

CP5804: SAS Week 3

Collaborative Tutorial

Session

Presented by Sisi

SAS Essentials

➤ Installation:

❑ SAS University Edition: https://www.sas.com/en_au/software/university-edition/download-software.html

❑ VirtualBox: https://www.virtualbox.org/wiki/Download_Old_Builds_5_2,
(Tip: please download 5.2.X builds instead of the latest version)

➤ Weekly materials:

❑ Assessments → SAS Self-Training Content

❑ Assessments → Assessment 5 (SAS Coursework)

Troubleshooting VirtualBox

- Check if you have enabled VT-X in BIOS in startup. Could refer to the link for details: https://docs.fedoraproject.org/en-US/Fedora/13/html/Virtualization_Guide/sect-Virtualization-Troubleshooting-Enabling_Intel_VT_and_AMD_V_virtualization_hardware_extensions_in_BIOS.html;
- If VT-X is already enabled, check if Hyper_V is turned off. Try running the following in Command Prompt (Run): **dism.exe /Online /Disable-Feature:Microsoft-Hyper-V**. Could refer to the link for details: <https://superuser.com/questions/1153470/vt-x-is-not-available-but-is-enabled-in-bios>;
- If the issue still exists, possibly try using VMware instead of VirtualBox.

SAS Assessments

➤ Submission due:

Week 1 - 3 due on **Wednesday of Week 4 (01/04/2020)**

Week 4 due on **Wednesday of Week 5 (08/04/2020)**

Week 5 due on **Wednesday of Week 6 (15/04/2020)**

Week 6 due on **Wednesday of Week 7 (22/04/2020)**

Week 7 non-graded

SAS Assessments

➤ Marking criteria (2/2):

☐ For activity answer sheets less than or equal to 10 questions:

- 0 – not submitted or submit non-relevant results
- 1 – more than half, but all attempted
- 1.5 – more than 2, less than half
- 2 – less than 2 incorrect answer

☐ For activity answer sheets more than 10 questions:

- 0 – not submitted or submit non-relevant results
- 1 – more than half, but all attempted
- 1.5 – more than 4, less than half
- 2 – less than 4 incorrect answer

SAS Help

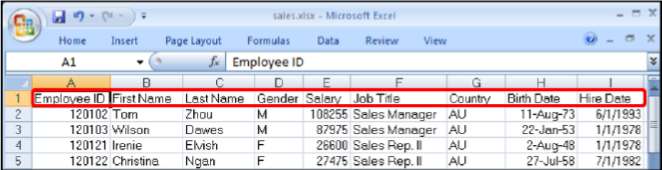
- Q&A for SAS
- Setup Instructions: Assessments → SAS Self-Training Content → Video Instructions for SAS Setup
- Demo

SAS Week 3 Activity Answer Sheet

➤ 7.01:

- ☐ Vary among different SAS working environments; VALIDVARNAME option
- ☐ PG_Ch07, page 5, slide 9.

SAS Variable Names



Excel column headings are used as variable names.

- The SAS windowing environment replaces blanks and special characters with underscores.
- By default, SAS Studio and SAS Enterprise Guide allow blanks and special characters in variable names.


9

- In the SAS windowing environment, the VALIDVARNAME= option is set to V7 by default. V7 does not allow variable names to contain special characters.
- In SAS Studio and SAS Enterprise Guide, the VALIDVARNAME= option is set to ANY by default. ANY enables variable names to contain special characters. If a variable name contains special characters, the variable name must be expressed as a SAS name literal.

SAS Week 2 Activity Answer Sheet

➤ 7.02:

- ☐ Execute p107d02a, p107d02b;
- ☐ PG_Ch07, page 13, slides 22-23;
- ☐ SAS UE does not support excel or pcfiles engine;
- ☐ Xlsx engine does not have name ranges.



Subsetting a Worksheet

Regardless of the LIBNAME engine used, the PRINT procedure can be used to display a subset of the worksheet.

```
proc print data=orionx.'Australia$'n noobs;  
  where Job_Title contains 'IV';  
  var Employee_ID Last_Name  
      Job_Title Salary;  
run;
```

EXCEL and PCFILES engines

```
proc print data=orionx.Australia noobs;  
  where Job_Title contains 'IV';  
  var Employee_ID Last_Name  
      Job_Title Salary;  
run;
```

XLSX engine

23

p107d02a
p107d02b

If the sheet name contains special characters or embedded blanks, a name literal must be used to refer to the sheet, even when you use the XLSX engine.

SAS Week 3 Activity Answer Sheet

➤ 8.03:

- ❑ PG_Ch08, page 12, slide 23
- ❑ Compilation phase:
 - ❖ creates an input buffer to hold one record at a time from the raw data file
 - ❖ creates the program data vector (PDV) to hold one observation: variable name, type, byte size, no initial value
 - ❖ creates the descriptor portion of the output data set

sas
THE POWER TO KNOW.

Compilation Phase

During compilation, SAS does the following:

- scans the step for syntax errors
- translates each statement into machine language
- creates an *input buffer* to hold one record at a time from the raw data file

The diagram illustrates the Input Buffer as a sequence of slots representing bytes. The first record (labeled '1') occupies slots 1 through 16, ending with a null byte (0) at slot 16. The second record (labeled '2') starts at slot 17 and occupies slots 17 through 24, also ending with a null byte (0) at slot 24. Ellipses (...) follow the final slot, indicating further data.

Input Buffer																1							2
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6		8	9	0	...			


- creates the program data vector (PDV) to hold one observation
- creates the descriptor portion of the output data set

23

SAS Week 3 Activity Answer Sheet

➤ 8.04:

- ☐ PG_Ch08, page 25, slide 51;
- ☐ Default for *input* statement **8**;
- ☐ **\$** to specify character;
- ☐ **No default** for the *length* statement;
- ☐ Sample: p108d02


The logo for SAS, featuring the letters 'sas' in a stylized font with a tagline 'The Power to Know'.

LENGTH Statement

The *LENGTH* statement defines the type and length of a variable.

LENGTH variable(s) <\$> length ...;

```
data work.subset;  
  length First_Name $ 12 Last_Name $ 18  
         Gender $ 1 Job_Title $ 25  
         Country $ 2;  
  infile "&path\sales.csv" dlm=',';  
  input Employee_ID First_Name $ Last_Name $  
         Gender $ Salary Job_Title $ Country $;  
run;
```

 Put the LENGTH statement before the INPUT statement.

51 p108d02

The LENGTH statement is used primarily for character variables.

SAS Week 3 Activity Answer Sheet

➤ 8.06:

- ☐ PG_Ch08, page 31, slide 66;
- ☐ Interpret the log after executing the program;
- ☐ Keywords: option a -**invalid**, option b -**incorrectly**.

sas THE POWER OF DATA

Data Errors

A data error occurs when a data value does not match the field specification. The following information is written to the SAS log:

- a note describing the error
- a column ruler
- the input record
- the contents of the PDV

```
NOTE: Invalid data for Salary in line 4 31-33.
RULE:  -----1-----2-----3-----4-----5-
4      120122,Christina,Ngan,Rep. II,n/a,AU 36
Employee_ID=120122 First=Christin Last=Ngan Job_Title=Rep. II Salary=.
Country=AU _ERROR_=1 _N_=4
```

A missing value is assigned to the corresponding variable, and execution continues.


66

Even though these are referred to as *data errors*, they generate notes, not error messages.
 Syntax errors stop the DATA step, whereas data errors enable processing to continue.

SAS Week 3 Activity Answer Sheet

➤ 8.07:

- ☐ Date *formats*: PG_Ch05, page 8-9, slides 15-16
- ☐ Date *Informats*: PG_Ch08, page 41, slide 86.

Copyright © 2014, SAS Institute Inc. All rights reserved.


SAS Informats

Use date informats to read and convert dates to SAS date values.

Informat	Raw Data Value	SAS Data Value
MMDDYY.	010160	0
	01/01/60	
	01/01/1960	
	1/1/1960	
DDMMYY.	311260	365
	31/12/60	
	31/12/1960	
DATE.	31DEC59	-1
	31DEC1959	

86

SAS Week 3 Activity Answer Sheet


➤ 8.08:

❑ PG_Ch08, page 40, slide 83;

❑ Help → SAS Product Documentation;

➤ 9.03:

❑ Refer to program **p109d02**.



SAS Informats

SAS informats have the following form:

`<$><informat><w>.`

\$	Indicates a character informat.
<i>informat</i>	Names the SAS informat or user-defined informat.
<i>w</i>	Specifies the width or number of columns to read or specifies the length of a character variable.
.	Is required syntax.


✎ The width is typically not used with list input because SAS reads each field until it encounters a delimiter.

83

SAS Week 3 Activity Answer Sheet

➤ 9.05:

- ❑ Hint: Length function; PG_Ch09, page 34, slide 71;
- ❑ List **at least two** possible ways to solve the problem.



Defining Character Variables

Set the length of the variable **Freq** to avoid truncation.

```
data work.bonus;  
  set orion.sales;  
  length Freq $ 12;  
  if Country='US' then do;  
    Bonus=500;  
    Freq='Once a Year';  
  end;  
  else if Country='AU' then do;  
    Bonus=300;  
    Freq='Twice a Year';  
  end;  
run;
```

LENGTH variable(s) <\$> length;

✍ It is a good practice to use a LENGTH statement anytime that you create a new character variable.

71 p109d08

