```
FILENAME REFFILE '/home/u64189948/Self_Efficacy/AEMS-2019.csv';
PROC IMPORT DATAFILE=REFFILE
DBMS=CSV
OUT=WORK.IMPORT2;
GETNAMES=YES;
RUN;

PROC CONTENTS DATA=WORK.IMPORT2; RUN;

/* check data strecutre */
PROC CONTENTS DATA=WORK.IMPORT2;
RUN;

PROC MEANS DATA=WORK.IMPORT2 N NMISS MEAN STD MIN MAX;
VAR Age;
RUN;

PROC FREQ DATA=WORK.IMPORT2;
TABLES Source Gender / MISSING;
RUN;

/* summary statstics note 99, 9, and 88 values are NA*/
PROC MEANS DATA=WORK.IMPORT2 N MEAN STD MIN MAX;
CLASS Source Gender;
VAR Age;
RUN;

/*explore NA */
DATA WORK.IMPORT2_CLEAN;
    SET WORK.IMPORT2;

    ARRAY num_vars {*} _NUMERIC_; /* Select all numeric variables */
    DO i = 1 TO DIM(num_vars);
        IF num_vars[i] IN (9, 88, 99) THEN num_vars[i] = .;
    END;

    DROP i;
RUN;

/* selct high school student data set*/

DATA WORK.IMPORT2_SOURCE1;
    SET WORK.IMPORT2_CLEAN; /* Use the cleaned dataset */
    WHERE Source = 1;
RUN;

/*elimnate non relvant variables*/
DATA WORK.IMPORT2_CLEANED;
    SET WORK.IMPORT2_SOURCE1 (DROP=
        "Source.Name"n
        "Respondent ID"N
        "School Type"N
        DegreeYear
        Education
        RelationshipStatus
        Children
        WorkExperience
        TSE_TeachingSkills
        TSE_Absenteeism
        TSE_AdverseCommunity
        TSE_DifficultStudents
        TSE_DoWell
        TSE_DoWork
        Grade
        TSE_Dropout
        TSE_LackofSupport
        TSE_LocalInvolve
        TSE_LowInterest
        TSE_OnTask
        TSE_Safe
        TSE_StudentsEnjoy
```

```
        TSE_StudentsMemory
        TSE_StudentsTrust
        TSE_TeachingSkills
        TSE_WorkTogether
        Language
        LS_ChangeNothing
        LS_ImportantThings
        LS_LifeExcellent
        LS_LifeIdeal
        LS_LifeSatisfied
        Religion);
RUN;

PROC CONTENTS DATA=WORK.IMPORT2_CLEANED; RUN;

/* elimnate all missing observations/
/* Remove all observations with any missing values */
DATA WORK.IMPORT2_CLEANED_NO_MISSING;
    SET WORK.IMPORT2_CLEANED;
    IF CMISS(OF _ALL_) = 0;
RUN;


/* Check if missing values exist */
PROC MEANS DATA=WORK.IMPORT2_CLEANED_NO_MISSING N NMISS;
RUN;

PROC FREQ DATA=WORK.IMPORT2_CLEANED_NO_MISSING;
    TABLES _ALL_ / MISSING;
RUN;

/*SAS Code for Reverse Coding (4-Point Scale)*/

DATA WORK.IMPORT2_CLEANED_REVERSED;
    SET WORK.IMPORT2_CLEANED_NO_MISSING;

    /* Define an array with the variables to reverse code */
    ARRAY reverse_vars(*) CIO_SelfDepend
                          CIO_SelfDependMost
                          CIO_Identity
                          CIO_JobBetter
                          CIO_Competition
                          CIO_BetterTense
                          Gratitude_NotMuch
                          Gratitude_Time
                          Empathy_Waste
                          Empathy_Difficult
                          SR_GoalSettingHard
                          SR_GoalPlanTrouble
                          SR_Distracted
                          SR_Trouble
                          SR_Decisions
                          SR_Change
                          SR_Problems
                          SR_Focus
                          SR_Mistakes
                          SB_DifferentSchool
                          SB_NotInterested
                          SB_FeelDifferent
                          SB_DontBelong
                          SB_AcceptanceHard
                          MM_NoPurpose
                          SE_Insecure
                          SE_DontHandle
                          SE_NotCapable
                          SE_RarelyAchieve
                          SE_NewGiveUp
                          SE_AvoidDifficult
                          SE_NotTryComplicated;

    /* Reverse code: 4-point Likert scale */
```

```
    DO i = 1 TO DIM(reverse_vars);
        reverse_vars[i] = 5 - reverse_vars[i]; /* Reverse transformation */
    END;

    DROP i;
RUN;

PROC CONTENTS DATA=WORK.IMPORT2_CLEANED_REVERSED; RUN;

/* group items into single construec then calcuate mean*/

DATA WORK.IMPORT2_CONSTRUCT_MEANS;
    SET WORK.IMPORT2_CLEANED_NO_MISSING;

    /* Calculate mean for each construct */
    mean_CIO = MEAN(OF CIO_BetterTense, CIO_Competition, CIO_Cooperate, CIO_Family,
                    CIO_Identity, CIO_JobBetter, CIO_OwnThing, CIO_ParentsChildren,
                    CIO_PeerPrize, CIO_PeerWellbeing, CIO_PleasureTime,
                    CIO_RespectGrpDecisions, CIO_SelfDepend, CIO_SelfDependMost);

    mean_ER = MEAN(OF ER_EmotionControl, ER_ExpressPositive, ER_LessNegative, ER_NegativeExpress,
                   ER_NegativeThink, ER_PositiveChange, ER_PositiveThink_Num, ER_StressCalm);

    mean_Empathy = MEAN(OF Empathy_Criticizing, Empathy_Difficult, Empathy_Perspective,
                        Empathy_Sides, Empathy_TwoSides, Empathy_Upset, Empathy_Waste);

    mean_FG = MEAN(OF Forgive_BrokenEngagement, Forgive_Car_Num, Forgive_CousinArgument,
                   Forgive_CurseDiffReligion, Forgive_CurseSameReligion, Forgive_Loss_Num,
                   Forgive_Rumor, Forgive_SecretDisclosure, Forgive_Wall);

    mean_Grat = MEAN(OF Gratitude_Appreciate, Gratitude_LongList, Gratitude_NotMuch,
                     Gratitude_Thankful, Gratitude_Time, Gratitude_Variety);



    mean_MM = MEAN(OF MM_FeelSignificant, MM_LifeMeaning, MM_LifeMeaningful,
                   MM_LifeMeaningfulSense, MM_LifesPurpose, MM_Mission, MM_NoPurpose,
                   MM_Purpose, MM_SatisfyingPurpose, MM_UnderstandLife);

    mean_PS = MEAN(OF PS_CompareIdeas, PS_ExpressThoughts, PS_GatherInfo, PS_GiveReasons,
                   PS_IdentifyOptions, PS_InfoToSupport, PS_ListenIdeas, PS_MindOpen,
                   PS_MoreThanOne, PS_PlanInfo, PS_ResultsThink, PS_SupportDecisions);

    mean_RS = MEAN(OF RS_DefiningYou, RS_FeelGod, RS_Prayer, RS_Religion, RS_ReligionGrp);

    mean_SB = MEAN(OF SB_AcceptanceHard, SB_Activities, SB_BeMyself_Num, SB_CanTalk,
                   SB_DifferentSchool, SB_DontBelong, SB_FeelDifferent, SB_GoodWork,
                   SB_LikeMe, SB_NotInterested, SB_NoticeGood, SB_OpinionsSeriously,
                   SB_PartOfCommunity, SB_PeopleFriendly, SB_ProudSchool,
                   SB_TeachersInterested, SB_TeachersRespect, SB_TreatedRespect);

    mean_SE = MEAN(OF SE_AvoidDifficult, SE_DontHandle, SE_Insecure, SE_KeepTrying,
                   SE_NewGiveUp, SE_NotCapable, SE_NotTryComplicated, SE_PlansCertain,
                   SE_RarelyAchieve, SE_RightToWork, SE_TryHarder, SE_Unpleasant);

    mean_SR = MEAN(OF SR_Change, SR_Decisions, SR_Distracted, SR_Focus, SR_GoalPlan,
                   SR_GoalPlanTrouble, SR_GoalProgress, SR_GoalSettingHard, SR_Goals,
                   SR_MistakeOnce, SR_Mistakes, SR_MistakesLearn, SR_Problems,
                   SR_Resolution, SR_Trouble, SR_Willpower);

RUN;

PROC CONTENTS DATA=WORK.IMPORT2_CONSTRUCT_MEANS; RUN;

/* remove age >24*/
DATA WORK.IMPORT2_CONSTRUCT_MEANS_FILTERED;
    SET WORK.IMPORT2_CONSTRUCT_MEANS;

    /* Keep only participants with Age ≤ 24 */
    IF Age <= 24;
RUN;
```

```
/* construct only*/
DATA WORK.CONSTRUCT_MEAN_ONLY;
    SET WORK.IMPORT2_CONSTRUCT_MEANS(KEEP=
        mean_CIO mean_ER mean_Empathy mean_FG mean_Grat
        mean_Hope mean_MM mean_PS mean_RS mean_SB mean_SE mean_SR Age);
RUN;

PROC CONTENTS DATA=WORK.CONSTRUCT_MEAN_ONLY; RUN;


/*Visualization*/
/* Generate histograms for all variables */
%MACRO HISTOGRAMS;
    %LET VARS = mean_CIO mean_ER mean_Empathy mean_FG mean_Grat
                mean_PS mean_RS mean_SB mean_SE mean_SR Age;

    %LET COUNT = %SYSFUNC(COUNTW(&VARS));

    %DO I = 1 %TO &COUNT;
        %LET VAR = %SCAN(&VARS, &I);

        PROC SGPLOT DATA=WORK.CONSTRUCT_MEAN_ONLY;
            TITLE "Histogram of &VAR";
            HISTOGRAM &VAR / BINWIDTH=0.5;
            DENSITY &VAR / TYPE=NORMAL;
        RUN;
    %END;
%MEND HISTOGRAMS;

%HISTOGRAMS;

PROC FREQ DATA=WORK.CONSTRUCT_MEAN_ONLY;
    TABLES Age / NOCUM;
RUN;

/* keep only observation with age =1 - student less than 18 year old */
DATA WORK.CONSTRUCT_MEAN_ONLY_AGE1;
    SET WORK.CONSTRUCT_MEAN_ONLY;

    /* Keep only observations where Age = 1 */
    IF Age = 1;
RUN;

PROC CONTENTS Data =WORK.CONSTRUCT_MEAN_ONLY_AGE1; RUN;

/* final data setl* take of age variable */
DATA WORK.CONSTRUCT_MEAN_ONLY_AGE1_FINAL;
    SET WORK.CONSTRUCT_MEAN_ONLY_AGE1 (DROP=Age);
RUN;

PROC CONTENTS Data =WORK.CONSTRUCT_MEAN_ONLY_AGE1_FINAL; RUN;


PROC REG DATA=WORK.CONSTRUCT_MEAN_ONLY_AGE1_FINAL;
    MODEL mean_SE = mean_CIO mean_ER mean_Empathy mean_FG mean_Grat
                    mean_MM mean_PS mean_RS mean_SB mean_SR;
    TITLE "Linear Regression Analysis: Predicting mean_SE";
RUN;
QUIT;


 /* Regresstion*/
PROC REG DATA=WORK.CONSTRUCT_MEAN_ONLY_AGE1_FINAL;
    MODEL mean_SE = mean_CIO mean_ER mean_Empathy mean_FG mean_Grat
                    mean_MM mean_PS mean_RS mean_SB mean_SR;
    OUTPUT OUT=RESIDUALS PREDICTED=P_mean_SE RESIDUAL=Residual;
    TITLE "Linear Regression Analysis: Predicting mean_SE with Residuals";
RUN;
QUIT;
```

```
PROC UNIVARIATE DATA=RESIDUALS NORMAL;
    VAR Residual;
    QQPLOT Residual / NORMAL(MU=EST SIGMA=EST);
    HISTOGRAM Residual / NORMAL;
    TITLE "QQ Plot and Histogram of Residuals";
RUN;
```

```
PROC SGPLOT DATA=RESIDUALS;
    SCATTER X=P_mean_SE Y=Residual;
    REFLINE 0 / AXIS=Y;
    TITLE "Residuals vs. Predicted Values";
RUN;
```