

IMS Health & Quintiles are now



Innovative SQLs

Challenging SQL-phobia

Joey Murphy

- + Introduction
- + Conditional Derivations
- + Metadata Review
- + Rudimentary Calculations
- + Summary

Introduction

- The target audience for this presentation are programmers with a good understanding of SAS ® and some rudimentary understanding of Proc SQL code.
- The intention of this presentation is to:
 - Demonstrate some of the lesser used applications for the SQL procedure.
 - Challenge stigma against the regular usage of Proc SQLs.
 - Provide some industry context of appropriate scenarios to implement discussed Proc SQL examples.

Conditional Derivations

Proc SQLs allow you to apply conditional derivations:

```
proc sql;  
  create table demog_agegrp_ord as  
  select *,  
  case  
    when age >= 23 then 'Age >= 23'  
    when age < 23 then 'Age < 23'  
  end as agegrp 'Age category' length=20  
  from demog  
  order by agegrp, sex, subject;  
quit;
```

The demonstrated code:

1. Reads in DEMOG and outputs DEMOG_AGEGRP_ORD.
2. Keeps all input variables (with the * special character).
3. Derives the variable AGEGRP.
4. Sorts the output dataset by AGEGRP, SEX, SUBJECT.

Metadata Review

Method comparison

Consider the following scenario:

The variable APERIOD has been flagged upon review as not CDISC compliant. The study ADaM specifications are in poor condition and we need to know which ADaM domains will require an update.

To find the information we are looking for the following methods can be used:

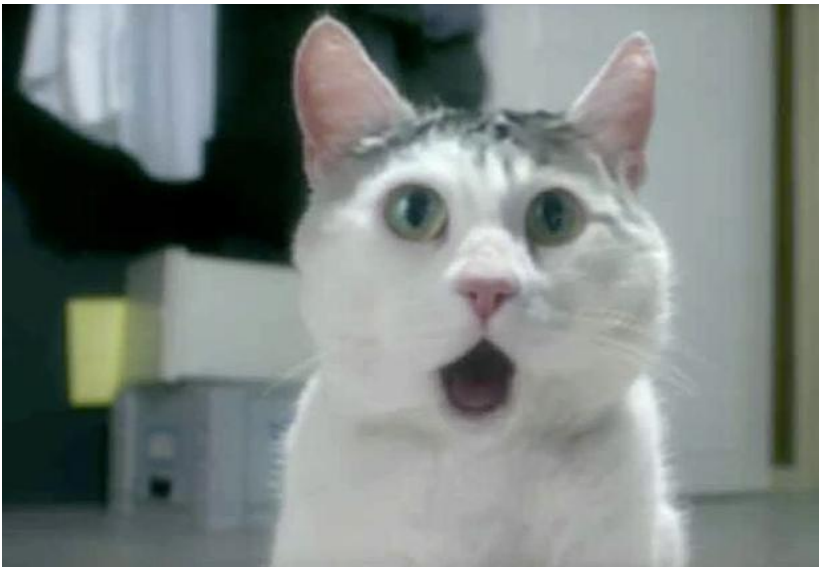
Conventional code	SQL code
<pre>data per_doms_alt(keep=memname); set sashelp.vcolumn(keep=memname libname name); where libname='ADAM' and name='APERIOD'; run;</pre>	<pre>proc sql; create table per_doms as select memname from dictionary.columns where libname='ADAM' and name='APERIOD'; quit;</pre>

Metadata Review

Run time difference

Consider the run time for each method:

Conventional code	SQL code
<pre>NOTE: DATA statement used (Total process time): real time 4.60 seconds cpu time 1.40 seconds</pre>	<pre>NOTE: PROCEDURE SQL used (Total process time): real time 0.09 seconds cpu time 0.00 seconds</pre>



Rudimentary Calculations

SQLs for summary statistics

SQLs can be practical for simple calculation such as min, max and frequency.

subject	sex	age	blwght	agegrp
E1	F	20	50kg	Age < 23
E4	M	22	70kg	Age < 23
E2	F	25	43kg	Age >= 23
E3	F	30		Age >= 23
E5	M	23		Age >= 23
E6	M	28	92kg	Age >= 23

```
proc sql;  
  create table demog_sum_stat as  
  select *,  
    count(age) as n,  
    mean(age) as mean,  
    median(age) as med,  
    min(age) as min,  
    max(age) as max  
  from demog_agegrp_wght  
  group by sex;  
quit;
```

This results as follows:

SUBJECT	SEX	AGE	BLWGHT	AGEGRP	N	MEAN	MED	MIN	MAX
E1	F	20	50kg	Age < 23	3	25	25	20	30
E2	F	25	43kg	Age >= 23	3	25	25	20	30
E3	F	30		Age >= 23	3	25	25	20	30
E6	M	28	92kg	Age >= 23	3	24.33333333	23	22	28
E4	M	22	70kg	Age < 23	3	24.33333333	23	22	28
E5	M	23		Age >= 23	3	24.33333333	23	22	28

Rudimentary Calculations

Critical calculations

See the following example taken from a live study:

```
proc sql;  
  create table max_lab as  
  select distinct usubjid, param, aval  
  from adlb  
  where ablf1^="Y" and aval^=.  
  group by usubjid, param  
  having aval=max(aval);  
quit;
```

This SQL does the following in one step:

1. Filters the input dataset.
2. Creates a unique list of records.
3. Outputs the maximum AVAL result per subject per laboratory parameter.

The application of the demonstrated “having” clause can be applied to multiple scenarios, such as determining:

- Study defined critical and analysis flags.
- Baseline values.
- Last alive dates.

Summary

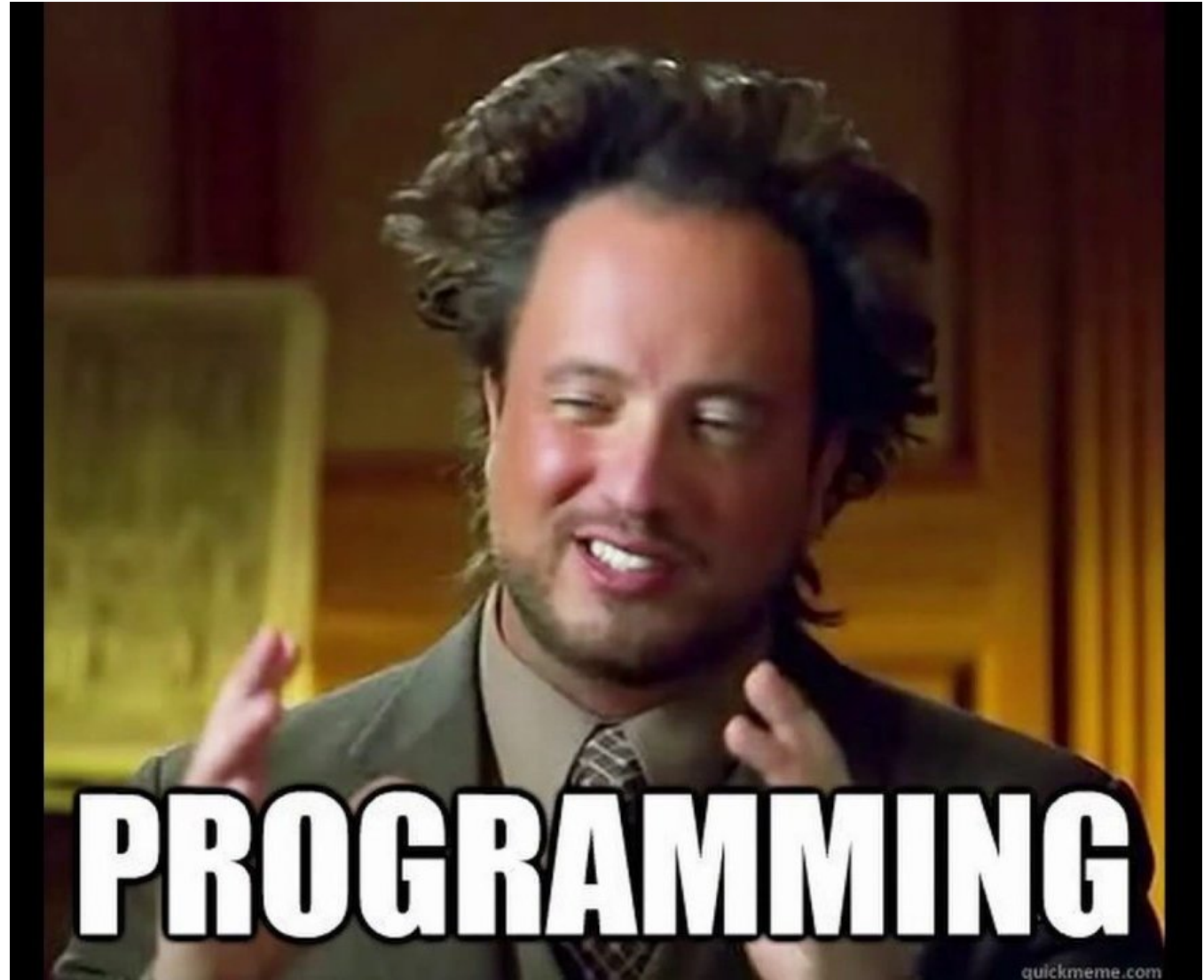
SQLs are not limited to many-to-many merges and creating macros variables.

They can be used to:

- + Derive variables
- + Review metadata and define datasets attributes
- + Generate summary statistics and perform basic derivations
- + Cut down on processing time

Like most things, “people fear what they don’t understand”, and lack of SQL understanding comes from lack of use. Let’s change the stigma and start incorporating an underused and underappreciated procedure.

Summary



Questions



Please Contact me for More Information



Joseph Murphy
Statistical Programmer II

joseph.murphy@iqvia.com
