

**Please login...**

- Take a seat at one of the work stations
- Login with your HawkID
- Locate SAS 9.3 in the Start Menu
  - Start / All Programs / SAS / SAS 9.3 (English)
- Make SAS “go”
  - Raise your hand if you need assistance

Day 2 of SAS® Summer Institute 2014

**will begin soon...**

**Procedures for Data Insight**

University of Iowa SAS User Group  
SAS Summer Institute  
August 20, 2014

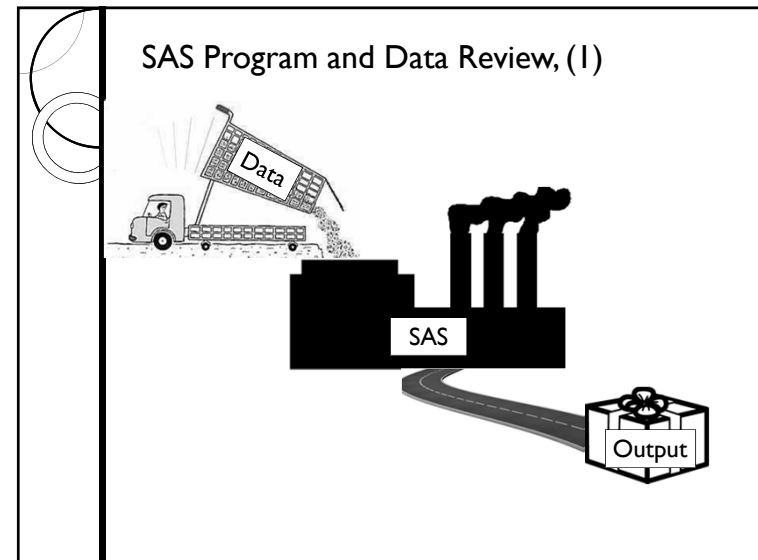
Presented by: Fred Ullrich  
Department of Health Management and Policy  
College of Public Health  
Liz Uribe  
Clinical Trials Data Management Center  
Beth Cook  
Clinical Trials Data Management Center

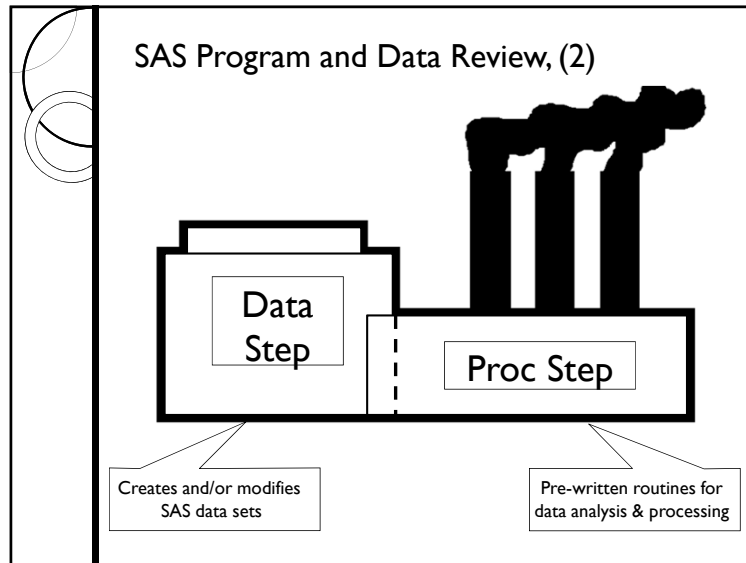
2<sup>nd</sup> day morning session (8:07 – 11:34)

**overview**

**“Procedures for Data Insight”**

<b>Part 1:</b>	<b>Part 2:</b>
<ul style="list-style-type: none"><li>• SAS program and data review</li><li>• PROC Overview</li><li>• Some PROC specifics</li></ul>	<ul style="list-style-type: none"><li>• Elements of style</li><li>• More PROC specifics</li><li>• Intro to data import/export</li></ul>





### SAS Program and Data Review, (3)

#### Sample Data Steps

**“inline” data**

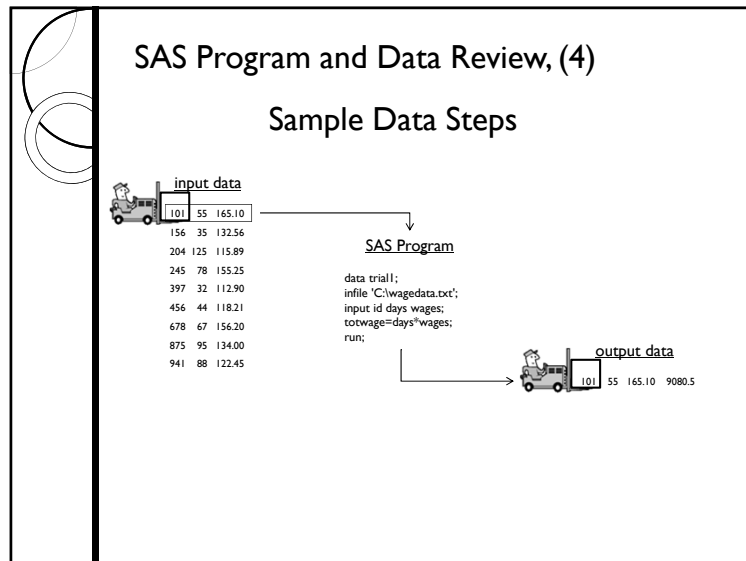
```
DATA wages;
INPUT id days wages;
totwage=days*wages;
CARDS;
101 55 165.10
156 35 132.56
204 125 115.89
245 78 155.25
397 32 112.90
456 44 118.21
678 67 156.20
875 95 134.00
941 88 122.45
;
run;
```

**“external” data**

```
data trial1;
infile 'C:\wagedata.txt';
input id days wages;
totwage=days*wages;
run;
```

c:\wagedata.txt

```
101 55 165.10
156 35 132.56
204 125 115.89
245 78 155.25
397 32 112.90
456 44 118.21
678 67 156.20
875 95 134.00
941 88 122.45
```



### SAS Program and Data Review, (5)

#### SAS Log Window

```
1 data wages;
2 infile 'C:\wagedata.txt';
3 input id days wages;
4 totwage=days*wages;
5 run;
```

**NOTE:** 9 records were read from the infile 'C:\junk\wagedata.txt'.  
The minimum record length was 18.  
The maximum record length was 18.

**NOTE:** The data set WORK.WAGES has 9 observations and 4 variables.

**NOTE:** DATA statement used (Total process time):  
real time 0.00 seconds  
cpu time 0.00 seconds

- “WORK” is the name of a library
- Libraries can contain many datasets
- Datasets in the “WORK” library are deleted at the end of your SAS session (i.e. they are temporary)
- To make your datasets “permanent” use a different library name
- Before you can do that, the library (name) must be associated with a directory (folder) accessible from your computer

## SAS Program and Data Review, (6)

*name you give the library*

```
LIBNAME mylib 'C:\Research\Data';
```

*LIBNAME statement*                      *location of the library*

**Sample Program:**  
**Create permanent dataset**

```
libname mylib 'c:\research\data';
data mylib.trial1;
infile 'C:\wagedata.txt';
input id days wages;
run;
```

**Sample Program:**  
**Use permanent dataset**

```
libname mylib 'c:\research\data';
proc print data=mylib.trial1;
run;
```

## SAS Program and Data Review, (7)

SAS comes with data libraries

## SAS Procedures\*

“Base” software	SAS/Stat software
• APPEND    • HTTP    • PWENCODE	• ACECLUS    • INBREED    • PLM    • SURVEYSELECT
• BMDP    • IMPORT    • RANK	• ANOVA    • KDE    • PLS    • TPSPLINE
• CALENDAR    • INFOMAPS    • REGISTRY	• BOXPLOT    • KRIGE2D    • POWER    • TRANSREG
• CATALOG    • ITEMS    • RELEASE	• CALIS    • LATTICE    • PRINCOMP    • TREE
• CDISC    • JAVAINFO    • REPORT	• CANCORR    • LIFEREG    • PRINQUAL    • TTEST
• CHART    • MEANS    • SCAPROC	• CANDISC    • LIFETEST    • PROBIT    • VARCLUS
• CIMPORT    • METADATA    • SOAP	• CATMOD    • LOESS    • QUANTREG    • VARCOMP
• COMPARE    • METALIB    • SORT	• CLUSTER    • LOGISTIC    • REG    • VARIOGRAM
• CONTENTS    • METAOPERATE    • SOURCE	• CORRESP    • MCMC    • ROBUSTREG
• CONVERT    • MIGRATE    • SQL	• DISCRIM    • MDS    • RSREG
• COPY    • OPTIONS    • STANDARD	• DISTANCE    • MI    • SCORE
• CORR    • OPTLOAD    • SUMMARY	• FACTOR    • MIANALYZE    • SEQDESIGN
• CPORT    • OPTSAVE    • TABULATE	• FASTCLUS    • MIXED    • SEQTEST
• DATASETS    • PDS    • TAPECOPY	• FREQ    • MODECLUS    • SIM2D
• DBCSTAB    • PDSCOPY    • TAPELABEL	• GAM    • MULTTEST    • SIMNORMAL
• DISPLAY    • PLOT    • TEMPLATE	• GENMOD    • NESTED    • STDIZE
• DOCUMENT    • PMENU    • TIMEPLOT	• GLMMIX    • NLIN    • STEPOISC
• EXPLODE    • PRINT    • TRANSPOSE	• GLM    • NL MIXED    • SURVEYFREQ
• EXPORT    • PRINTTO    • TRANSTAB	• GLMMOD    • NPARIWAY    • SURVEYLOGISTIC
• FCOMP    • PROTO    • UNIVARIATE	• GLMPOWER    • ORTHOREG    • SURVEYMEANS
• FORMS    • PRTRDEF    • VAXTOINTEG	• GLMSELECT    • PHREG    • SURVEYPHREG
• FREQ    • PRTEXP    • XSL (Preproduction)	• HPMIXED    • PLAN    • SURVEYREG

\*in two of 20 SAS product lines

## Basic PROC Syntax

“PROC”                      *Options (may be many)*

```
PROC Print data=sashelp.heart;
```

*Procedure name*

*Additional procedure statements* { `var sex weight;`

```
run;
```

## PROC PRINT

- The PRINT procedure is used to organize and display data in the 'output' window as lists or case reports.
- PRINT has many options for controlling the appearance of data in reports.
- PRINT lists data, but has some selection, grouping and summary capabilities.

## PROC CONTENTS

- The CONTENTS procedure does much more than generate text output.
  - Review: default text output order is to list variable names alphabetically
    - The VARNUM option sets the list to order by column position (as you would see it in the display manager view of the dataset).
- NOPRINT: Making use of 'metadata'
  - Simple lists (lists of datasets to process)
  - Simple lines (lines of repetitive code to create and submit)

## PROC SORT

- The SORT procedure is used to organize datasets typically in preparation for merging related data, or putting the data into a sequence that will match the 'BY' processing used in a future procedure.
- Sorts can be ASCENDING or DESCENDING and can include one, or all variables in a dataset.
- Sorts can create new datasets
- Sorts can be used to eliminate duplication (but this can be tricky)

## PROC FREQ

- The FREQ or FREQUENCY procedure is useful for examining categorical variables. It simply reports frequency counts and calculates some simple sums and percentages.
- Tables can be crossed: variables a X b
- Any table request can include a BY grouping request provided the data are pre-sorted.

## PROC MEANS

- Use for descriptive statistics
- Stats for All records is default.
- Can use BY or CLASS statements to report by categories in other variables (ie, Gender, Year, City, etc...).

--break--

Time for a 10-minute break....

Make sure you have "saved" your program

[www.uiowa.edu/~uisug](http://www.uiowa.edu/~uisug)

## Elements of style

sample program

```
data triall;
infile 'C:\wagedata.txt';
input id days wages;
wage_rate=wages/days;
if wage_rate>20 then lvl='hi';
   else lvl='lo';
run;
```

## Elements of style

(don't do this)

```
data triall;infile 'C:\wagedata.txt'; input id days wages/wage_rate
=wages/days;if wage_rate>20 then lvl='hi';else lvl='lo';run;
```

```
data triall; Infile 'C:\wagedata.txt';
Input      id
days      wages;
wage_rate  =
wages/     days;
if         wage_rate>20
then lvl='hi'; else  lvl='lo'; run;
```

```
data
triall;
Infile
'C:\wagedata.txt
';
input
id
days
wages;
wage_rate
=
wages/
days;
if
wage_rate>20
then
lvl=
'hi';
else
lvl='lo';
run;
```

### Elements of style (do this)

```

/* this is a sample program used to
demonstrate some of the basic
elements of programming style */

data triall;
infile 'C:\wagedata.txt';
input id days wages;
wage_rate=wages/days;

* "20" is industry standard for hi;
if wage_rate>20 then lvl='hi';
  else lvl='lo';
run;
    
```

- Large block comment at beginning describing program and purpose
- One statement per line
- Blank lines to separate sections of the program
- Short comment to explain code
- Indenting subordinate statements

### PROC FORMAT (build)

- Lets you define your own informats and formats for variables.

```

PROC Format;
  value <$> formatname;
  1='male';
  2='female';
run;
    
```

Annotations for PROC FORMAT:

- "PROC" label above PROC Format;
- Value Statement label on the left of value <\$> formatname;
- There could be options here label pointing to the space between PROC and Format;
- char format? label pointing to <\$> ;
- SAS "normal" name rules label pointing to formatname;
- data value "displayed" value label pointing to the values 1 and 2;

### Merging data

- The MERGE statement is used within a dataset to combine two or more SAS datasets.
- One or more datasets can be merged by a 'key' variable, or group of variables that creates a unique key
  - SAS will let you merge with repeats on the key, and it will note this in the log.

### Merging data

Patient Data			Visit Data		
patno	lname	sex	patno	visit #	wt
11	Jones	M	11	1	137
66	Smith	M	11	2	135
33	Brown	F	33	1	186
55	Harris	F	33	2	182
44	Anderson	F	33	3	176
22	Collins	M	66	1	157

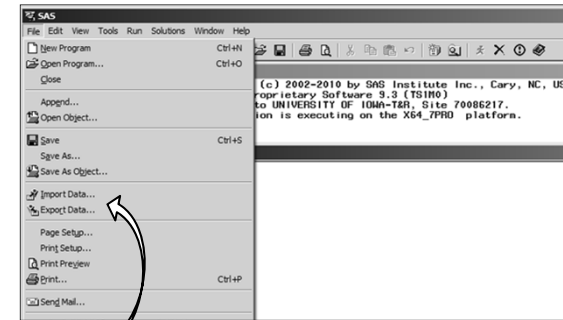
↓

"Merged" Data					
patno	lname	sex	visit #	wt	
11	Jones	M	1	137	
11	Jones	M	2	135	
22	Collins	M	.	.	
33	Brown	F	1	186	
33	Brown	F	2	182	
33	Brown	F	3	176	
44	Anderson	F	.	.	
55	Harris	F	.	.	
66	Smith	M	1	157	

## Import/Export Data

- SAS can import data from, and export data to, many different formats
  - MS-Excel
  - MS-Access
  - .csv
  - SPSS
  - Stata
  - many others
- A variety of methods for importing/exporting
- Best approach depends on variety of factors
  - Operating system (Linux, Windows, 32/64-bit)
  - SAS version (9.2, 9.3, 32/64-bit)
  - Originating/destination software (Excel, .csv, SPSS)
- Use the Wizard
  - Be careful, pay attention

## Import/Export Data



**Wizards!**

## PROC SQL

### Structured Query Language (SQL)

A language used for managing data in many different computer applications (primarily database applications). It has been available in SAS since the late 1980's and can be used for a wide variety of purposes including nearly everything we have done this morning.

#### Syntax:

```
PROC SQL;  
  SQL statements;  
quit;
```

## PROC TRANSPOSE

- The TRANSPOSE procedure is what we use to flip data on its side.
- It is recommended to do this in small chunks so that it is easy for you to understand and explain to others.
- With experience you can transpose multiple variables simultaneously.

## PROC TRANSPOSE

### Raw Data

Variable	Bob	Tim	Kim	Ann	Pat
age	23	25	21	26	24
height	62	61	66	71	69
weight	120	125	160	220	205
score	88	93	100	75	98

### "Transposed" Data

Name	Age	Height	Weight	Score
Bob	23	62	120	88
Tim	25	61	125	93
Kim	21	66	160	100
Ann	26	71	220	75
Pat	24	69	205	98

Review and questions

THANKS to our SAS experts  
for their onsite assistance

-- break for lunch --

- **Thank you** for participating in our session.
- We hope these materials will be helpful.
  
- Enjoy a lunch break and return for the next session featuring SAS Enterprise Guide.