When you have completed this HW, submit via Moodle the following:

1) Your SAS program (.sas) that contains the answers to the fill in the blanks as comments. Be sure to include the question number and letter for each comment.
2) Your SAS log
3) Your Results, generate your results using ods as a .pdf or .rtf file

1. Examine the below code. Given the DROP statement in the dataset.
    a. Will the program calculate Compensation and BonusMonth correctly?
    b. Why or Why not?

```
data work.comp;
  set orion.sales;
  drop Gender Salary Job_Title Country
       Birth_Date Hire_Date;
  Bonus=500;
  Compensation=sum(Salary,Bonus);
  BonusMonth=month(Hire_Date);

run;
```

2. Read in **orion.nonsales**, a non-validated data set ("dirty data").
    a. Submit the below code to create the dataset work.bonus1 and review the results.

```
data work.bonus;
  set orion.nonsales;
  if Country='US' then Bonus=500;
  else Bonus=300;

run;
```

        i. Why is **Bonus** set to 300 in observations 125, 197, and 200?
        ii. Write at least one proc that is NOT a proc print to check the different Bonus amounts assigned to the different Country codes in the dataset.
        iii. How many people receive a bonus of $300?

    b. Utilize the code below for one solution to testing for invalid data.

```
/* Solution 1: Test for multplie values of country */
data work.bonus;
  set orion.nonsales;
  if Country in ('US','us')
                then Bonus=500;
  else Bonus=300;

run;
```

        i. How many people receive a bonus of $500?

    c. Utilize the code below for a second solution to testing for invalid data.

```
/* Solution 2: Call the upcase function in the expression */
```

```
data work.bonus;
  set orion.nonsales;
  if upcase(Country)='US' then Bonus=500;
  else Bonus=300;

run;
```

i. How many people receive a bonus of $500?

d. It is best practice to clean the data at the source but in some cases that is not possible. However you can write a statement before the if then/else statements to ensure that all of your Country values are “AU” or “US”. Add a data step that reads in orion.nonsales to create work.bonus4. Within this one data step

i. use the upcase function to “clean” the codes under Country

ii. write one if/then statement and one else statement to give US employees a bonus of 500 and AU employees a bonus of 300.

iii. Provide a proc step that demonstrates you have implemented your if/then/else statements correctly.

iv.

3. Write a DATA step that reads orion.customer to create work.birthday.

a. In the DATA step, create three new variables: Bday2009, BdayDOW2009, and Age2009 where

i. 'Bday2009' is the combination of the month of 'Birth_Date', the day of 'Birth_Date', and the constant of 2009 in the MDY function.

ii. 'BdayDOW2009' is the day of the week of 'Bday2009'.

iii. 'Age2009' is the age of the customer in 2009. Subtract 'Birth_Date' from 'Bday2009' and then divide by 365.25.

b. Include only the variables 'Customer_Name', 'Birth_Date', 'Bday2009', 'BdayDOW2009', and 'Age2009'.

c. Format 'Bday2009' to resemble a two-digit day, a three-letter month, and a four-digit year. 'Age2009' should be formatted to appear with no digits after the decimal point.

d. Write a PRINT procedure that only prints the customer named 'Raedene'.

i. What observation is she? _____

ii. How old is she? _____

iii. On which day of the week is her birthday (give the actual day, not just a number)? _____

e. Write a PRINT procedure that only prints the customers whose birthday is in June.

i. From the log, how many observations are there? _____

4. In this question, you will use the Orion star SAS datasets 'Employee_Addresses' and 'Employee_Donations'. The first dataset contains addresses of Orion Star employees and the second dataset contains the amounts of charitable donations made by some employees in different quarters of the year. Note that every employee in Employee_Donations has made at least one donation to some charity (i.e. there are no employees who have missing values for all four quarters).

    a. Create a temporary dataset named 'Donations' based on the dataset Employee_Donations'. Do not include the variables 'Recipients' and 'Paid_By'.

    b. Create a temporary dataset named 'Addresses' based on the dataset Employee_Addresses'. Don't include the variables 'Employee_Name' and 'Street_ID'. Only include US employees.

    c. Sort the two temporary datasets in a. and b. by 'Employee_ID'. Do not overwrite the original temporary datasets.

    d. Merge the two new sorted datasets by Employee_ID to create a new temporary dataset called “Donate_Address”.

        i. How many observations are in the new dataset “Donate_Address”? _____

    e. OUTPUTTING TO MULTIPLE DATA SETS: Write another DATA step that merges the two sorted datasets in c. by Employee_ID. In particular do the following:

        i. In this DATA step, write all the observations where the value of 'Employee_ID' is found in (sorted) Addresses but not in (sorted) Donations to a new dataset called 'AddOnly'.

        ii. Write all the observations where the value of 'Employee_ID' is found in (sorted) Donations but not in (sorted) Addresses to a new dataset called 'DonOnly'.

        iii. Do not include the missing variables in the new datasets. For example, the dataset 'AddOnly' will have missing values for all observations for the variables Qtr1, ... Qtr4. Do not include these in the 'AddOnly' dataset. Similarly, DonOnly will have missing values for all the Address-related variables. Don't include these in 'DonOnly'. In order to do this, you will have to use the DROP= and KEEP= options in the DATA statement.

        iv. How many observations and variables are in 'AddOnly'?

            Observations_____ Variables_____

        v. How many observations and variables are in 'DonOnly'?

            Observations_____ Variables_____

        vi. Using your fill-in-the-blank answers in parts c. and d., how many employees have a valid address AND made a charitable donation? (i.e. how many 'matches' are there?) _____

    f. Now write a short DATA step to confirm your answer in e.

How many hours did it take you to complete this homework assignment? _____