1. In SAS, begin your library by creating a new folder under your (Home) directory. Make sure to define the correct path (i.e. /home/uXXXXXXXX/Name)

**LIBNAME Exercise "/home/u62103066/Exercise";**

1. Use **PROC IMPORT** to import each **Excel**file provided (Comment2.xlsx, FallTable2.xlsx, InterviewTable2.xlsx, MedicalHistory2.xlsx and SurgicalHistory2.xlsx)

**PROC IMPORT DBMS=XLSX**

**OUT=Comment**

**Datafile="/home/u62103066/Exercise/Comment2.xlsx"**

**Replace;**

**RUN;**

**PROC IMPORT DBMS=XLSX**

**OUT=FallTable**

**Datafile="/home/u62103066/Exercise/FallTable2.xlsx"**

**Replace;**

**RUN;**

**PROC IMPORT DBMS=XLSX**

**OUT=InterviewTable**

**Datafile="/home/u62103066/Exercise/InterviewTable2.xlsx"**

**Replace;**

**RUN;**

**PROC IMPORT DBMS=XLSX**

**OUT=MedicalHistory**

**Datafile="/home/u62103066/Exercise/MedicalHistory2.xlsx"**

**Replace;**

**RUN;**

**PROC IMPORT DBMS=XLSX**

**OUT=SurgicalHistory**

**Datafile="/home/u62103066/Exercise/SurgicalHistory2.xlsx"**

**Replace;**

**RUN;**

1. Sort all tables by **SubjectID**

**PROC SORT DATA=Comment; BY SubjectID; RUN;**

**PROC SORT DATA=FallTable; BY SubjectID; RUN;**

**PROC SORT DATA=InterviewTable; BY SubjectID; RUN;**

**PROC SORT DATA=MedicalHistory; BY SubjectID; RUN;**

**PROC SORT DATA=SurgicalHistory; BY SubjectID; RUN;**

1. Merge all of the tables by **SubjectID**

**DATA AllTables;**

**MERGE Comment FallTable InterviewTable MedicalHistory SurgicalHistory;**

**BY SubjectID;**

**RUN;**

1. Use **PROC FREQ** to identify the distribution of response categories for the following variables:

**SDOsteoDiag SDRheumaDiag SDOADiag SDLupusDiag SDHyperthyDiag SDHypothyDiag SDLiverDiag SDBreastDiag SDUterineDiag SDMtMyelomaDiag SDCancerOtherDiag**

 Also use **PROC FREQ**to determine the distribution of individuals who had a diagnosis of BOTH osteoporosis (**SDOsteoDiag**) and osteoarthritis(**SDOADiag**).

**PROC FREQ Data=AllTables; Tables SDOsteoDiag SDRheumaDiag SDOADiag SDLupusDiag SDHyperthyDiag SDHypothyDiag SDLiverDiag SDBreastDiag SDUterineDiag SDMtMyelomaDiag SDCancerOtherDiag**

**; RUN;**

**PROC FREQ Data=AllTables; Where SDOsteoDiag = "Yes" and SDOADiag = "Yes"; RUN;**

1. Use a **DO**loop to recode all "missing data" and "Don't Know" responses to a period. You can repeat**PROC FREQ** to check if this worked

**Data AllTables2;**

**SET AllTables;**

**Array DX(\*) SDOsteoDiag SDRheumaDiag SDOADiag SDLupusDiag SDHyperthyDiag SDHypothyDiag SDLiverDiag SDBreastDiag SDUterineDiag SDMtMyelomaDiag SDCancerOtherDiag;**

**DO i = 1 to DIM(DX);**

**IF DX(i) = "missing data" THEN DX(i) = ".";**

**ELSE IF DX(i) = "Don't Know" THEN DX(i) = ".";**

**ELSE DX(i) = DX(i);**

**END;**

**RUN;**

**PROC FREQ Data=AllTables2; Tables SDOsteoDiag SDRheumaDiag SDOADiag SDLupusDiag SDHyperthyDiag SDHypothyDiag SDLiverDiag SDBreastDiag SDUterineDiag SDMtMyelomaDiag SDCancerOtherDiag**

**; RUN;**

1. Recode "Yes" and "No" responses to 1's and 0's as well

**Data AllTablesRC;**

**SET AllTables2;**

**Array RX(\*) SDOsteoDiag SDRheumaDiag SDOADiag SDLupusDiag SDHyperthyDiag SDHypothyDiag SDLiverDiag SDBreastDiag SDUterineDiag SDMtMyelomaDiag SDCancerOtherDiag;**

**DO i = 1 to DIM(RX);**

**IF RX(i) = "Yes" THEN RX(i) = 1;**

**ELSE IF RX(i) = "No" THEN RX(i) = 0;**

**ELSE RX(i) = RX(i);**

**END;**

**RUN;**

1. Create a new subset of the original merged dataset, including only observations with outcome code 7710 (SIOutcome)

**Data SIOutcomeDS;**

**Set AllTables;**

**Where SIOutcome = 7710;**

**RUN;**

1. For these individuals, use **PROC MEANS** to measure the central tendancy (mean, median) and distribution (SD, Q1, Q3) for the following variables: **SDOsteoAge SDRheumaAge SDOAAge SDHypothyAge SDBreastAge**

**PROC MEANS Data = SIOutcomeDS MEAN MEDIAN STD Q1 Q3;**

**VAR SDOsteoAge SDRheumaAge SDOAAge SDHypothyAge SDBreastAge;**

**RUN;**

**PROC Means Data = SIOutcomeDS MEAN MEDIAN STD Q1 Q3;**

**VAR SDOsteoAge SDRheumaAge SDOAAge SDHypothyAge SDBreastAge;**

**Where SDHysterectomy = "Yes";**

**RUN;**

**PROC Means Data = SIOutcomeDS MEAN MEDIAN STD Q1 Q3;**

**VAR SDOsteoAge SDRheumaAge SDOAAge SDHypothyAge SDBreastAge;**

**Where SDHysterectomy = "No";**

**RUN;**