

Power Data Explorer (PDE)

- Data Exploration in an All-In-One Dynamic Report Using SAS & EXCEL

Harry Chen, Qian Zhao, Janssen R&D China

Lisa Lyons, Janssen R&D US

Introduction

Introduction

- ▶ Getting to know your data is the first key step for every data analyst
- ▶ We may need to do the repetitive, tedious and time-consuming job to
 - ▶ write ad-hoc procs or data steps to summarize and compare different data
 - ▶ develop sophisticated macros that create listings or even 'dynamic' html files to summarize all SAS datasets that may run multiple pages
- ▶ Now we have an “easy-to-explore” overview of multiple SAS datasets in an all-in-one, dynamic report
 - ▶ Require minimum user input, no ad hoc programming but general adaptive template
 - ▶ To know elements (data, variable, value), structure/hierarchy (cross-table) and relationship across data, and other patterns of the data
 - ▶ To compare differences, one vs. one, or one vs. multiple (one 'base', multiple 'compare')
 - ▶ To have Pivot table report in one place, no scattered SAS listings, no hundred pages of lst/html report

Macro call

Macro call

```
%powerDataExplorer(  
    summaryClassXlsFile    = &inputfile  
    ,outdir = &rptdrv\PDEV\&username\PDE  
    ,_PlanningSheet2Vardef = Y  
    ,_debug = Y /*keep output sasdatasets?*/  
    ,_nMaxClassVar=10  
    ,_nMaxIDVar=2  
    ,_nOutMaxMinID=1  
    ,_maxLevLimitValuelist= 200 /*default limit if not specified in input file*/  
    ,_maxLevLimitCrossTable= 300 /*default limit if not specified in input file*/  
    ,_doSubset =Y             /*sample the input data?*/  
    ,_compare=Y             /*compare between two data folders? */  
);
```

SAS Dataset	Description
 _01_datalist.sas7bdat	Dataset level metadata including studyid, data standard (SDTM or ADaM), datasets, labels, filesize, number of observations, number of variables, number of subjects and CDISC class, e.g. Events, Interventions, Findings, BDS, Occurrence.
 _02_varlist.sas7bdat	Variable level metadata including variable level attributes, number of subjects, observations
 _03_valuelist.sas7bdat	Value level metadata including frequency counts per value
 _04_tablevalues.sas7bdat	Table of relationship values across variables, e.g. LBCAT*LBSCAT*LBTESTCD*LBTEST*LBORRESU*LBORRES

Input Files

Input Files

studyID	dataStandard	inDir	Compare
StudyA	SDTM	&topdrv\studyA\Data_Prod\Metadata	compare
StudyA	SDTM	&topdrv\studyA\Data_Prod	
StudyB	SDTM	&topdrv\studyB\Data_Prod\Metadata	base
StudyB	SDTM	&topdrv\studyB\Data_Prod	

Data	types	var	idvar	where	formatVars	formats
all	visitnum*visitdy*visit*epoch*--tptnum*--tpt					
all	--cat*--scat*--sscat*--SPEC*--testcd*--test*--POS*--ORRESU*--STRESU*--stnrlo*--stnrhi					
all	--cat*--scat*--sscat*--testcd*--test*--STRESU*--STAT	--stresn	usubjid	--seq		
all	--cat*--scat*--sscat*--testcd*--test*--STRESU*visitnum*visitdy*visit*epoch*--STAT	--stresn	usubjid	--seq		
all	--cat*--scat*--sscat*--testcd*--test*--ORRESU*--ORRES				#RES #RESC #RESU	\$num2x. \$parenthesis.
all	RDOMAIN*IDVAR*QNAM*QLABEL*QORIG*QEVAL					
dm	ARMCD*ARM*ACTARMCD*ACTARM SEX Race	age	usubjid		SEX	\$total.
AE MH CM	--CAT*--scat*--sscat*--OCCUR*--PRESP					
EX	visitnum*visitdy*visit*epoch*--tptnum*--tpt --CAT*--SCAT*--SSCAT*--TRT*--DOSU*--DOSE--DOSFRM*--DOSFRQ*--ROUTE*--DOSADJ					
FA	--cat*--scat*--sscat*--testcd*--test*--OBJ*--ORRES*--STRESC				#RES #RESC	\$num2x.
IE	IECAT*IESCAT*IETESTCD*IETEST*IESTRESC					
all	#DTC				#DTC	\$num2x.
CD	CODELST*CODEVAL*RNK*DECOD			CODELST in ('ACN' "ACT")		

Input files (cont'd)

► Pattern Based Adaptive Match

- # , any character(including empty)
- - , any non-empty single character
- _ALL_ , all data/variable
- Cross table, adaptively match data

Pattern: --CAT*--SPEC*--TESTCD*--TEST*--POS*--STRESU*--ORRESU

example:

VS: VSTESTCD*VSTEST*VSPOS*VSSTRESU*VSORRESU (no VSCAT and VSSPEC)

LB: LBCAT*LBSPEC*LBTESTCD*LBTEST *LBSTRESU*LBORRESU (no LBPOS)

► Flexible inclusion/exclusion

- Select input data/variable
- Select output data/variable for valuelist

Input files (cont'd)

Demo

Output

Output

1_Data

2_Var

3_Value

4a_CrossTable
4b_crossTableMinMax

Across Domain
Formatted Value

Proc Contents

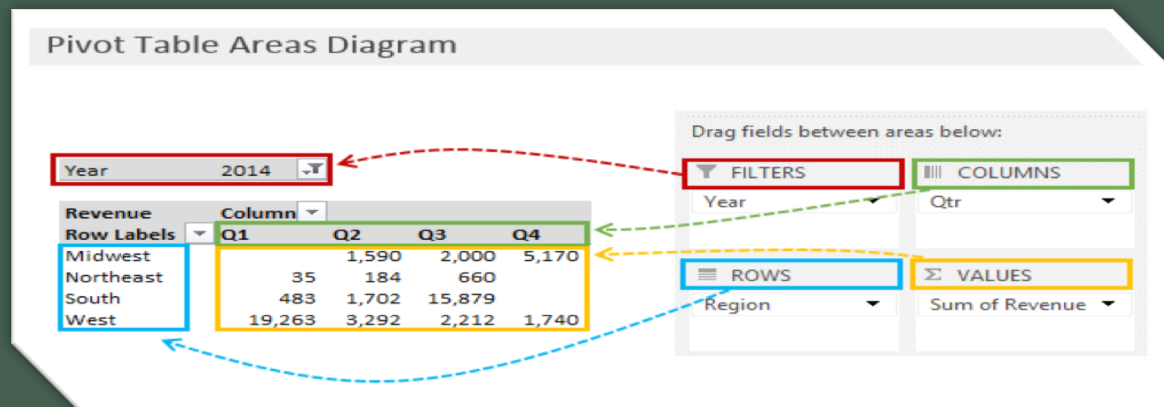
Proc Freq

Proc Summary/Format

Proc Compare

Output (cont'd)

- All-in-one, Dynamic Pivot Table Report
 - Hierarchy and grouping - Sheet, Slicer, Filter, Flag, Rows, Columns
 - Compare data cuts and differences across studies (one vs. one, or more vs. one)
 - Instant searching/filtering at the dataset, variable, value and relationship level
 - Across domain Comparison
 - Identify pattern via formatted value
 - Table Adjustable per user specific needs



Output (cont'd)

► All-in-one, Dynamic Pivot Table Report Data Sample

studyID		CLASSNM		Refresh Data				
studyA		Events						
studyA-COMPARE ONLY		Findings						
studyA-DIF		Interventions						
studyB		Relationship						
		Special Purpose						
		Trial Design						

CLASSNM	MEMNAME	MEMLABEL	NVAR	NOBS	_NSubject	FILESIZE	CRDATE	studyID
Events	TI		(blank)	(blank)	(blank)	(blank)	1/17/1960 23:06	studyA-DIF
	TS		(blank)	2	(blank)	(blank)	1/17/1960 23:06	studyA-DIF
	TV		(blank)	12	(blank)	(blank)	1/17/1960 23:06	studyA-DIF
	VALDEF		(blank)	6	(blank)	(blank)	1/16/1960 21:10	studyA-DIF
	VARDEF		(blank)	39	(blank)	16384	1/16/1960 21:10	studyA-DIF
	VS		(blank)	6414	218	6569984	1/17/1960 23:06	studyA-DIF
	AE	Adverse Events	35	311	74	540672	5/22/2015 10:19	studyB
				736	220	1228800	6/8/2015 9:25	studyA
	CE	Clinical Events	13	7	7	32768	6/8/2015 9:25	studyA
				8	7	32768	5/22/2015 10:19	studyB
	DS	Disposition	15	603	147	638976	5/22/2015 10:19	studyB
				1680	375	1753088	6/8/2015 9:25	studyA
	DV	Protocol Deviations	7	3	3	32768	6/8/2015 9:25	studyA
				7	7	32768	5/22/2015 10:19	studyB
	HO	Hospitalization	11	5	4	32768	5/22/2015 10:19	studyB
				7	6	32768	6/8/2015 9:25	studyA

1_Data	2_Var	3_Value	4a_CrossTable	4b_CrossTableMinMax1	4c_CrossTableVarsSummary	+
--------	-------	---------	---------------	----------------------	--------------------------	---

Output (cont'd)

- All-in-one, Dynamic Pivot Table Report Variable Sample

The screenshot displays a software interface for a dynamic pivot table report. At the top, there are several filter panels: 'studyID' (with options like studyA, studyA-DIF, studyB, studyA-BASE ONLY, studyA-COMPARE ONLY), 'CLASSNM' (with options like Events, Relationship, Findings, Interventions, Special Purp...), 'MEMNAME' (with options like AE, CD, CE, CM, CO, COMPMETH), 'CORE' (with options like Exp, Perm, Req), 'DERVCOPY' (with a UserIncludedFlag dropdown), 'VARROLE' (with options like Identifier, Record Qualifier, Synonym Qualifier, Timing, Topic, Variable Qualifier), 'DATATYPE' (with options like integer, text, float), 'UserExcludedFI...' (with a dropdown), and 'Nlevel' (with a dropdown). Below these filters is a data table with 13 columns: MEMNAME, NAME, LABEL, TYPE, CODELST, LENGTH, FORMAT, Nlevel, VARNUM, NMiss, Nobs, NSubje, and studyID. The table contains data for various medical events (AE) across different studies (studyA, studyB).

MEMNAME	NAME	LABEL	TYPE	CODELST	LENGTH	FORMAT	Nlevel	VARNUM	NMiss	Nobs	NSubje	studyID
AE	AEACN	Action Taken with Study Treatment	char	ACN	20	\$20.	2	19	0	736	220	studyA
	AEBSYCD	Body System or Organ Class Code	num	MedDRA	8	\$.	4	19	0	311	74	studyB
	AEBSYCD	Body System or Organ Class Code	num	MedDRA	8	\$.	19	15	0	311	74	studyB
	AEBSYCD	Body System or Organ Class Code	num	MedDRA	8	\$.	22	15	0	736	220	studyA
	AEBSYCD	Body System or Organ Class Code	num	MedDRA	8	\$.	22	15	0	736	220	studyA
	AEBSYCD	Body System or Organ Class Code	num	MedDRA	8	\$.	19	14	0	311	74	studyB
	AEBSYCD	Body System or Organ Class Code	num	MedDRA	8	\$.	22	14	0	736	220	studyA
	AECONTRT	Concomitant or Additional Treatment Given	char	NY	1	\$1.	2	28	0	311	74	studyB
	AECONTRT	Concomitant or Additional Treatment Given	char	NY	1	\$1.	2	28	0	736	220	studyA
	AEDECOD	Dictionary-Derived Term	char	MedDRA	200	\$200.	146	8	0	311	74	studyB
	AEDECOD	Dictionary-Derived Term	char	MedDRA	200	\$200.	227	8	0	736	220	studyA

Output (cont'd)

- All-in-one, Dynamic Pivot Table Report Value Level Sample

studyID	NAME	_Nlevel
studyA	ACTARM ACTAR... AEACN AECON... AEENDY AEENRF AELLT	(bl... -5 -4 -3 -2 -1
studyA-BASE ONLY	AEOUT AEREL AESCO... AESDIS... AESDTH AESER AESHOSP	1 2 3 4 5 6
studyA-COMPARE ONLY	AESLIFE AESMIE AESTDY AETOX... AGE AGEU ARM	7 8 9 10 11 12
studyA-DIF	ARMCD CEACN... CEENDY CESTDY CLASSN... CMCAT CMDOSE	13 14 15 16 17 18
studyB	CMDOS... CMDOS... CMDOSU CMDUR CMENDY CMENRF CMINDC	19 20 21 22 23 24

MEMNAME	NAME	_Nlevel	_LEVEL	VALUE	studyID	Sum of _FREQ_
AE	AEACN	(blank)	(blank)	DRUG INTERRUPTED	studyA-BASE ONLY	1
				DRUG WITHDRAWN	studyA-BASE ONLY	4
		-2 (blank)		DOSE NOT CHANGED	studyA-DIF	348
				NOT APPLICABLE	studyA-DIF	82
		2		1 DOSE NOT CHANGED	studyA	611
				2 NOT APPLICABLE	studyA	125
		4		1 DOSE NOT CHANGED	studyB	263
				2 DRUG INTERRUPTED	studyB	1
				3 DRUG WITHDRAWN	studyB	4
				4 NOT APPLICABLE	studyB	43
	AECONTR	(blank)	(blank)	N	studyA-DIF	314

Output (cont'd)

► All-in-one, Dynamic Pivot Table Report Cross-Tabs Sample

studyID

studyA
studyA- COMPARE ONLY
studyA- DIF
studyB
studyA- BASE ONLY

classVarType

VISITNUM*VISIT*EPOCH*QSTPTNUM*QSTPT
VISITNUM*VISIT*EPOCH*RFTPTNUM*RFTPT
VISITNUM*VISIT*EPOCH*VSTPTNUM*VSTPT
VSDTC
VSTESTCD*VSTEST*VSORRESU*VSORRES
VSTESTCD*VSTEST*VSPOS*VSORRESU*VSSTRESU
VSTESTCD*VSTEST*VSSTRESU

_classVarType VSTESTCD*VSTEST*VSPOS*VSORRESU*VSSTRESU						MEMNAME	
Sum of _FREQ_	classValue1	classValue2	classValue3	classValue4	classValue5	studyID	VS
	DIABP	DIASTOLIC BLOOD	SUPINE	mmHg	mmHg	studyA	2399
						studyA- DIF	1550
						studyB	849
	HEIGHT	HEIGHT	SUPINE	cm	cm	studyA	110
						studyA- DIF	81
						studyB	29
				in	cm	studyA	222
						studyA- DIF	137
						studyB	85
	PULSE	PULSE RATE	SUPINE	BEATS/MIN	BEATS/MIN	studyA	2399
						studyA- DIF	1550
						studyB	849
	SYSBP	SYSTOLIC BLOOD	SUPINE	mmHg	mmHg	studyA	2399
						studyA- DIF	1550
						studyB	849
	TEMP	TEMPERATURE	SUPINE	C	C	studyA	3
				F	C	studyA- COMPARE ONLY	3
						studyA	8
						studyA- DIF	-1
						studyB	9

MEMNAME

VS
AE
CD
CE
CM
CO
DA
DIF

Output (cont'd)

Demo

Summary

Summary

- ❖ An all-in-one dynamic report
- ❖ Minimum input, general adaptive template and smart match
- ❖ Easy to run/review, one SAS submit and one Pivot table output
- ❖ Instant filtering at the dataset, variable, value and relationship level through cross-tabulations
- ❖ Compare data cuts and differences across studies
- ❖ Especially helpful for more efficient planning and integration
- ❖ SAS & EXCEL (future other data visualization application can be adopted)
- ❖ COMPUTER Calculation & HUMAN Interaction
- ❖ REPITITION & FLEXIBILITY

Acknowledgement

We would like to thank the Janssen Autocode team Ilse Augustyns, Nick Masel, Phyllis Wolf and Sandy Lei, for the contribution to the macros described in this paper and valuable comments.

Power Data Explorer (PDE) - Data Exploration in an All-In-One Dynamic Report Using SAS & EXCEL

QUESTIONS?

Harry Chen
Janssen R&D China.
Phone: 86-21-3337-8892
e-mail: hchen104@its.jnj.com

Qian Zhao
Janssen R&D China.
Phone: 86-21-3337-7624
e-mail: qzhao1@its.jnj.com

Lisa Lyons
Janssen R&D US
Phone: (609)-802-6283
e-mail: llyons3@its.jnj.com



Backup Slides

Summary by flexible user-defined format

studyID	_classVarType
studyA	RFENDTC
studyA-BASE ONLY	RFICDTC
studyA-COMPARE ONLY	RFPENDTC
studyA-DIF	RFSTDTC
studyB	RFXENDTC
	RFXSTDTC
	SCDTC

_classVarType	Sum of _FREQ_	MEMNAME	AE	CE	CM	CO	DA	DM	DP	DS	EG	EX	FA	GT
studyA	117	1503	266	22	700	1370								
studyA	1	1032			2									
studyA	18	307												
studyA	1336	7	1376	3090	2675	756	1990	375	2286	608	22671			
studyA		7												781

```

%macro umPatternFormat();
  proc fcmp outlib=work.functions.smd;
    function num2x(text $) $;
      if compress(TEXT, '-T:', 'd') = '' then do;
        x = TRANSLATE(TEXT, 'xxxxxxxx', '0123456789');
      end;
      else x = TEXT;
      return(x);
  endsub;
run;

options cmplib={work.functions};
proc format;
  value $num2x(default=200)
    other=[num2x()];
run;

%mend;

```

SAS 94 new feature

Data	types	var	idvar	where	formatVars	formats
all	#DTC				#DTC	\$num2x.

Summary by flexible user-defined format

studyID

- studyA
- studyA-COMPARE ONLY
- studyA-BASE ONLY
- studyA-DIF
- studyB

_classVarType

- PCCAT*PCSCAT*PCTESTCD*PCTEST*PCORRESU*PCORRES
- PCCAT*PCSCAT*PCTESTCD*PCTEST*PCSTRESU*PCSTAT
- PCCAT*PCSCAT*PCTESTCD*PCTEST*PCSTRESU*VISITNUM*VISIT*EPOCH*PCSTAT
- PCDTC
- PEDTC
- PETESTCD*PETEST
- PETESTCD*PETEST*PEORRES

_classVarType PCCAT*PCSCAT*PCTESTCD*PCTEST*PCORRESU*PCORRES

Sum of _FREQ_

MEMNAME	classValue1	classValue2	classValue3	classValue4	classValue5	classValue6	classValue7	studyID	PC
ANALYTE	drg1	DrugName	(ng/mL)	xx.x	xxx	xxxx	studyA	5	
								19	
								9	
								3	
								26	
								1	
								4	
								5	
								3	
								26	
drg2SBVR	DrugName2	(ng/mL)	<5.00	x.xx	xx.x	xxx	studyA	7	
GS331007	GS-331007	(ng/mL)	xx.x	xxx	xxxx	studyA	7		

Data	types	var	idvar	where	formatVars	formats
all	--cat*--scat*--sscat*--testcd*--test*--ORRESU*--ORRES				#RES #RESC #RESU	\$num2x. \$parenthesis.

Compare across domain

[illegible]

Data	types
all	USUBJID

Compare across domain – ‘Time and Events Schedule’

studyID	_classVarType
studyA	VISITNUM*VISIT+EPOCH*HUTPTNUM*HUTPT
studyA - BASE ONLY	VISITNUM*VISIT+EPOCH*HUTPTNUM*HUTPT
studyA - COMPARE ON...	VISITNUM*VISIT+EPOCH*HUTPTNUM*HUTPT
studyA - DIF	VISITNUM*VISIT+EPOCH*HUTPTNUM*HUTPT
studyB	VISITNUM*VISIT+EPOCH*HUTPTNUM*HUTPT

_classVarType (Multiple Item)	Sum of _FREQ_	MEMNAME	DP	EX	FA	GT	HU	II	LB	PC	PE	QS
1.000 SCREENING	999.000 NO_POINT_DI	DA		1143					55			
	1200.000 12WEEK								51			
	1202.000 END								59			
	1203.000 AFTER								9			
	1204.000 2MONTHS_Af								59			
	-2.000 BEFORE											
10.000 WITHDRAWAL	999.000 NO_POINT_DI		758		332	20173		24910		111		
2.000 BASELINE	999.000 NO_POINT_DI					277	12	17562		60		67
	-1.000 PRE - DOSE							310		57		21269
3.000 DAY3	999.000 NO_POINT_DI	1240					3720	383				310
4.000 DAY7	999.000 NO_POINT_DI						3708	16260		47		10201
5.000 WEEK2	0.000 REFERENCE									18		
	2.000 AFTER_PREVH									18		
	-991.000 MOST_RECEN							616				
	-992.000 2ND_MOST_R							616				
	999.000 NO_POINT_DI											
6.000 WEEK3	999.000 NO_POINT_DI						3696		16196		61	10118
7.000 WEEK4	0.000 REFERENCE						3684		307			10146
	2.000 AFTER_PREVH									18		
	-991.000 MOST_RECEN							620				
	-992.000 2ND_MOST_R							620				
	999.000 NO_POINT_DI	1238					3720		16934		59	20802
8.000 WEEK8	0.000 REFERENCE									18		
	2.000 AFTER_PREVH									18		
	-991.000 MOST_RECEN							614				
	-992.000 2ND_MOST_R							614				
	999.000 NO_POINT_DI	610					3708		17537		71	20822

Data types

all visitnum*visitdy*visit*epoch*--tptnum*--tpt