### PROC REPORT

Shannon Pileggi

**STAT 330** 



Overview

### OUTLINE

Overview

Detail Report

Summary Report with GROUP

Summary Report with ACROSS

More

#### Overview

Overview •000 Detail Report

PROC	Detail	Summary	Control	N	sum	mean	std	%
PRINT	<b>√</b>	Х	$\checkmark$	<b>√</b>	<b>√</b>	X	X	Х
MEANS	X	$\checkmark$	X	<b>√</b>	$\checkmark$	$\checkmark$	$\checkmark$	X
FREQ	X	$\checkmark$	X	<b>√</b>	X	X	X	$\checkmark$
REPORT	<b>√</b>	<b>√</b>	<b>√</b>	<b>√</b>	<b>√</b>	<b>√</b>	<b>√</b>	<b>√</b>
TABULATE	X	$\checkmark$	$\checkmark$	<b>√</b>	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
SQL	<b>√</b>	$\checkmark$	X	<b>√</b>	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

- ▶ Detail: display a row for each observation
- ► Summary: display a row for a group of observations
- ► Control: many layout/format/display options in output
- ▶ SQL: can additionally combine and sort data

◆□▶◆□▶◆□▶◆□▶ □ り<00</p>

3 / 29

STAT 330: Lecture 16

### Patents data

- number of utility patent ("patents for inventions") grants from 2011, by county
- demographic variables from the American Community Survey
  - some variables may be missing for smaller counties
- San Jose, CA (Santa Clara County)
  - ▶ 3<sup>rd</sup> largest city in CA, 10<sup>th</sup> largest city in US
  - leads all US cities in generating patents

On your own: Explore the patents data in SAS.

# Syntax

Overview

0000

```
PROC REPORT DATA = dataset;
COLUMN item1 item2 ...;
DEFINE item1 / options;
DEFINE item2 / options;
RUN;
SAS Code
```

#### Default properties:

- Permanent formats/labels automatically applied
- Character vars left justified, numeric vars right-justified
- ▶ an item can be a variable or a statistic
- ► COLUMN specifies *items* to use and their order of appearance
- ▶ DEFINE specifies the *item's* use and display

# DEFINE usages

DEFINE var1 / usage;

Usage	Detail	Summary	Description
DISPLAY	$\checkmark$	X	creates 1 row per obs
ORDER	$\checkmark$	X	creates 1 row per obs, ordered
ANALYSIS	<b>√</b>	✓	calculates statistics
COMPUTED	$\checkmark$	$\checkmark$	creates new variable
GROUP	X	<b>√</b>	values placed in rows
ACROSS	X	$\checkmark$	values placed in columns

Summary Report with GROUP

....

Overview

Detail Report

•0000000

7 / 29

Summary Report with ACROSS

More

### SAS Code

```
PROC REPORT DATA = patents ;
RUN:
PROC REPORT DATA = patents ;
   COLUMN state county patents;
RUN:
PROC REPORT DATA = patents ;
   COLUMN state county patents;
   DEFINE state / DISPLAY ;
   DEFINE county / DISPLAY ;
   DEFINE patents / ANALYSIS;
RUN:
          SAS Code -
```

- with no statements, prints all data
- COLUMN specifies variables to print (and order of display)
- equivalent output to previous PROC
- Default usages: DTSPLAY for character vars ANALYSTS for numeric vars

# Getting started - example output

Detail Report

0000000

```
PROC REPORT DATA = patents;

COLUMN state county patents;

DEFINE state / DISPLAY;

DEFINE county / DISPLAY;

DEFINE patents / ANALYSIS;

RUN;

SAS Code
```

```
Number
                                     of
   State
               US county name
                                   patents
AI ARAMA
            Baldwin County
                                         8
ALABAMA
            Calhoun County
ALABAMA
            Cullman County
ALABAMA
            DeKalb County
                                         2
                                         2
ALABAMA
            Elmore County
                                         2
ALABAMA
            Etowah County
AI ARAMA
            Houston County
AI ARAMA
            Jefferson County
                                       51
AI ARAMA
            Lauderdale County
                                         5
AI ARAMA
            Lee County
                                       24
                                       27
ALABAMA
            Limestone County
            Madison County
ALABAMA
                                      122
            Marshall County
ALABAMA
                                         6
```

Overview

More

Overview

Summary Report with GROUP

## ORDER usage and SPANROWS

```
SAS Code
PROC REPORT DATA = patents SPANROWS;
   COLUMN state county patents;
   DEFINE state / ORDER;
   DEFINE county / ORDER;
   DEFINE patents / ANALYSIS;
RUN:
            SAS Code _
```

- ▶ With ORDER:
  - rows arranged by ascending values
  - repetitious printing is suppressed
- SPANROWS merges cells with same values

State	US county name	Number of patents
ALABAMA	Baldwin County	8
	Calhoun County	1
	Cullman County	4
	DeKalb County	2
	Elmore County	2
	Etowah County	2
	Houston County	3
	Jefferson County	51
	Lauderdale County	5
	Lee County	24
	Limestone County	27
	Madison County	122
	Marshall County	6
	Mahila Caustu	12

STAT 330: Lecture 16 10 / 29

#### Discussion

US county name	Number of patents	Population estimate
Harrison County	3	69436
Raleigh County	1	79127
Wood County	10	87120
Cabell County	2	96653
Monongalia County	18	98528
Berkeley County	9	105750
Kanawha County	12	192315

### Identify the *usage* of:

- 1. county
- 2. patents
- 3. population

#### Discussion

Overview

```
PROC REPORT DATA = patents ;
WHERE state = "ALABAMA";
COLUMN county patents population;
DEFINE county / DISPLAY;
DEFINE patents / ORDER;
DEFINE population / ANALYSIS;
RUN;
SAS Code
```

On your own: What do the blank values of patents represent?

	Number	
US county name	of patents	Population estimate
St. Clair County	0	84398
Talladega County		81664
Calhoun County	1	117797
Montgomery County		232032
DeKalb County	2	71375
Elmore County		80162
Etowah County		104303
Walker County		66661
Houston County	3	102369
Cullman County	4	80536
Lauderdale County	5	92781
Marshall County	6	94166
Morgan County	7	119953

# Spanning Column Headings

```
COLUMN ("header1" items) ("header2" items)
```

SAS Code \_\_\_\_\_

```
PROC REPORT DATA = patents SPANROWS;

COLUMN ("Location" state county)

patents

("Demographics" population age education income);

DEFINE state / order;

DEFINE county / order;

RUN;
```

SAS Code \_\_\_\_\_

		Location		Demographics			
	State	US county name	Number of patents	Population estimate	Median Age	Education level of bachelor's degree or more (%)	Median household income
	ALABAMA	Baldwin County	8	186717	42	28.3	50900
		Calhoun County	1	117797	39.6	15.2	39037
TAT 330: Lec	ture 16	Cullman County	4	80536	414	14.2	40054

13 / 29

 Overview
 Detail Report
 Summary Report with GROUP
 Summary Report with ACROSS
 Mo

 0000
 0000000●
 00000
 00000
 00000

#### Formats and Labels

```
DEFINE item / F=myfmt. "mylabel"
```

SAS Code -

```
PROC REPORT data=patents spanrows;

COLUMN state county patents population age education income;

DEFINE state / ORDER;

DEFINE county / ORDER;

DEFINE patents / "Patents" F=COMMA15.;

DEFINE population / "Population" F=COMMA15.;

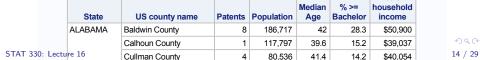
DEFINE education / "% >= Bachelor";

DEFINE income / F=DOLLAR15.;

RUN:
```

\_ SAS Code \_\_\_

Median



Summary Report with GROUP

•0000

Overview

Overview

Detail Report

Detail Report

Summary Report with GROUP

Summary Report with ACROSS

Mor

15 / 29

Summary Report with ACROSS

More

# **GROUP** example

```
PROC REPORT DATA = patents SPANROWS;

COLUMN region division patents;

DEFINE region / GROUP;

DEFINE division / GROUP;

RUN;

SAS Code
```

region	division	Number of patents
Midwest	East North Central	13126
	West North Central	5630
Northeast	Middle Atlantic	13841
	New England	8656
South	East South Central	1601
	South Atlantic	11427
	West South Central	8050
West	Mountain	6876
	Pacific	34771

#### GROUP usage

- summarizes data
- collapses observations with same values
- places values on rows
- orders rows
- suppresses repetitious printing



### Default statistics

#### Defaults for numeric variables:

00000

- ANALYSIS usage
- SUM statistic

```
Equivalent Statements —
DEFINE patents / "Patents";
DEFINE patents / ANALYSIS "Patents";
DEFINE patents / ANALYSIS SUM "Patents";
DEFINE patents / SUM "Patents";
          Equivalent Statements -
```

# Statistics keywords

Overview

CSS	CV	MAX	MEAN	MIN
MODE	N	NMISS	RANGE	STDEV
STDERR	SUM	SUMWGT	USS	VAR
PCTN	PCTSUM			
MEDIAN P50	P1	P5	P10	P25 Q1
P75 Q3	P90	P95	P99	QRANGE

Summary Report with GROUP

00000

4 D > 4 B > 4 B > 4 B > 9 Q P

Overview

```
SAS Code
PROC REPORT DATA = patents SPANROWS;
   COLUMN region division N PCTN patents, (SUM MEAN) income;
   DEFINE region / GROUP;
   DEFINE division / GROUP;
   DEFINE patents / ANALYSIS "Patents";
  DEFINE income / ANALYSIS MEAN "Ave Income" F=DOLLAR10.;
   DEFINE PCTN / "Percent" F=PERCENT8.1;
   DEFINE MEAN / "Mean" F=COMMA10.1;
   DEFINE SUM / "Sum" F=COMMA10.;
RUN;
                         SAS Code
```

Summary Report with GROUP

00000

- ▶ N and PCTN can be specified COLUMN statement
- All other statistics must be associated with a numeric variable
  - Single statistic: specify in DEFINE
  - Multiple statistics: specify with comma in COLUMN

STAT 330: Lecture 16 19 / 29

Summary Report with ACROSS

# Specifying statistics, example output

	Patents					
region	division	N	Percent	Sum	Mean	Ave Income
Midwest	East North Central	138	17.1%	13,126	95.1	\$49,746
	West North Central	55	6.8%	5,630	102.4	\$53,606
Northeast	Middle Atlantic	100	12.4%	13,841	138.4	\$55,697
	New England	37	4.6%	8,656	233.9	\$59,568
South	East South Central	64	7.9%	1,601	25.0	\$44,716
	South Atlantic	192	23.8%	11,427	59.5	\$50,498
	West South Central	92	11.4%	8,050	87.5	\$47,738
West	Mountain	52	6.4%	6,876	132.2	\$47,901
	Pacific	78	9.7%	34,771	445.8	\$52,987

Summary Report with GROUP

Overview

Overview

Detail Report

Detail Report

Summary Report with GROU

Summary Report with ACROSS

Mor

Summary Report with ACROSS

•0000

More

# ACROSS example

\_ SAS Code \_

PROC REPORT DATA = patents SPANROWS;
COLUMN region division N edu25 patents;

DEFINE region / GROUP;

DEFINE division / GROUP;
DEFINE edu25 / ACROSS;

DEFINE patents / "Patents" F=COMMA15.; RUN:

SAS Code \_\_\_

#### ACROSS usage

- summarizes data
- collapses
   observations
   with same
   values
- places ordered values on columns
- default statistic is N

				edu	125	
	region	division	N	0	1	Patents
	Midwest	East North Central	138	76	62	13,126
		West North Central	55	13	42	5,630
	Northeast	Middle Atlantic	100	46	54	13,841
Lectu	re 16	New England	37	5	32	8,656

### Analysis variables within ACROSS

\_ SAS Code

```
PROC REPORT DATA = patents SPANROWS;

COLUMN region division N edu25,patents;

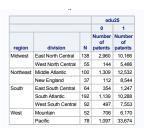
DEFINE region / GROUP;

DEFINE division / GROUP;

DEFINE edu25 / ACROSS;

DEFINE patents / F=COMMA15.;

RUN;
```



SAS Code \_\_\_\_

SAS Code \_\_

PROC REPORT DATA = patents SPANROWS;

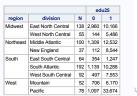
COLUMN region division N patents,edu25;

DEFINE region / GROUP;

DEFINE division / GROUP;

DEFINE edu25 / ACROSS;

DEFINE patents / " " F=COMMA15.;



## Multiple analysis variables within ACROSS

```
PROC REPORT DATA = patents SPANROWS;

COLUMN region division N edu25, (patents income);

DEFINE region / GROUP;

DEFINE division / GROUP;

DEFINE edu25 / ACROSS;

DEFINE patents / "Patents" SUM F=COMMA15.;

DEFINE income / "Income" MEAN F=DOLLAR15.;

RUN;
```

SAS Code

SAS Code -

			edu25				
			0		•	1	
region	division	N	Patents	Income	Patents	Income	
Midwest	East North Central	138	2,960	\$44,660	10,166	\$55,981	
	West North Central	55	144	\$44,922	5,486	\$56,294	
Northeast	Middle Atlantic	100	1,309	\$45,682	12,532	\$64,229	
	New England	37	112	\$44,829	8,544	\$61,871	



24 / 29

East South Central

354 \$41,679

1.247 \$49.154

# Multiple statistics for analysis variables within ACROSS

Use commas and parentheses to nest multiple statistics for an analysis variable within an across variable:

```
COLUMN AcrossVar, AnalysisVar, (stat1 stat2);
```

Use commas and parentheses to nest multiple statistics for multiple analysis variables within an across variable:

```
COLUMN AcrossVar, (AnalysisVar1 AnalysisVar2), (stat1 stat2);
```

Overview

#### Discussion

Overview

Number of patents						
region						
Midwest	west Northeast South		West			
18756	22497	21078	41647			

```
PROC REPORT DATA = patents;

COLUMN Patents ? Region;

DEFINE Region / ?;

RUN;

SAS Code
```

### Fill in the ?:

1. \* ACROSS

More

- 2. , ACROSS
- 3. \* GROUP
- 4. , GROUP

STAT 330: Lecture 16 26 / 29

Summary Report with GROUP

Overview

Detail Report

イロト (部) (を) (を)

Summary Report with ACROSS

More •0000

#### **Breaks**

```
PROC REPORT DATA = patents SPANROWS;

COLUMN region division N patents income;

DEFINE region / GROUP;

DEFINE division / GROUP;

DEFINE patents / ANALYSIS "Patents" F=COMMA10.;

DEFINE income / ANALYSIS MEAN "Ave Income" F=DOLLAR10.;

BREAK AFTER region / SUMMARIZE;

RBREAK AFTER / SUMMARIZE;

RUN;
```

SAS Code \_\_\_



## More things you can do with PROC REPORT

	>=25% with Bachelor's degree						
	Yes			No			
Region	N	Sum	Mean	N	Sum	Mean	Difference in Sums
Midwest	104	15,652	150.5	89	3,104	34.9	12,548
Northeast	86	21,076	245.1	51	1,421	27.9	19,655
South	155	19,088	123.1	193	1,990	10.3	17,098
West	72	39,844	553.4	58	1,803	31.1	38,041
All	417	95,660	229.4	391	8,318	21.3	87,342

- Highlight cells
- Customize break lines
- Calculate variables. that aren't in the input data set

See SAS code corresponding to lecture for full details.

Overview

# Syntax to calculate a new variable

```
SAS Code -
PROC REPORT DATA = mydata;
   COLUMN var1 var2 newvar:
   DEFINE var1 / analysis;
  DEFINE var2 / analysis;
   DEFINE newvar / COMPUTED :
   COMPUTE newvar:
       newvar = expression;
   ENDCOMP:
RUN;
            SAS Code _
```

There are many ways to write the *expression*. One way is to use | \_Cn\_ | where n is the column number.

### PROC REPORT vs PROC TABULATE

Detail Report

	PROC REPORT	PROC TABULATE
Create summary tables	<b>√</b>	<b>√</b>
Create detail tables	<b>√</b>	X
Lines between groups	$\checkmark$	X
Calculate new item	$\checkmark$	X
Multiple nested variables	X	<b>√</b>
Statistics options	less	more



STAT 330: Lecture 16

Overview

More 0000