



**BC Cancer Agency**

CARE + RESEARCH

*An agency of the Provincial Health Services Authority*

**Cancer Surveillance & Outcomes**

# Reporting from Base SAS Tips & Tricks

Fareeza Khurshed  
BC Cancer Agency

# Overview

---

- Index for large data
- Summarizing Data
- Getting Data to Excel



**BC Cancer Agency**

CARE + RESEARCH

*An agency of the Provincial Health Services Authority*

**Cancer Surveillance & Outcomes**

# Index

---

- Think of book index or library catalogue or search function



**BC Cancer Agency**



























CARE + RESEARCH

*An agency of the Provincial Health Services Authority*

**Cancer Surveillance & Outcomes**

# Index

---

-   Procedures
  -   The APPEND Procedure
  -   The CALENDAR Procedure
  -   The CALLRFC Procedure
  -   The CATALOG Procedure
  -   The CHART Procedure
  -   The CIMPORT Procedure
  -   The COMPARE Procedure
  -   The CONTENTS Procedure
  -   The COPY Procedure
  -   The CORR Procedure
  -   The CPORT Procedure
  -   The CQUEST Procedure



# Index

---

- Example with sashelp.zipcode file
  - 41761 Zipcodes (version dependent)
  - 58 States
  - $58/41761 = 0.14\%$

```
libname zip 'C:\temp';
```

```
data zip.zipcode;
```

```
    set sashelp.zipcode; *Copy file;
```

```
run;
```



# Index

---

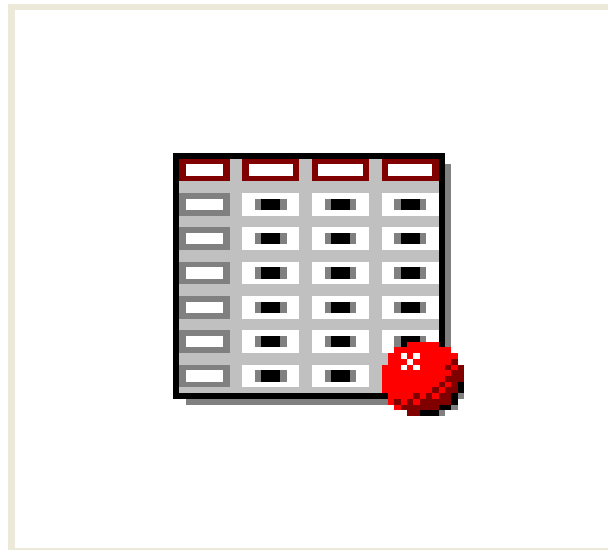
- Can be created using Proc SQL, Proc Datasets, or a Data step
- Proc Datasets example

```
PROC DATASETS library=zip;*Library name;  
    MODIFY zipcode; *Dataset name;  
    Index create state_code;*Single field;  
Run;quit;
```

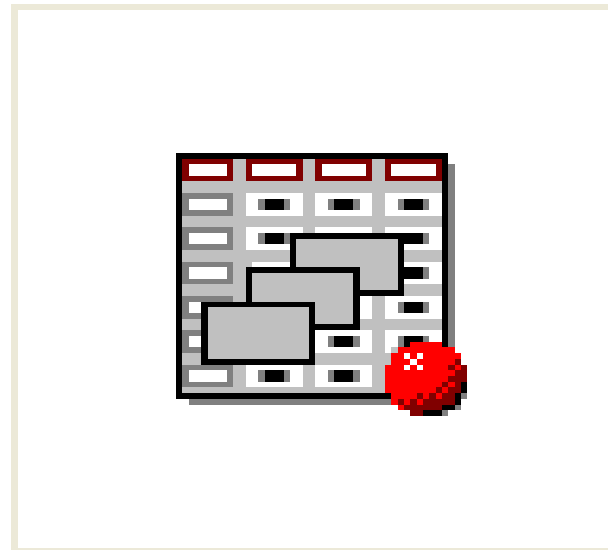


# Create Index

---



zipcode.sas7bdat



zipcode.sas7bndx



# Check for Index

---

```
Proc contents data=zip.zipcode;  
Run;
```

## Alphabetic List of Indexes and Attributes

#	Index	# of Unique Values
1	STATECODE	58





# When to Index

---

- Specific field(s) that data can be filtered on
- Field will be filtered on using a WHERE, BY or KEY clause
- Reference file should be less than 20% of file
- Data set required for analysis is large
- Permission to modify data



# Using an Index

---

- SAS Does It!
- Should be used in WHERE, BY or KEY
- Option msglevel=i;

```
1999 ;  
INFO: Index booking_form_received_date selected for WHERE clause optimization.  
2000 run ;
```



# Index – Example

---

- Dataset is all surgeries performed in BC from
- 3GB, 1.8 million rows, 102 variables
- Wait time for surgery
- Cohort analysis based on when surgery booked



# Index – Example

---

- Analysis is done on Health Authority, Health Service Delivery Area and Hospitals by Date
- Which to use as index?
  - Main filter for report
  - Reduces data into smallest subset
- Date best choice



# Index – Example

---

```
*Sort Dataset;  
proc sort data=spr.spr_report_data_detailed;  
    by booking_form_received_date;  
run;
```

```
*Create an index on the booking_form_received_date;  
proc datasets library=spr NODETAILS NOLIST;  
    modify spr_report_data_detailed;  
    index create booking_form_received_date;  
run;QUIT;
```

```
*Query Dataset;  
data spr_select_dates;  
    set spr.spr_report_data_detailed;  
    where booking_form_received_date between "&start_date."d and "&end_date."d;  
run ;
```

8:57  
minutes

3 reports

~27 minutes



# Index – Example

---

\*Sort Dataset;

```
proc sort data=spr.spr_report_data_detailed;  
  by booking_form_received_date;  
run;
```

15:00  
minutes

\*Create an index on the booking\_form\_received\_date;

```
proc datasets library=spr NODetails NOLIST;  
  modify spr_report_data_detailed;  
  index create booking_form_received_date;  
run;QUIT;
```

\*Query Dataset;

```
data spr_select_dates;  
  set spr.spr_report_data_detailed;  
  where booking_form_received_date between "&start_date."d and "&end_date."d;  
run ;
```

10:00  
minutes

25:00  
minutes

3 reports  
~ 45 minutes



# Index – Example

---

\*Sort Dataset;

```
proc sort data=spr.spr_report_data_detailed;  
  by booking_form_received_date;  
run;
```

\*Create an index on the booking\_form\_received\_date;

```
proc datasets library=spr NODetails NOLIST;  
  modify spr_report_data_detailed;  
  index create booking_form_received_date;  
run;QUIT;
```

9:00  
minutes

\*Query Dataset;

```
data spr_select_dates;  
  set spr.spr_report_data_detailed;  
  where booking_form_received_date between "&start_date."d and "&end_date."d;  
run ;
```

3:30  
minutes

12:30  
minutes      3 reports  
~19:30 minutes



# Index – Example

---

```
*Sort Dataset;  
proc sort data=spr.spr_report_data_detailed;  
    by booking_form_received_date;  
run;
```

10:00  
minutes

```
*Create an index on the booking_form_received_date;  
proc datasets library=spr NODetails NOLIST;  
    modify spr_report_data_detailed;  
    index create booking_form_received_date;  
run;QUIT;
```

9:00  
minutes

```
*Query Dataset;  
data spr_select_dates;  
    set spr.spr_report_data_detailed;  
    where booking_form_received_date between "&start_date."d and "&end_date."d;  
run ;
```

7 seconds

19:07      3 reports  
minutes    ~19:21 minutes

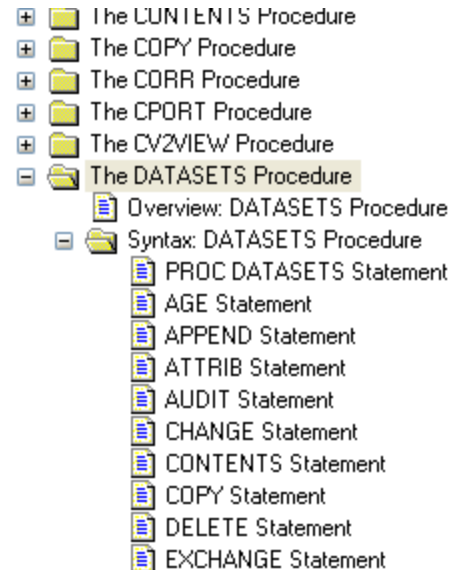




# Index

---

- Simple Index – single field
- Multiple Single Indices
- Composite Index – multiple field



# Index - References

---

## Paper 123-29 Creating and Exploiting SAS Indexes

Michael A. Raithel, Westat, Rockville, MD



**BC Cancer Agency**

CARE + RESEARCH

*An agency of the Provincial Health Services Authority*

**Cancer Surveillance & Outcomes**

# Summarizing Data

---

- Summary stats such as N/Mean/Median
- Summary at different levels
  - Subtotals
  - Summary by different categories



# Summarizing data

---

- Proc means
  - Use class statement
  - Can control interactions
  - Can use **WAYS** and **TYPE** statements to control output levels



# Proc Means

---

```
title 'Example of Ways Statement' ;  
proc means data=sashelp.class n mean median maxdec=2;  
    class sex age;  
    WAYS 0 1 2 ;  
    var height;  
run;
```

WAYS controls the interaction levels

0 – Overall Total

1 – Individual Total for each class level

2 – Two Way Interaction between each class level



# Proc Means

## Example of Ways Statement

The MEANS Procedure

Analysis Variable : Height				
N Obs	N	Mean	Median	
19	19	62.34	62.80	

Analysis Variable : Height					
Age	N Obs	N	Mean	Median	
11	2	2	54.40	54.40	
12	5	5	59.44	59.00	
13	3	3	61.43	62.50	
14	4	4	64.90	63.90	
15	4	4	65.63	66.50	
16	1	1	72.00	72.00	

Analysis Variable : Height				
Sex	N Obs	N	Mean	Median
F	9	9	60.59	62.50
M	10	10	63.91	64.15

Analysis Variable : Height					
Sex	Age	N Obs	N	Mean	Median
F	11	1	1	51.30	51.30
	12	2	2	58.05	58.05
	13	2	2	60.90	60.90
	14	2	2	63.55	63.55
	15	2	2	64.50	64.50
M	11	1	1	57.50	57.50
	12	3	3	60.37	59.00
	13	1	1	62.50	62.50
	14	2	2	66.25	66.25
	15	2	2	66.75	66.75
	16	1	1	72.00	72.00



# Proc Means

```
title 'Example of Type Statement' ;  
proc means data=sashelp.cars ;  
class make type cylinders;  
types make * (type cylinders);  
var mpg_city;  
run;
```

## Example of Type Statement

### The MEANS Procedure

Analysis Variable : MPG_City MPG (City)					
Make	Cylinders	N Obs	N	Mean	Median
Acura	4	2	2	23.00	23.00
	6	5	5	18.00	18.00
BMW	6	16	16	19.00	19.50
	8	4	4	17.50	18.00
Hyundai	4	6	6	27.50	27.50
	6	6	6	18.50	19.00

Analysis Variable : MPG_City MPG (City)					
Make	Type	N Obs	N	Mean	Median
Acura	SUV	1	1	17.00	17.00
	Sedan	5	5	20.40	20.00
	Sports	1	1	17.00	17.00
BMW	SUV	2	2	16.00	16.00
	Sedan	13	13	19.23	19.00
	Sports	4	4	18.25	18.00
	Wagon	1	1	19.00	19.00
Hyundai	SUV	1	1	20.00	20.00
	Sedan	10	10	23.70	26.00
	Sports	1	1	19.00	19.00



# Proc Means – Output to dataset

---

```
title 'Example of Type Statement' ;  
ods table summary=ex_type1;  
proc means data=sashelp.cars ;  
    where make in ('Acura', 'BMW', 'Hyundai');  
class make type cylinders;  
types make * (type cylinders);  
var mpg_city;  
output out=ex_type2;  
run;
```





# Proc Means - Output

Make	Type	Cylinders	N Obs	N	Mean	Std Dev	Minimum	Maximum
Acura		4	2	2	23	1.4142135624	22	24
Acura		6	5	5	18	1.2247448714	17	20
BMW		6	16	16	19	1.5916448515	16	21
BMW		8	4	4	17.5	1	16	18
Hyundai		4	6	6	27.5	1.6431676725	26	29
Hyundai		6	6	6	18.5	1.2247448714	17	20

**ex\_type1**

	Make	Type	Cylinders	_TYPE_	_FREQ_	_STAT_	MPG (City)
	Acura		4	5	2	N	2
	Acura		4	5	2	MIN	22
	Acura		4	5	2	MAX	24
	Acura		4	5	2	MEAN	23
	Acura		4	5	2	STD	1.4142135624
	Acura		6	5	5	MIN	17

**ex\_type2**



# References – Proc Means

---

- SAS Documentation



**BC Cancer Agency**

CARE + RESEARCH

*An agency of the Provincial Health Services Authority*

**Cancer Surveillance & Outcomes**

# ODS Tagsets – ExcelXP

---

- Generates XML readable by Excel
- Requires template code download from [support.sas.com](http://support.sas.com)
- [Available here](#)
- Run it once and good to go (until you need an update)



# ODS tagsets.ExcelXP

---

- Some useful features:
  - Formatting of rows, columns or cells using proc report
  - Automatic insertion of filters to let user filter the result
  - Control which tabs results are placed on



# Tagsets.ExcelXP

---

```
ods tagsets.excelxp file="C:\temp\sashelp_class.xls"  
  style=journal  
options(  
  fittopage='yes' orientation='landscape' scale='150'  
  sheet_interval='proc' autofilter='1-3'  
  sheet_name='Class Summary');
```

**SAS CODE GOES HERE;**

```
ods tagsets.excelxp close;
```



# Tagsets.ExcelXP

---

```
proc report data=sashelp.class nowd split='*';  
  column (name) ('Demographics' sex age height weight);  
  
  define name/center 'Name';  
  define sex/ center 'Sex';  
  define age/center 'Age*(yrs)';  
  define height/center 'Height*(in)';  
  define weight/center 'Weight*(lbs)';  
  
  compute name;  
    if find(name,'l')>0 then  
      call define(_row_, 'style', 'style=[background=Grey]');  
  endcomp;  
  
run;quit;
```



# Tagsets.ExcelXP

sashelp\_class.xls [Read-Only]

	A	B	C	D	E
1	Demographics				
2	Name	Sex	Age (yrs)	Height (in)	Weight (lbs)
3	Alice	F	13	56.5	84
4	Barbara	F	13	65.3	98
5	Carol	F	14	62.8	102.5
6	Jane	F	12	59.8	84.5
7	Janet	F	15	62.5	112.5
8	Joyce	F	11	51.3	50.5
9	Judy	F	14	64.3	90
10	Louise	F	12	56.3	77
11	Mary	F	15	66.5	112
12	Alfred	M	14	69	112.5
13	Henry	M	14	63.5	102.5
14	James	M	12	57.3	83
15	Jeffrey	M	13	62.5	84
16	John	M	12	59	99.5
17	Philip	M	16	72	150
18	Robert	M	12	64.8	128
19	Ronald	M	15	67	133
20	Thomas	M	11	57.5	85
21	William	M	15	66.5	112
22					



# Tagsets.ExcelXP

---

- Add in Titles/Remove styles

```
ods tagsets.excelxp file="C:\temp\sashelp_class.xls"  
options(  
  embedded_titles='yes' fittopage='yes' orientation='landscape' scale='150'  
  sheet_interval='none'  
  sheet_name='Summary');
```

```
title 'Frequency of Age and Sex';  
proc freq data=sashelp.class;  
  tables age sex;  
run;
```

```
title 'Summary Stats of Weight and Height';  
proc means data=sashelp.class maxdec=2;  
  var weight height;  
run;
```

```
ods tagsets.excelxp close;
```





# Tagsets.ExcelXP

	A	B	C	D	E	F
1	<i>Frequency of Age and Sex</i>					
2						
3	<b>Age</b>	<b>Frequency</b>	<b>Percent</b>	<b>Cumulative Frequency</b>	<b>Cumulative Percent</b>	
4	<b>11</b>	2	10.53	2	10.53	
5	<b>12</b>	5	26.32	7	36.84	
6	<b>13</b>	3	15.79	10	52.63	
7	<b>14</b>	4	21.05	14	73.68	
8	<b>15</b>	4	21.05	18	94.74	
9	<b>16</b>	1	5.26	19	100	
10						
11	<b>Sex</b>	<b>Frequency</b>	<b>Percent</b>	<b>Cumulative Frequency</b>	<b>Cumulative Percent</b>	
12	<b>F</b>	9	47.37	9	47.37	
13	<b>M</b>	10	52.63	19	100	
14						
15	<i>Summary Stats of Weight and Height</i>					
16						
17	<b>Variable</b>	<b>N</b>	<b>Mean</b>	<b>Std Dev</b>	<b>Minimum</b>	<b>Maximum</b>
18	Weight	19	100.03	22.77	50.50	150.00
19	Height	19	62.34	5.13	51.30	72.00
20						
21						



# Tagsets.ExcelXP

---

- MultiSheet Example

```
ods tagsets.excelxp file="C:\temp\sashelp_class.xls"  
options(  
  embedded_titles='yes' fittopage='yes' orientation='landscape' scale='150'  
  sheet_interval='by'  
  sheet_label=");
```

```
title 'Example of Multiple Sheets';
```

```
proc sort data=sashelp.class out=class; by sex; run;
```

```
proc print data=class noobs label;
```

```
  by sex;
```

```
run;
```

```
ods tagsets.excelxp close;
```



# Tagsets.ExcelXP

Results Viewer - sashelp\_class.xls

Example of Multiple Sheets

Sex=F

Name	Age	Height	Weight
Alice	13	56.5	84
Barbara	13	65.3	98
Carol	14	62.8	102.5
Jane	12	59.8	84.5
Janet	15	62.5	112.5
Joyce	11	51.3	50.5
Judy	14	64.3	90
Louise	12	56.3	77
Mary	15	66.5	112

Table 1 - Sex=F / Table 2 - Sex=M



# References

---

- [http://support.sas.com/rnd/base/ods/odsmarkup/excelxp\\_help.html](http://support.sas.com/rnd/base/ods/odsmarkup/excelxp_help.html)
- [http://support.sas.com/rnd/base/ods/odsmarkup/excelxp\\_demo.html](http://support.sas.com/rnd/base/ods/odsmarkup/excelxp_demo.html)
- Paper 170-2011 Creating Stylish Multi-Sheet Microsoft Excel Workbooks the Easy Way with SAS
  - Vince DelGobbo, SAS Institute Inc. Cary, NC



**BC Cancer Agency**

CARE + RESEARCH

*An agency of the Provincial Health Services Authority*

**Cancer Surveillance & Outcomes**

# Conclusion

---

- Speed up queries with an index
- Use Types and Ways to summarize data with one step
- Exporting to Excel with Tagsets.ExcelXP



# Acknowledgements

---

Thanks for the support of the BC Cancer  
Agency

Thanks for the support of the Biostatisticians  
within Cancer Surveillance & Outcomes



**BC Cancer Agency**

CARE + RESEARCH

*An agency of the Provincial Health Services Authority*

**Cancer Surveillance & Outcomes**

# Questions

---



**BC Cancer Agency**

CARE + RESEARCH

*An agency of the Provincial Health Services Authority*

**Cancer Surveillance & Outcomes**

# Contact Info

---

[fkhurshed@bccancer.bc.ca](mailto:fkhurshed@bccancer.bc.ca)

**Fareeza Khurshed, MSc.**

*Biostatistician*

Cancer Surveillance & Outcomes

Population Oncology

BC Cancer Agency

801- 686 West Broadway

Vancouver, BC V5Z 1G1



BC Cancer Agency

CARE + RESEARCH

An agency of the Provincial Health Services Authority

Cancer Surveillance & Outcomes