Detail Report Summary Report with GROUP Summary Report with ACROSS More Overview

#### PROC REPORT

Shannon Pileggi

**STAT 330** 

1 / 31

STAT 330: Lecture 16

Summary Report with GROUP

**OUTLINE** 

Overview

Overview

Detail Report

Detail Report

Summary Report with GROUP

Summary Report with ACROSS

More

STAT 330: Lecture 16

Overview

2 / 31

Summary Report with ACROSS More

Detail Report

Summary Report with GROUP

Summary Report with GROUP

Summary Report with ACROSS

Summary Report with ACROSS

More

More

#### Overview

Detail Report

Overview

•000

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

#### 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.

メロトメ劇トメミトメミト ヨ 3 / 31

STAT 330: Lecture 16

STAT 330: Lecture 16

4 / 31

Overview Detail Report Summary Report with GROUP Summary Report with ACROSS More 0000

# **Syntax**

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

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

5 / 31

Detail Report Summary Report with GROUP Summary Report with ACROSS Overview More 000

# **DEFINE** usages

DEFINE var1 / usage;

Usage	Detail	Summary	Description
DISPLAY	<b>√</b>	Х	creates 1 row per obs
ORDER	$\checkmark$	X	creates 1 row per obs, ordered
ANALYSIS	<b>√</b>	<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

STAT 330: Lecture 16 6 / 31

Overview Detail Report Summary Report with GROUP Summary Report with ACROSS More •0000000

STAT 330: Lecture 16

#### Detail Report

Overview Detail Report Summary Report with GROUP Summary Report with ACROSS More 0.000000

# Getting started

PROC REPORT DATA = patents ; RUN ;

\_ SAS Code -

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: DISPLAY for character vars ANALYSIS for numeric vars

STAT 330: Lecture 16 8 / 31

**▼ロト ◆母 ト ◆ 恵 ト ◆ 恵 ・ 夕 久 ○** STAT 330: Lecture 16

Detail Report Summary Report with GROUP Summary Report with ACROSS More Overview 0000000

# Getting started - example output

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

PROC REPORT DATA = patents ; COLUMN state county patents; DEFINE state / DISPLAY ; DEFINE county / DISPLAY; DEFINE patents / ANALYSIS ; RUN ;

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

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

4回 > 4回 > 4 直 > 4 直 > 1 更 のQで 9 / 31

STAT 330: Lecture 16

Overview

Summary Report with GROUP

Summary Report with ACROSS

More

Detail Report

00000000

Summary Report with GROUP

Summary Report with ACROSS

More

#### Discussion

Detail Report

00000000

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

# 00000000

Summary Report with GROUP

# ORDER usage and SPANROWS

Detail Report

Overview

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
	Mobile County	40

Summary Report with ACROSS

10 / 31

STAT 330: Lecture 16

Overview

#### Discussion

\_ SAS Code \_

```
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?

US county name	Number 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

◆ロト ◆個 ト ◆ 恵 ト ◆ 恵 ・ 夕 へ ○

◆□▶◆□▶◆□▶◆□▶ □ り STAT 330: Lecture 16

STAT 330: Lecture 16

Overview Detail Report Summary Report with GROUP Summary Report with ACROSS More 0000 00000 000000 000000

# Spanning Column Headings

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

```
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				Demo	graphics	
	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
STAT 330: Lect	ure 16	Cullman County	4	80536	41.4	14.2	40054

 Overview
 Detail Report
 Summary Report with GROUP
 Summary Report with ACROSS
 More

 0000
 00000000
 000000
 000000
 000000

Overview

Detail Report

Summary Report with GROUP

Summary Report with ACROSS

More

4 D > 4 B > 4 E > 4 E > 9 Q ()

 Overview
 Detail Report
 Summary Report with GROUP
 Summary Report with ACROSS

 0000
 000000●
 000000
 000000

#### 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 \_\_

	State	US county name	Patents	Population	Median Age	% >= Bachelor	Median household income
	ALABAMA	Baldwin County	8	186,717	42	28.3	\$50,900
		Calhoun County	1	117,797	39.6	15.2	\$39,037
STAT 330: Lectu	re 16	Cullman County	4	80,536	41.4	14.2	\$40,054

√ Q (~ 14 / 31

CDOUD

Detail Report

Overview

13 / 31

Summary Report with GROUP

Summary Report with ACROSS

More 0000

Number

# **GROUP** example

PROC REPORT DATA = patents SPANROWS;

COLUMN region division patents;

DEFINE region / GROUP;

DEFINE division / GROUP;

RUN;

SAS Code

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

#### GROUP usage

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

4□ > 4₫ > 4월 > 4월 > 월

STAT 330: Lecture 16

STAT 330: Lecture 16

Summary Report with GROUP Summary Report with ACROSS Overview Detail Report More 000000

#### Default statistics

Defaults for numeric variables:

- ► ANALYSIS usage
- ► SUM statistic

```
_ Equivalent Statements ___
DEFINE patents / "Patents";
DEFINE patents / ANALYSIS "Patents";
DEFINE patents / ANALYSIS SUM "Patents";
DEFINE patents / SUM "Patents";
         _ Equivalent Statements ___
```

STAT 330: Lecture 16

Summary Report with GROUP

17 / 31

4□ > 4₫ > 4 ≧ > 4 ≧ > 9 < 0</p>

Summary Report with ACROSS

More

19 / 31

# Specifying statistics

STAT 330: Lecture 16

Detail Report

```
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;
```

- ▶ 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

Summary Report with GROUP Summary Report with ACROSS Overview Detail Report More 000000

# Statistics keywords

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

STAT 330: Lecture 16 18 / 31

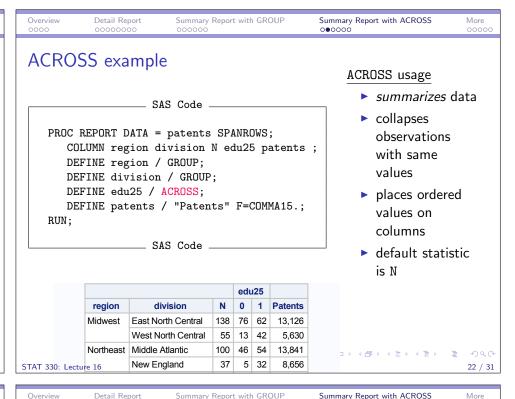
Overview Detail Report Summary Report with GROUP Summary Report with ACROSS More 00000

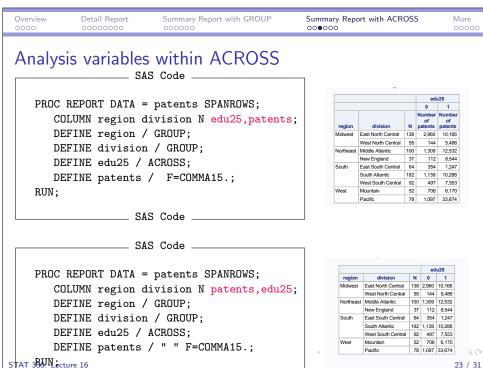
# Specifying statistics, example output

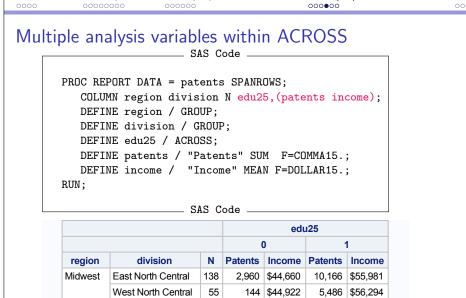
					Patents	
region	division	N	Percent	Sum	Mean	Ave Income
Midwest	East North Central	138	17.1%	13,126	95.1	\$49,746
	West North Central	55	6.8%	5,630	102.4	\$53,606
Northeast	Middle Atlantic	100	12.4%	13,841	138.4	\$55,697
	New England	37	4.6%	8,656	233.9	\$59,568
South	East South Central	64	7.9%	1,601	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

4□▶ 4□▶ 4 = ▶ 4 = ▶ 9 < 0</p> STAT 330: Lecture 16









100

1,309 \$45,682

112 \$44.829

354 \$41.679

12.532 \$64.229

8.544 \$61.871

1.247 \$49.154

24 / 31

Northeast | Middle Atlantic

STAT 330: Lect South

New England

East South Central

 Overview
 Detail Report
 Summary Report with GROUP
 Summary Report with ACROSS
 More

 0000
 0000000
 00000€0
 00000€0
 00000€0

# 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 Detail Report Summary Report with GROUP Summary Report with ACROSS More

Overview

Detail Report

Summary Report with GROUP

Summary Report with ACROSS

More

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

PROC REPORT DATA = patents;
COLUMN Patents ? Region;
DEFINE Region / ?;
RUN;

SAS Code 4.,

808 103.978 \$50.941

Fill in the ?:

1. \* ACROSS

2. , ACROSS

3. \* GROUP

4. , GROUP

28 / 31

 STAT 330: Lecture 16
 26 / 31

Overview Detail Report Summary Report with GROUP Summary Report with ACROSS More 00000 **Breaks** SAS Code 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 \_ 52 6.876 \$47,901 

Detail Report Summary Report with GROUP Summary Report with ACROSS More Overview 00000

## 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.

<ロ > → □ > → □ > → □ > → □ > → ○ へ ○ 29 / 31

STAT 330: Lecture 16

Summary Report with GROUP Overview Detail Report

Summary Report with ACROSS More 00000

#### PROC REPORT vs PROC TABULATE

	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

◆□▶ ◆□▶ ◆□▶ ◆□ ● りゅ○

STAT 330: Lecture 16

Detail Report Summary Report with GROUP Summary Report with ACROSS Overview 00000

## 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.

30 / 31

STAT 330: Lecture 16