std=round(std, 0.001);

  nc=cats(n);
  msd=catt(r_mean, '(', r_std, ')');
  med=cats(median);
  quart=catx(',', 'q1, q3');
  range=catx(',', 'q1, q3');
  keep nc msd med quart range;
RUN;
```

TYPE	_FREQ_	N	MEAN	STD	MEDIAN	Q1	Q3	MIN	MAX
0	19	19	13.315789474	1.4926721594	13	12	15	11	16

Handling Continuous Variables (p2)

```
PROC TRANSPOSE data=mean1  
out=mean1_t;  
var nc msd med quart range;  
RUN;
```

```
DATA mean2;  
set mean1_t;  
ord=_n_;  
statc=col1;  
run;
```

```
DATA name;  
set drive0;  
where NAME=upcase("age");  
ord=0;  
RUN;
```

```
DATA mean_age;  
length param $200. var $20.;  
merge name mean2;  
by ord;  
if ord=0 then param=cats(label);  
else param=put(ord, mnfmt.);  
var='AGE';  
keep param statc;  
RUN;
```

Continuous Summary Statistics

Age	
N	XXX
Mean (SD)	XX.XXX (XX.XXX)
Median	XX
q1, q3	XX, XX
min, max	XX, XX

PARAM	STATC
Age	
N	19
Mean (SD)	13.316 (1.493)
Median	13
q1, q3	12, 15
min, max	12, 15

Handling Continuous Variables (p3)

```
%MACRO mean_ds(dsin=, var=,
driver=);

PROC SUMMARY data=adsl;
  var &var.;
  OUTPUT OUT=outstat
  N=n MEAN=mean STD=std
  MEDIAN=median Q1=q1 Q3=q3
  MIN=min MAX=max;
RUN;

DATA mean1;
  set outstat;
  r_mean=round(mean, 0.001);
  r_std=round(std, 0.001);

  nc=cats(n);
  msd=catt(r_mean,
(‘,r_std,’));
  med=cats(median);
  quart=catx(‘, ‘,q1, q3);
  range=catx(‘, ‘,q1, q3);
  keep nc msd med quart range;
RUN;
```

```
PROC TRANSPOSE data=mean1 out=mean1_t;
  var nc msd med quart range;
RUN;

DATA mean2;
  set mean1_t;
  ord=_n_;
  statc=col1;
run;

DATA name;
  set &driver.;
  where NAME=upcase("&var.");
  ord=0;
RUN;

DATA mean_&var.;
  length param $200. var $20.;
  merge name mean2;
  by ord;
  if ord=0 then param=cats(label);
  else param=put(ord, mnfmt.);
  var="&var.";
RUN;
%MEND mean_ds;
```

Continuous
Summary Statistics

Let metadata drive analyses



```
DATA _null_;  
    set drive0;  
    if type=1 then do;  
        call execute('%mean_ds(dsin=' || memname || ',  
var=' || name || ', driver=drive0)');  
    end;  
  
    if type=2 then do;  
        call execute('%freq_ds(dsin=' || memname || ',  
var=' || name || ', driver=drive0)');  
    end;  
RUN;
```

Output table

```
PROC FORMAT;
  invaline paramord
  'AGE' = 1
  'SEX' = 2
  other = .;

RUN;

DATA tbl_out;
  set mean_: freq_:;
  paramord=input(var. paramord.);
  if nmiss(paramord)=0;

RUN;

PROC SORT data=tbl_out(keep=paramord
param statc);
  BY paramord;

RUN;
```

Ordered Table

Parameter	Statistics
Age	
N	XXX
Mean (SD)	XX.XXX (XX.XXX)
Median	XX
q1, q3	XX, XX
min, max	XX, XX
Sex N(%)	
F	XX (XX.XX)
M	XX (XX.XX)

PARAM	STATC	PARAMORD
Age		1
N	19	1
Mean (SD)	13.316 (1.493)	1
Median	13	1
q1, q3	12, 15	1
min, max	11, 16	1
Sex N(%)		2
F	9 (47.37)	2
M	10 (52.63)	2

Example

```
%include './getfreq.sas';
%include './getmeans.sas';

%macro getdrive(drive=);
options mprint mlogic symbolgen spool;

data _null_;
  set &drive;
  if type=1 then do;
    call execute('%mean_ds(dsin='||memname||',
var='||name||', driver=&drive)');
  end;
  if type=2 then do;
    call execute('%freq_ds(dsin='||memname||',
var='||name||', driver=&drive)');
  end;
run;

%mend getdrive
```

```
PROC SORT Data=sashelp.class out=class0; by
name; run;

DATA class;
  set class0;
  label NAME="First Name" SEX="Sex"
AGE="Age" Height="Height (cm)"
WEIGHT="Weight (kg)";
RUN;

PROC CONTENTS Data=class out=drive0
%getdrive(drive=drive0);

PROC FORMAT;
  invalue paramord
  'AGE' = 1
  'SEX' = 2
  other = .;

RUN;

DATA tbl_out;
  set mean_: freq_;
  paramord=input(var. paramord.);
  if nmiss(paramord)=0;
RUN;

PROC SORT data=tbl_out(keep=paramord param
statc);
  BY paramord;
RUN;
```


Editing The Code

```
%MACRO mean_ds(dsin=, var=,  
driver=);
```

```
PROC SUMMARY data=adsl;  
  var &var.;  
  OUTPUT OUT=outstat  
  N=n MEAN=mean STD=std  
  MEDIAN=median Q1=q1 Q3=q3  
  MIN=min MAX=max;  
RUN;
```

```
DATA mean1;  
  set outstat;  
  r_mean=round(mean, 0.001);  
  r_std=round(std, 0.001);  
  r_med=round(median, 1);  
  
  nc=cats(n);  
  msd=catt(r_mean, ' (,r_std,')');  
  med=cats(median);  
  quart=catx(' ', 'q1, q3');  
  range=catx(' ', 'q1, q3');  
  keep nc msd med r_med quart  
range;  
RUN;
```

Parameter	Statistics
Decimal Variable	
N	XXX
Mean (SD)	XX.XXX (XX.XXX)
Median	XX
q1, q3	XX, XX
min, max	XX, XX

```
PROC TRANSPOSE data=mean1 out=mean1_t;  
  var nc msd med quart range;  
RUN;
```

```
DATA mean2;  
  set mean1_t;  
  ord=_n_;  
  statc=col1;  
run;
```

```
DATA name;  
  set &driver.;  
  where NAME=upcase("&var.");  
  ord=0;  
RUN;
```

```
DATA mean_&var.;  
  length param $200. var $20.;  
  merge name mean2;  
  by ord;  
  if ord=0 then param=cats(label);  
  else param=put(ord, mnfmt.);  
  var="&var.";  
RUN;  
%MEND mean_ds;
```

Benefits

- ▶ Any output that requires this format can be generated by calling this code
- ▶ Call execute works on any and all variables
- ▶ %include can be used for both categorical and continuous macro sections
- ▶ Reduce human errors and QC time

Take home message

- ▶ Procedure 1 for categorical variables
- ▶ Procedure 2 for continuous variables
- ▶ Use Metadata

Q&A



[This Photo](#) by Unknown Author is licensed under [CC BY-NC](#)