SAS Test Questions:

For the following questions you may provide pseudocode, fully functional code or a step by step walk through of your methodology where appropriate

1. **Please describe how you would sort a dataset without duplicate patient IDs using SAS**

```
proc sort data=dataset NODUPKEY;
by id;
run;
```

2. **Write a code to join two datasets together without duplicate entries using a SAS data step and proc SQL**

```
data want;
set have1 have2;
run;
proc SQL;
select distinct *
     from want;
quit;
```

3. **How would you transpose a SAS dataset from columns to multiple rows per patient ID?**
   proc transpose data=dataset;

```
proc transpose data=dataset;
by patientID NOTSORTED;#otherwise you will need to sort it
run;
```

4. **Given a patient ID number and a series of columns containing diagnosis codes write an array statement that checks across all diagnosis fields for a specific diagnosis code**

```
proc SQL;
select distinct *
     from want;
quit;
```

5. **Write a macro combining questions 2 and 4**

```
#Q2
%macro;
%mend;
#Q4
%macro varexist(data,var,info);
%mend;
```

6. **Briefly describe what methods you would use to model the following types of dependent variables using regression analysis and why**
   a. **Cost data -**
   b. **Binary value for whether a patient does or does not have a disease - chi square**
   c. **Time to treatment from diagnosis of a disease - time series**

7. **Using a %DO loop within the appropriate macro statement identify variables from a dataset with variables TX1 – TX5, startdt1 – startdt5, and enddt1 – enddt5. Do three things:**

   a. **Search for the value where the treatment (TX) = "DRUGX"**

b. **identify the start and end dates (startdt, enddt) for "DRUGX"**

c. **calculate months on therapy for "DRUGX" where months is a continuous measure**

8. **Write a short code for formatting and then taking the difference between two date variables. The date should be formatted in MMDDYY format**

```
input(date, MMDDYY10.);
diff = INTCK(DAY, Date_1, Date_2);
```

9. **What SAS procs would you use for Kaplan Meier analysis?**

```
proc LIFETEST
```

10. **Write a code to perform a Kaplan Meier plot for time to death for three different populations on the same graph. Explain the logic for each variable used in your script**

```
ODS HTML;
ODS GRAPHICS On;
proc lifetest data=mydata.Life plots=(s);
   time SURVIVAL * CENSOR (0);
   strata GROUP;#3 groups
   title 'Survival Analysis -- Compare 3 Groups';
run;
ODS HTML CLOSE;
ODS GRAPHICS OFF;
quit;
```