/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*

\* Project : BIOS 645 Course

\*

\* Program name : Homework 1

\*

\* Author : Esther Mun

\*

\* Date created : 2021-02-01

\*

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

ods rtf file = '/home/u41545557/Bios511/Output/Homework1.rtf';

/\*

Imports the txt file into a work dataset

\*/

proc import datafile = '/home/u41545557/Bios511/Data/Chinese\_health\_&\_family\_life\_study.txt'

out = health\_life

dbms = dlm

replace;

getnames=yes;

delimiter = '09'x;

run;

/\* Printed the dataset to make sure the dataset imported correctly. \*/

/\* proc print data = health\_life; \*/

/\* run; \*/

/\* Univariate descriptives for each variable. Proc univariate will be used for continuous variables. \*/

\* Descriptive tables and a plot for R\_age continuous variable. ;

title1'Simple Descriptive Statistics for the Age of the Responding Women';

PROC MEANS DATA=health\_life N MIN MEDIAN MAX MEAN STD SKEW KURT MAXDEC=3;

\* MAXDEC= limits decimal places printed;

VAR R\_age;

RUN;

title1'Histogram and Density Curve for the Distribution of the Age of the Responding Women ';

PROC SGPLOT DATA=health\_life;

HISTOGRAM R\_age;

DENSITY R\_age /TYPE=NORMAL; \* draws a normal distribution over histogram;

RUN;

\* Descriptive tables and a plot for R\_income continuous variable. ;

title1'Simple Descriptive Statistics for the Monthly Income of the Responding Women';

PROC MEANS DATA=health\_life N MIN MEDIAN MAX MEAN STD SKEW KURT MAXDEC=3;

\* MAXDEC= limits decimal places printed;

VAR R\_income;

RUN;

title1'Histogram and Density Curve for the Distribution of the Monthly Income of the Responding Women';

PROC SGPLOT DATA=health\_life;

HISTOGRAM R\_income;

DENSITY R\_income /TYPE=NORMAL; \* draws a normal distribution over histogram;

RUN;

\* Descriptive tables and a plot for R\_height continuous variable. ;

title1'Simple Descriptive Statistics for the Height of the Responding Women';

PROC MEANS DATA=health\_life N MIN MEDIAN MAX MEAN STD SKEW KURT MAXDEC=3;

\* MAXDEC= limits decimal places printed;

VAR R\_height;

RUN;

title1'Histogram and Density Curve for the Distribution of the Height of the Responding Women';

PROC SGPLOT DATA=health\_life;

HISTOGRAM R\_height;

DENSITY R\_height /TYPE=NORMAL; \* draws a normal distribution over histogram;

RUN;

\* Descriptive tables and a plot for A\_height continuous variable. ;

title1"Simple Descriptive Statistics for the Height of the Responding Women's Partner";

PROC MEANS DATA=health\_life N MIN MEDIAN MAX MEAN STD SKEW KURT MAXDEC=3;

\* MAXDEC= limits decimal places printed;

VAR A\_height;

RUN;

title1"Histogram and Density Curve for the Distribution of the Height of the Responding Women's Partner";

PROC SGPLOT DATA=health\_life;

HISTOGRAM A\_height;

DENSITY A\_height /TYPE=NORMAL; \* draws a normal distribution over histogram;

RUN;

\* Descriptive tables and a plot for A\_income continuous variable. ;

title1"Simple Descriptive Statistics for the Monthly Income of the Responding Women's Partner";

PROC MEANS DATA=health\_life N MIN MEDIAN MAX MEAN STD SKEW KURT MAXDEC=3;

\* MAXDEC= limits decimal places printed;

VAR A\_income;

RUN;

title1"Histogram and Density Curve for the Distribution of the Monthly Income of the Responding Women's Partner";

PROC SGPLOT DATA=health\_life;

HISTOGRAM A\_income;

DENSITY A\_income /TYPE=NORMAL; \* draws a normal distribution over histogram;

RUN;

/\* Univariate descriptives for each variable. Proc freq will be used for categorical variables. \*/

title1'Frequency Table for the Region Where the Responding Women Live';

PROC FREQ DATA=health\_life;

TABLES R\_region;

RUN;

title1'Frequency Table for the Education Level of the Responding Women';

PROC FREQ DATA=health\_life;

TABLES R\_edu;

RUN;

title1'Frequency Table for the Self-reported Health Status of the Responding Women';

PROC FREQ DATA=health\_life;

TABLES R\_health;

RUN;

title1'Frequency Table for the Self-reported Happiness of the Responding Women';

PROC FREQ DATA=health\_life;

TABLES R\_happy;

RUN;

title1"Frequency Table for the Level of Education of the Women's partner ";

PROC FREQ DATA=health\_life;

TABLES A\_edu;

RUN;

/\* Bivariate Descriptives for 3 variable pairs \*/

\* Bivariate Descriptive for a pair of continuous variables;

title1'Correlation Matrix of the Relationship Between Age and Monthly Income of the Responding Women';

PROC CORR DATA=health\_life NOSIMPLE;

VAR R\_age R\_income;

RUN;

title1'Scatterplot and Regression Line of the Relationship Between Age and Monthly Income of the Responding Women';

PROC SGPLOT DATA=health\_life ;

SCATTER X=R\_age Y=R\_income;

REG X=R\_age Y=R\_income; \* this draws a regression line over the scatterplot;

RUN;

/\* Bivariate Descriptive for a pair of categorical variables

There shouldn't be a plot since they are categorical variables

\*/

title1"Frequency Table Showing the Relationship Between the Women's Education Level and their Partners ";

PROC FREQ DATA=health\_life;

TABLES R\_edu \* A\_edu /NOCOL NOPCT;

/\* WEIGHT freq; \*/

RUN;

\* Bivariate Descriptive for a pair of a continuous and categorical variable;

title1'Descriptive Statistics for the Income of the Women Grouped by their Education Level';

PROC MEANS DATA=health\_life N MIN MEDIAN MAX MEAN STD SKEW KURT MAXDEC=3;

VAR R\_income; \* means etc. for mortality;

CLASS R\_edu; \* as a function of ocean yes or no;

RUN;

title1"Vertical Boxplots for the Relationship Between the Women's Income and Education Level";

PROC SGPLOT DATA=health\_life;

VBOX R\_income /CATEGORY=R\_edu;

RUN;

ods rtf close;