**Part A. Q1**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | MSEL score | folic acid intake(𝜇g) | Child’s birthyear | Mom’s age at birth | Child’s genotype (0,1) | | Mom’s Pre-Preg BMI (1,2,3,4) | | | | Child’s sex(0,1) | | Home max educ(0,1) | |
| N | 867 | 967 | 966 | 966 |  | |  | | | |  | |  | |
| Mean | 74.434 | 720.865 | 2003.19 | 30.364 |  | |  | | | |  | |  | |
| Std Dev | 28.589 | 422.766 | 1.950 | 5.767 |  | |  | | | |  | |  | |
| Min | -50 | -3 | 1998 | 16 |  | |  | | | |  | |  | |
| Max | 147.5 | 2196.69 | 2008 | 49 |  | |  | | | |  | |  | |
| Freq |  |  |  |  | 621 | 107 | 35 | 488 | 234 | 194 | 191 | 775 | 441 | 525 |
| Percent % |  |  |  |  | 85.3 | 17.4 | 3.68 | 51.31 | 24.61 | 20.40 | 19.77 | 80.23 | 45.65 | 54.35 |
| T-score | N/A | 2.97 | 3.81 | 1.61 | 0.71 | | -2.64 | | | | -1.34 | | 5.06 | |
| Pearson’s R | N/A | 0.100 | 0.128 | 0.055 | 0.026 | | -0.090 | | | | -0.046 | | 0.170 | |
| P-value | N/A | 0.0031 | 0.0002 | 0.1067 | 0.4793 | | 0.0085 | | | | 0.1792 | | <.0001 | |
| Sig. (α = 0.05) \*yes /no | N/A | \* | \* | / | / | | \* | | | | / | | \* | |

Table 1: Summary Statistics, Pearson’s correlation coefficients, t-statistics, and p-values

**Methods**

To explore the distribution of the data, descriptive statistics are calculated and analyzed. Including mean and standard deviation for numeric variables, and frequency counts and percentages for categorical variables. To explore the pairwise association between MSEL early learning composite score and each of the other variables of interest, several methods are used. Including scatter plots for MSEL score against each of the numeric variables, boxplots for MSEL score by categorical variables, Pearson’s correlation coefficient analysis, and simple univariate linear regression analysis. To determine the statistical significance at α = 0.05 for the pairwise associations, equivalent p-values are calculated from the Pearson’s correlation coefficient or the t-score from linear regression.

**Results**

Summary statistics shows that the age range for mother’s age at child’s birth is within a medically viable range (19-49). The mean score for MSEL is 74.434, with the maximum possible score being 162 and a sample maximum of 147.5. There are more male subjects (80.23%) than female subjects (19.77%) in the sample. Initial visual inspection from the three scatter plots and the four scatter plots shows positive linear trends in the associations of MSEL with folic acid intake, MSEL with child’s birthyear, and MSEL with highest education at home. A negative linear trend is found between MSEL and mother’s pre-pregnancy BMI. No apparent trends are seen between the three associations MSEL and mother’s age at child’s birth, MSEL and child’s genotype, and MSEL and child’s sex. Further analyses using Pearson’s correlation analysis and simple linear univariate regression confirm the above visual impressions, showing statistically significant p-values at α = 0.05 for variables folic acid intake (p=0.0031), child’s birthyear (p=0.0002), mother’s pre-pregnancy BMI (p=0.0085), and maximum education at home (p<.0001).

**Code**

libname library "Y:\";

**proc** **format**;

value mthfr677f

**0** = "CC or CT genotype"

**1** = "TT genotype";

value prepregbmi\_catf

**1** = "< 18.5"

**2** = "18.5 to <25.0"

**3** = "25.0 to <30.0"

**4** = "30.0+ ";

value child\_malef

**0** = "Female"

**1** = "Male";

value maxedu\_bsf

**0** = "Some college or less"

**1** = "Bachelor's degree or higher";

**run**;

**data** charge;

set library.charge\_278;

format mthfr677 mthfr677f.;

format prepregbmi\_cat prepregbmi\_catf.;

format child\_male child\_malef.;

format maxedu\_bs maxedu\_bsf.;

label subject\_id = "Subject ID";

label msel\_elcs = "MSEL early learning composite score";

label folicacid\_t1 = "Maternal folic acid (𝜇g) intake during the first trimester";

label mthfr677 = "Child's MTHFR677 Genotype";

label prepregbmi\_cat = "Mother's pre-pregnancy BMI";

label yob = "Child's year of birth";

label child\_male = "Child's sex";

label agemomyrs = "Mother's age at birth (years)";

label maxedu\_bs = "Maximum education in the home";

**run**;

**proc** **means** data = charge;

var msel\_elcs folicacid\_t1 yob agemomyrs;

**run**;

**proc** **freq** data = charge;

tables mthfr677 prepregbmi\_cat child\_male maxedu\_bs;

**run**;

**%macro** scatter(y, x);

proc sgplot data = charge;

scatter y = &y x = &x;

reg y = &y x = &x;

**%mend** scatter;

%***scatter*** (msel\_elcs, folicacid\_t1);

%***scatter*** (msel\_elcs, yob);

%***scatter*** (msel\_elcs, agemomyrs);

**%macro** vbox(x);

proc sgplot data = charge;

vbox msel\_elcs / category = &x;

**%mend** vbox;

%***vbox***(mthfr677);

%***vbox***(prepregbmi\_cat);

%***vbox***(child\_male);

%***vbox***(maxedu\_bs);

**proc** **corr** data = charge;

var msel\_elcs folicacid\_t1 mthfr677 prepregbmi\_cat yob child\_male agemomyrs maxedu\_bs;

**run**;

**%macro** reg(x);

proc reg data = charge;

model msel\_elcs = &x;

title "Regress MSEL Score on &x";

**%mend** reg;

%***reg***(folicacid\_t1);

%***reg***(mthfr677);

%***reg***(prepregbmi\_cat);

%***reg***(yob);

%***reg***(child\_male);

%***reg***(agemomyrs);

%***reg***(maxedu\_bs);

**Output**

Graphical user interface, application, table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart

Description automatically generatedChart, scatter chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Table

Description automatically generated with low confidence

Table

Description automatically generated

Graphical user interface, application, table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated with low confidence

Table

Description automatically generated