Name Lan Liu 75/100

Q1 "What type of data files are not considered raw data?"

D: SQL Tables

Q3 "Which of the following is a valid libref name that can be used to create a permanent SAS data set?"

A: working

Q4 "Which DATA statement will create a permanent SAS data set called DOGS assuming that all SAS libraries have been properly defined?"

C. DATA sasdata.dogs

Q11 "Which statement is synonymous with a DATALINES statement?"

C. CARDS

Q12 "Which SAS statement enables you to refer to an external raw data file?"

C. INFILE

Q14 "Assuming that the raw data are arranged in neat columns, what is an advantage of column input?"

```
D. All of the above (missing data indicated by spaces, embedded blanks, character data longer than 8 characters)
```

Q16 "Given this note in the SAS log, what could you add to fix the INPUT statement so that the ID variable would be read correctly including all digits and hyphens?"

```
INPUT ID GPA Age;
NOTE: Invalid data for ID in line 1 1-9.

RULE: ---+---2
1 5437-2212 3.84 21
ID=. GPA=3.84 Age=21 _ERROR_=1 _N_=1
C. An informat
```

C. All illioillat

Q18 "Which informat would be appropriate to read the value 07/04/1776?"

A. MMDDYY8



Answer should be B because Year has 4 digits.

Q30 "Explain the difference between using a LIBNAME statement versus using an INFILE statement."

LIBNAME creates permanent datasets and allow us to import multiple datasets from a directory instead of loading them one by one by typing their filename, which is rather unnecessary. The libname is a global statement that remains valid for the current SAS session.

INFILE can only be used for one file at a time to specify the location of the file you want to read.

You use a LIBNAME statement to reference a SAS data set, and an INFILE statement to reference an external raw data file.

Q34" Describe one advantage of using formatted input over column input."

Column input can be used when the data values are arranged in neatly defined columns and the columns do not contain embedded spaces. However, formatted input allows us to read data with special characters

With formatted input you can read nonstandard numeric data such as dates or dollar amounts with commas and dollar signs

Q45

- 45. The file CancerRates.dat contains data on the top 10 cancer sites in the United States from the Centers for Disease Control and Prevention (CDC) website. These statistics are condensed across genders and races. The variables are ranking, cancer site, and incidence rate per 100,000 people.
 - a. Open the raw data file CancerRates.dat in a simple editor such as WordPad. In a comment in your program, state the number of variables and observations.
 - b. Read the raw data file into SAS. View the log to verify that your data set has the same number of variables and observations as you stated in part a).
 - c. Print the data set.
 - d. Copy the CancerRates.dat data set to a different location such as your desktop or a flash drive and read it into SAS a second

time from that new location.

Α

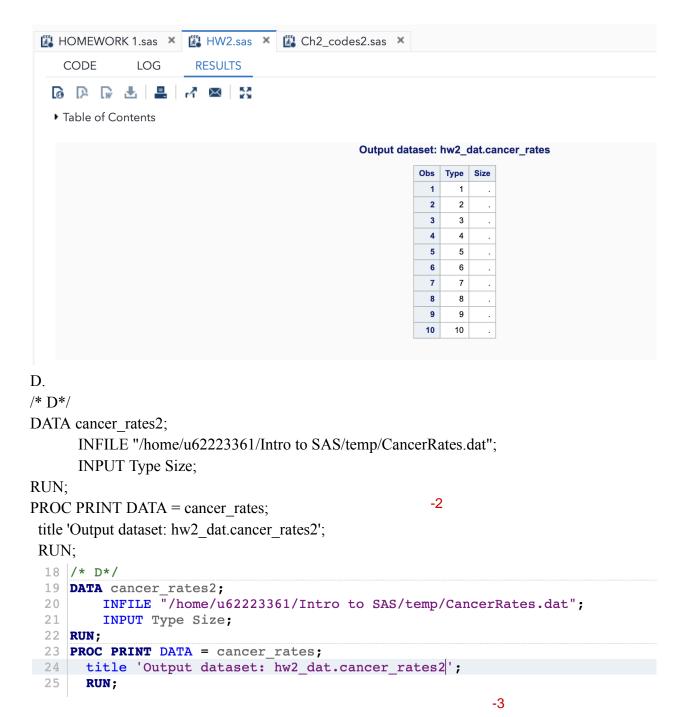
ID should be one of variables. -3

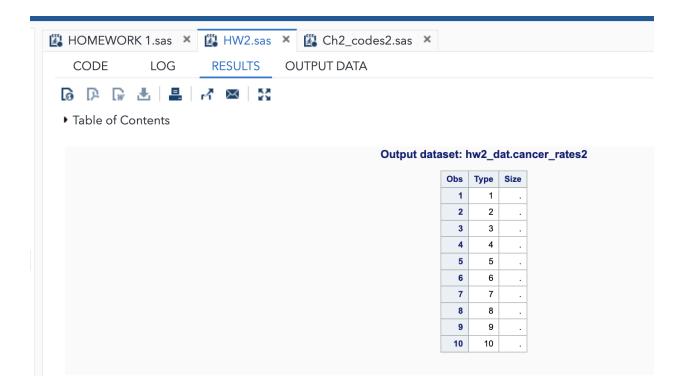
Number of variables: 2 (excluding ID/row number);

Number of observations: 10

```
В.
```

```
DATA cancer rates;
       INFILE "/home/u62223361/Intro to SAS/HW2/CancerRates.dat";
       INPUT Type Size;
RUN;
    8 DATA cancer rates;
           INFILE "/home/u62223361/Intro to SAS/HW2/CancerRates.dat";
    9
           INPUT Type Size;
  10
  11 RUN;
  🚜 HOMEWORK 1.sas 🗴 🚜 *HW2.sas 🗴 🚜 Ch2_codes2.sas 🗴
     CODE
                LOG
                         RESULTS
                                    OUTPUT DATA
                                                            Id is one of variable; use the proper variable
   names such the second variable as cancer
   ▼ Errors, Warnings, Notes
                                                            type/site; the second variable should be
        NOTE: Invalid data for Something in line 5 5-10.
                                                            character variable. -5
        NOTE: Invalid data for Something in line 6 5-12.
        NOTE: Invalid data for Something in line 7 5-11.
        NOTE: Invalid data for Something in line 8 5-7.
        NOTE: Invalid data for Something in line 9 5-10.
        NOTE: Invalid data for Something in line 10 5-11.
        NOTE: 10 records were read from the infile "/home/u6223361/Intro to SAS/HW2/CancerRates.dat".
       NOTE: The data set WORK.CANCER_RATES has 10 observations and 2 variables.
        NOTE: DATA statement used (Total process time):
C.
/* C*/
PROC PRINT DATA = cancer rates;
 title 'Output dataset: hw2_dat.cancer_rates';
 RUN;
 13 /* C*/
 14 PROC PRINT DATA = cancer rates;
 15
         title 'Output dataset: hw2 dat.cancer rates';
 16
         RUN;
                                                                     failed to read the data in. -3
```





Q54

- 54. Data on previous winners of the Oscars are stored in a Microsoft Excel file named Oscars.xlsx. The variables in this file are ID, year, host, best picture, best actor, best actress, best director, and best screenplay.
 - a. Examine the Microsoft Excel file Oscars.xlsx and read it into a permanent SAS data set using the IMPORT procedure.
 - Print a report that describes the contents of the data set including the attributes of the variables and data set.
 - c. In a comment in your program, discuss any limitations of the functionality of the resulting data set.
 - d. Print the Oscars.xlsx data file using the XLSX LIBNAME engine. In a comment in your program, discuss any limitations of using this method to read in the data.

A.

```
31 /* A*/
 32
 33 %web drop table(WORK.OSCAR);
 34
 35
 36 FILENAME REFFILE '/home/u62223361/Intro to SAS/HW2/Oscars.xlsx';
 37
 38 PROC IMPORT DATAFILE=REFFILE
 39
        DBMS=XLSX
 40
        OUT=WORK.OSCAR;
 41
        GETNAMES= NO;
 42
        INPUT
 43
            ID 2
 44
            YEAR 4.
            HOST $
 45
            BEST PIC $
 46
                                          PROC Import dose not take "Input" statement
            BEST ACTOR $
 47
            BEST ACTRESS $
 48
 49
            BEST DIRECTOR $
 50
            BEST SCREENPLAY $;
 51 RUN;
 52
 53 data OSCAR; set WORK.OSCAR;
 54 rename
 55 A = ID
 56 B = YEAR
 57 C= HOST
 58 D = BEST PIC
 59 E = BEST ACTOR
60 F = BEST ACTRESS
 61 G = BEST DIRECTOR
 62 H = BEST SCREENPLAY;
 63
 64 RUN;
 65
/* A*/
%web drop table(WORK.OSCAR);
FILENAME REFFILE '/home/u62223361/Intro to SAS/HW2/Oscars.xlsx';
PROC IMPORT DATAFILE=REFFILE
     DBMS=XLSX
     OUT=WORK.OSCAR;
     GETNAMES= NO;
     INPUT
```

ID 2

YEAR 4.
HOST \$
BEST_PIC \$
BEST_ACTOR \$
BEST_ACTRESS \$
BEST_DIRECTOR \$
BEST_SCREENPLAY \$;

RUN;

data OSCAR; set WORK.OSCAR; FILENAME REFFILE '/r

OSCAR; set WORK.OSCAR; REFFILE '/home/u63514490/SASIntro/Data/ch2/Oscars.xlsx';

rename
A = ID

PROC IMPORT DATAFILE=REFFILE

A - ID B = YEAR C = HOSTDBMS=XLSX
OUT=Oscars;
GETNAMES=no;

 $D = BEST_PIC$ RUN;

$$\begin{split} E &= BEST_ACTOR \\ F &= BEST_ACTRESS \end{split} \qquad \text{Try use the above to read the data in;} \end{split}$$

G = BEST_DIRECTOR H = BEST_SCREENPLAY;

RUN;

PROC PRINT DATA = OSCAR;

title 'Output dataset: hw2_dat.OSCAR';

RUN;



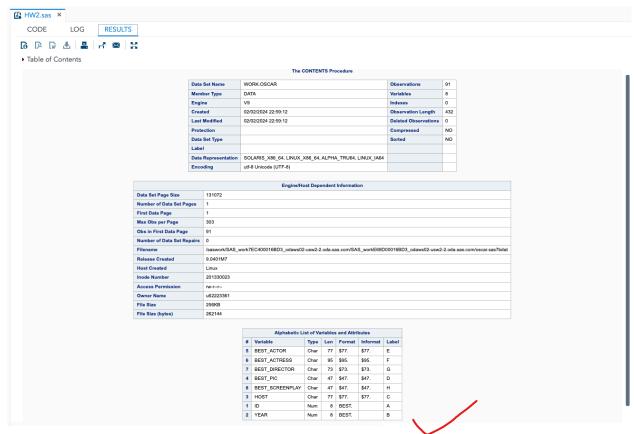
Table of Contents

					Output dataset: hy	w2_dat.OSCAR		
Obs	ID	YEAR	HOST	BEST_PIC	BEST_ACTOR	BEST_ACTRESS	BEST_DIRECTOR	BEST_SCREENPLAY
- 1	91	2019	No official host	Green Book	Rami Malek (Bohemian Rhapsody)	Olivia Colman (The Favourite)	Alfonso Cuarón (Roma)	Green Book
2	90	2018	Jimmy Kimmel	The Shape of Water	Gary Oldman (Darkest Hour)	Frances McDormand (Three Billboards Outside Ebbing, Missouri)	Guillermo del Toro (The Shape of Water)	Get Out
3	89	2017	Jimmy Kimmel	Moonlight	Casey Affleck (Manchester by the Sea)	Emma Stone (La La Land)	Damien Chazelle (La La Land)	Manchester by the Sea
4	88	2016	Chris Rock	Spotlight	Leonardo DiCaprio (The Revenant)	Brie Larson (Room)	Alejandro G. Iñárritu (The Revenant)	Spotlight
5	87	2015	Neil Patrick Harris	Birdman or (The Unexpected Virtue of Ignorance)	Eddie Redmayne (The Theory of Everything)	Julianne Moore (Still Alice)	Alejandro G. Iñárritu (Birdman or (The Unexpected Virtue of Ignorance))	Birdman or (The Unexpected Virtue of Ignorance)
6	86	2014	Ellen DeGeneres	12 Years a Slave	Matthew McConaughey (Dallas Buyers Club)	Cate Blanchett (Blue Jasmine)	Alfonso Cuaron (Gravity)	Her
7	85	2013	Seth MacFarlane	Argo	Daniel Day Lewis (Lincoln)	Jennifer Lawrence (Silver Linings Playbook)	Ang Lee (Life of Pi)	Django Unchained
8	84	2012	Billy Crystal	The Artist	Jean Dujardin (The Artist)	Meryl Streep (The Iron Lady)	Michel Hazanavicius (The Artist)	Midnight in Paris
9	83	2011	James Franco, Anne Hathaway	The King's Speech	Colin Firth (for The King's Speech)	Natalie Portman (for Black Swan)	Tom Hooper (for The King's Speech)	The King's Speech
10	82	2010	Steve Martin, Alec Baldwin	The Hurt Locker	Jeff Bridges (for Crazy Heart)	Sandra Bullock (for The Blind Side)	Kathryn Bigelow (for The Hurt Locker)	The Hurt Locker
11	81	2009	Hugh Jackman	Slumdog Millionaire	Sean Penn (for Milk)	Kate Winslet (for The Reader)	Danny Boyle (for Slumdog Millionaire)	Milk
12	80	2008	Jon Stewart	No Country for Old Men	Daniel Day-Lewis (for There Will Be Blood)	Marion Cotillard (for La Vie en Rose (La môme))	Joel Coen and Ethan Coen (for No Country for Old Men)	Juno
13	79	2007	Ellen DeGeneres	The Departed	Forest Whitaker (for The Last King of Scotland)	Helen Mirren (for The Queen)	Martin Scorsese (for The Departed)	Little Miss Sunshine
14	78	2006	Jon Stewart	Crash	Phillip Seymour Hoffman (for Capote)	Reese Witherspoon (for Walk the Line)	Ang Lee (for Brokeback Mountain)	Crash
15	77	2005	Chris Rock	Million Dollar Baby	Jamie Foxx (for Ray)	Hilary Swank (for Million Dollar Baby	Clint Eastwood (for Million Dollar Baby)	Eternal Sunshine of the Spotless Mind
16	76	2004	Billy Crystal	The Lord of the Rings: The Return of the King	Sean Penn (for Mystic River)	Charlize Theron (for Monster)	Peter Jackson (for The Lord of the Rings: The Return of the King)	Lost in Translation
17	75	2003	Steve Martin	Chicago	Adrien Brody (for The Pianist)	Nicole Kidman (for The Hours)	Roman Polanski (for The Pianist)	Talk to Her
18	74	2002	Whoopi Goldberg	A Beautiful Mind	Denzel Washington (for Training Day)	Halle Berry (for Monster's Ball)	Ron Howard (for A Beautiful Mind)	Gosford Park
19	73	2001	Steve Martin	Gladiator	Russel Crowe (for Gladiator)	Julia Roberts (for Erin Brockovich)	Steven Soderbergh (for Traffic)	Almost Famous
20	72	2000	Billy Crystal	American Beauty	Kevin Spacey (for American Beauty)	Hilary Swank (for Boys Don't Cry)	Sam Mendes (for American Beauty)	American Beauty
21	71	1999	Whoopi Goldberg	Shakespeare in Love	Roberto Benigni (for Life Is Beautiful)	Gwyneth Paltrow (for Shakespeare in Love)	Steven Spielberg (for Saving Private Ryan)	Shakespeare in Love
22	70	1998	Billy Crystal	Titanic	Jack Nicholson (for As Good as it Gets)	Helen Hunt (for As Good as it Gets)	James Cameron (for Titanic)	Good Will Hunting
23	69	1997	Billy Crystal	The English Patient	Geoffrey Rush (for Shine)	Frances McDormand (for Fargo)	Anthony Minghella (for The English Patient)	Fargo
24	68	1996	Whoopi Goldberg	Braveheart	Nicolas Cage (for Leaving Las Vegas)	Susan Sarandon (for Dead Man Walking)	Mel Gibson (for Braveheart)	The Usual Suspects
25	67	1995	David Letterman	Forrest Gump	Tom Hanks (for Forrest Gump)	Jessica Lange (for Blue Sky)	Robert Zemeckis (for Forrest Gump)	Pulp Fiction
26	66	1994	Whoopi Goldberg	Schindler's List	Tom Hanks (for Philadelphia)	Holly Hunter (for The Piano)	Steven Spielberg (for Schindler's List)	The Piano
27	65	1993	Billy Crystal	Unforgiven	Al Pacino (for Scent of a Woman)	Emma Thompson (for Howards End)	Clint Eastwood (for Unforgiven)	The Crying Game
28	64	1992	Billy Crystal	The Silence of the Lambs	Anthony Hopkins (for The Silence of the Lambs)	Jodie Foster (for The Silence of the Lambs)	Jonathan Demme (for The Silence of the Lambs)	Thelma & Louise
29	63	1991	Billy Crystal	Dances with Wolves	Jeremy Irons (for Reversal of Fortune)	Kathy Bates (for Misery)	Kevin Costner (Dances with Wolves)	Ghost

B.

70 PROC CONTENTS DATA=WORK.OSCAR; RUN;

PROC CONTENTS DATA=WORK.OSCAR; RUN;



 \mathbf{C}

/* C: The limitations include the missing talk show host represented as a string instead of a missing value.

The actresses and actors that won are associated with a picture, but best_actress should just be the

actresses' name to avoid confusion. Also, I had to rename the data separately.

The Import statment, unlike the engine, also cannot allow multiple users to use the same Excel file.

LIBNAME Oscars XLSX '/home/u62223361/Intro to SAS/HW2/Oscars.xlsx';
PROC PRINT DATA = Oscars.Awards;
TITLE 'PROC PRINT OF OSCAR DATA';

RUN:

/* Limitations of using this method: does not offer many options because it is reading it as a SAS dataset (according to the textbook).



Here are the points for the HW #4 HW Q1(5),3 (5), 4 (5), 11(5), 12(5), 14(5), 16(5), 18(5), 30(10), 34 (10), Q 45 (20), 54(20) with a total of 100

Notes From Teacher: Please pay attention to the following.

- 1) read the question carefully and use the proper variable
- 2) check the output and make sure that SAS code works as you have intended
- 3) include SAS log as part of HW
- 4) have your name on the first line of page