How many times do you go to see a doctor in a year?	How many times do you go to see a dentist in a year?	How many types of medication do you regularly take including vitamin supplements.	Please provide your age.
1	1	1	22
1	2	3	25
0	1	0	32
2	1	3	30
0	1	1	28
0	1	0	21
0	0	0	23
1	1	1	23
1	1	3	30
2	1	5	31
1	1	5	45
2	1	2	36
6	1	7	35
4	1	3	24
3	1	6	52
5	1	8	48
0	0	0	33
0	1	0	22
5	0	4	21
1	1	2	26
1	1	0	25
1	1	1	25
2	2	2	29
1	0	0	31
0	0	0	20
0	0	5	35
3	1	3	35

2	1	4	34
1	1	1	30
1	1	7	46
4	1	5	44
1	1	1	24

```
DATA HEALTH;
INPUT
Doctor_Visits  1-2
Dentist_Visits  3-4
Medication  5-6
Age   7-9
;
DATALINES;
1 1 1 22
1 2 3 25
0 1 0 32
2 1 3 30
0 1 1 28
0 1 0 21
0 0 0 23
1 1 1 23
1 1 3 30
2 1 5 31
1 1 5 45
2 1 2 36
6 1 7 35
4 1 3 24
3 1 6 52
5 1 8 48
0 0 0 33
0 1 0 22
5 0 4 21
1 1 2 26
1 1 0 25
1 1 1 25
2 2 2 29
1 0 0 31
0 0 0 20
0 0 5 35
3 1 3 35
2 1 4 34
1 1 1 30
1 1 7 46
4 1 5 44
1 1 1 24

RUN;
```

```
43 PROC MEANS DATA =HEALTH N MEAN VAR STD MIN Q1 MEDIAN Q3 MAX QRANGE MAXDEC=2;
44 RUN;
45
46 PROC GPLOT DATA = HEALTH;
47 PLOT Age*Doctor_Visits;
48 PLOT Age*Dentist_Visits;
49 PLOT Age*Medication;
50 RUN;
51
52 PROC UNIVARIATE DATA=HEALTH NORMAL PLOT;
53
54 RUN;
```

The MEANS Procedure

Variable	N	Mean	Variance	Std Dev	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Quartile Range
Doctor_Visits	32	1.63	2.69	1.64	0.00	0.50	1.00	2.00	6.00	1.50
Dentist_Visits	32	0.88	0.24	0.49	0.00	1.00	1.00	1.00	2.00	0.00
Medication	32	2.59	5.73	2.39	0.00	0.50	2.00	4.50	8.00	4.00
Age	32	30.78	73.02	8.54	20.00	24.00	30.00	35.00	52.00	11.00

Data Analysis

1. Correlation Coefficient value is 0.13986.

   They have a weak correlation because they have about 0.13986 correlation value which is less than 0.5 (|r| <0.5 then weak correlation).

```
44 PROC CORR DATA=HEALTH;
45     VAR Doctor_Visits Dentist_Visits;
46 RUN;
```

The CORR Procedure

2 Variables:	Doctor_Visits Dentist_Visits

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Doctor_Visits	32	1.62500	1.64120	52.00000	0	6.00000
Dentist_Visits	32	0.87500	0.49187	28.00000	0	2.00000

Pearson Correlation Coefficients, N = 32 Prob > |r| under H0: Rho=0		
	Doctor_Visits	Dentist_Visits
Doctor_Visits	1.00000	0.13986 0.4452
Dentist_Visits	0.13986 0.4452	1.00000

2. ANOVA results from the PROC REG and what they mean

(a) p-value is 0.4452 which is a greater number
(b) p-value is not small means that the model is not statistically significant.
It means that the relationship between two parameters is not significant. They are not much correlated.

3. Regression results and what they mean (p-values on parameters,significance, residuals)

```
48 PROC REG DATA=HEALTH;
49 MODEL Doctor_Visits = Dentist_Visits;
50 RUN;
```

The CORR Procedure

2 Variables:	Doctor_Visits Dentist_Visits

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Doctor_Visits	32	1.62500	1.64120	52.00000	0	6.00000
Dentist_Visits	32	0.87500	0.49187	28.00000	0	2.00000

Pearson Correlation Coefficients, N = 32 Prob > |r| under H0: Rho=0		
	Doctor_Visits	Dentist_Visits
Doctor_Visits	1.00000	0.13986 0.4452
Dentist_Visits	0.13986 0.4452	1.00000

The REG Procedure
Model: MODEL1
Dependent Variable: Doctor_Visits

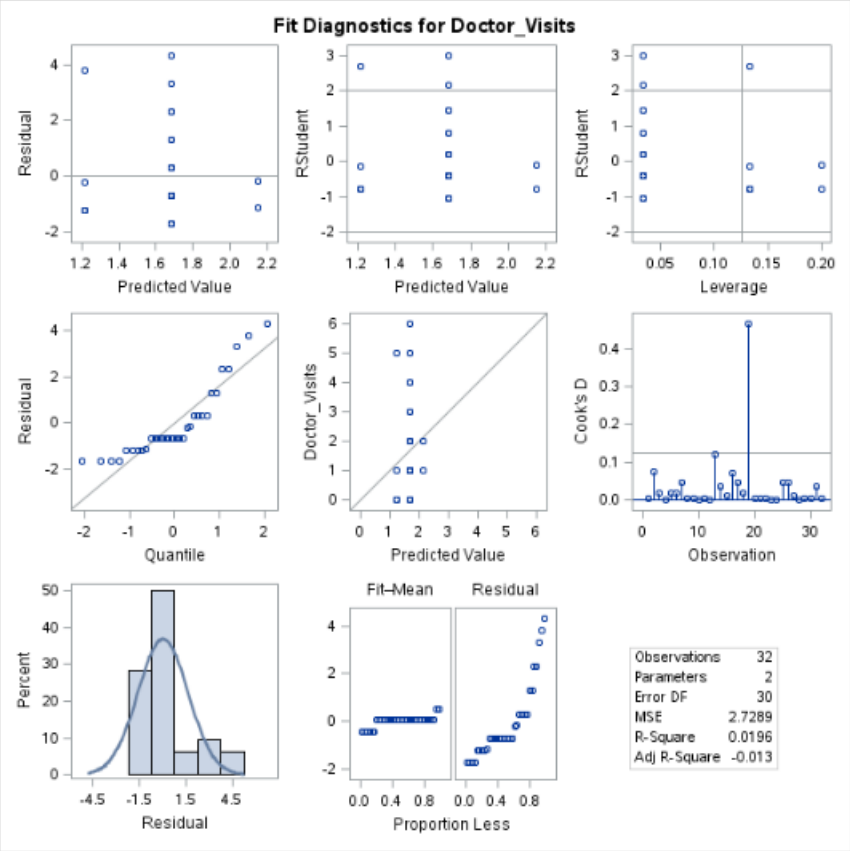
Number of Observations Read	33
Number of Observations Used	32
Number of Observations with Missing Values	1

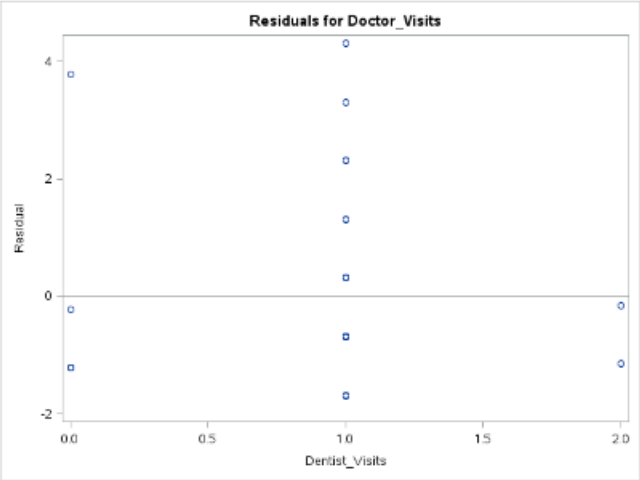
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.63333	1.63333	0.60	0.4452
Error	30	81.86667	2.72889		
Corrected Total	31	83.50000			

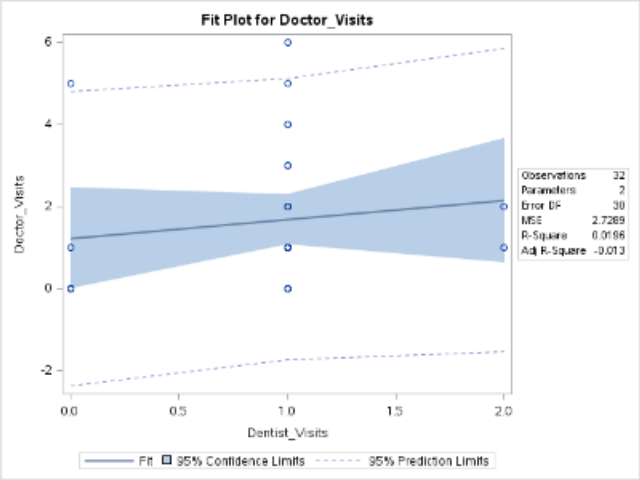
Root MSE	1.65193	R-Square	0.0196
Dependent Mean	1.62500	Adj R-Sq	-0.0131
Coeff Var	101.65753		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.21667	0.60320	2.02	0.0527
Dentist_Visits	1	0.46667	0.60320	0.77	0.4452

The REG Procedure
Model: MODEL1
Dependent Variable: Doctor_Visits







The regression equation

slope : 0.46667

intercept = 1.21667

y= 0.46667 x + 1.21667

The p-values for the parameters and explain what the p-values mean

$\beta_0 =$ 0.0527, changes in the predictor is not much related to changes in the response since little bit greater than alpha 0.05. It is not much meaningful to the model.

$\beta_1 =$ 0.4452 changes in the predictor is not related to changes in the response since the parameter's p-value is greater than alpha 0.05. It is not meaningful to the model.

The residual analysis, look at the normal probability plot and analyze and the residuals by regression for the dependent variable

The distribution of the residuals looks neither particularly symmetric, nor is perfectly close to normal distribution. It looks linear but there is a chance of existing outliers.

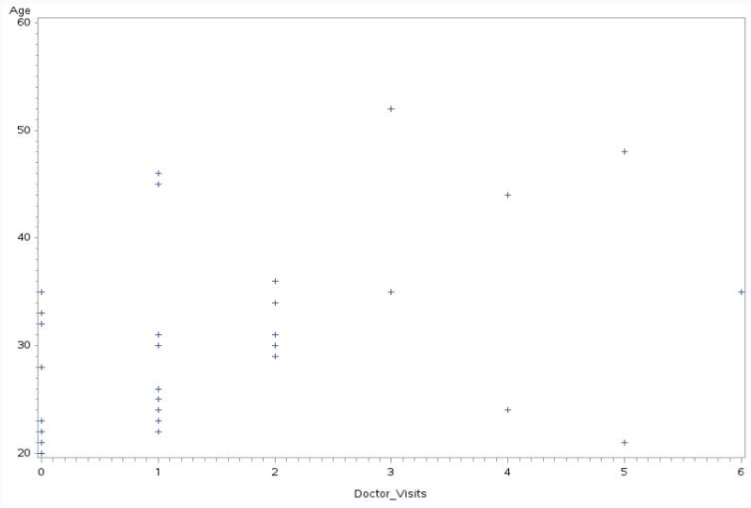
It shows that there are non-zero values for the residuals based on the fitted value. For example, a fitted value of 1.2 has 4 residual and 1.7 has 4,3,2, and etc. It does not seem to have a residual mean of 0.

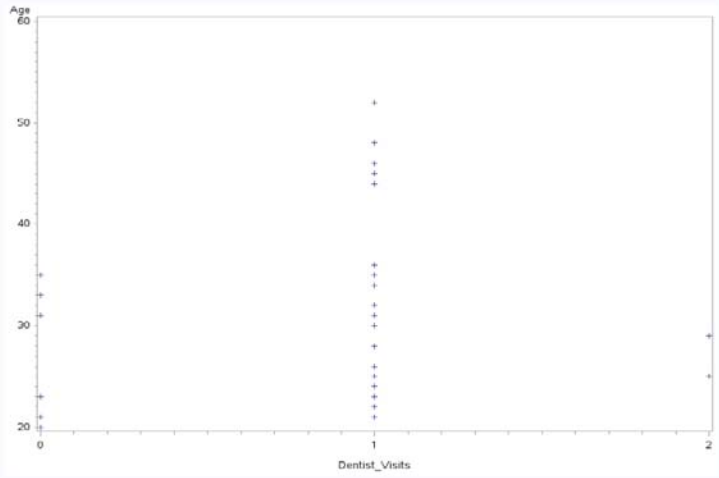
ii. Make sure all assumptions are met. Explain what this does to your results.

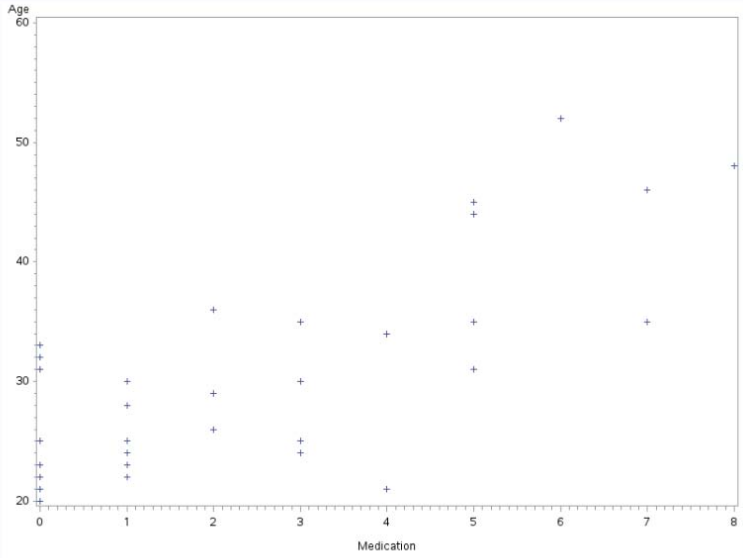
Assumption of linearity was met by making linear regression equation and line. Assumption of Independence is met by an assumption of a valid survey data. Assumption of normality is met by drawing a histogram of the residuals. There are few more assumptions and the all assumptions are met by the above test. A statistical test starts from an assumption and the result is not true if there exists a wrong assumption. Thus, the result is true because all the assumptions are met.

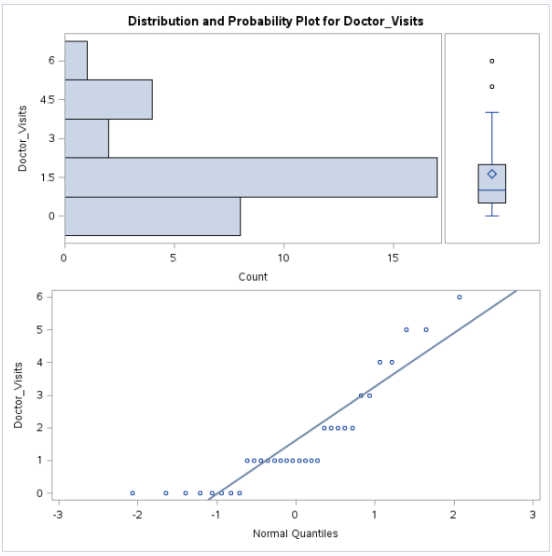
iii. Explain why your sample is good or not good.

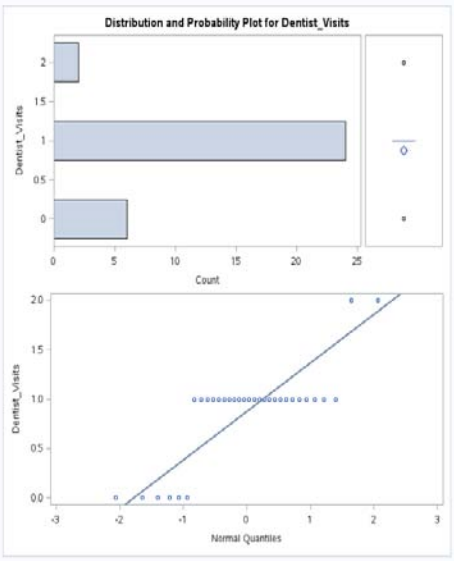
According to the test above, my sample is not good because my survey does not provide an important correlation or information. The test result came out that there is no significant correlation between the two variables. However, there is still a chance that it probably has a significant information or analysis from different types of tests in different variab

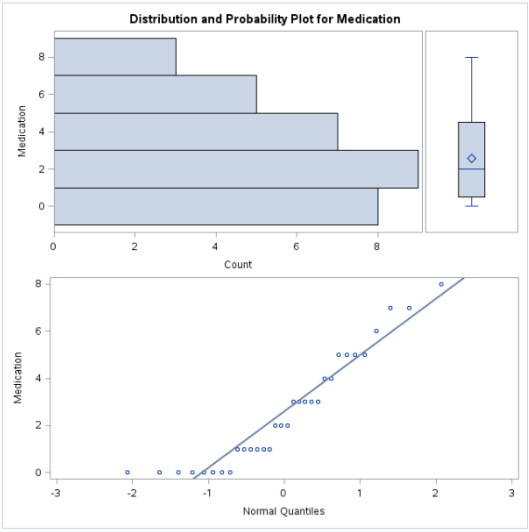

Age
60
50
40
30
20
0
1
2
3
4
5
6
Doctor_Visits


Age
60
50
40
30
20
0
1
2
Dentist_Visits


Age
60
50
40
30
20
0
1
2
3
4
5
6
7
8
Medication


Distribution and Probability Plot for Doctor_Visits
Doctor_Visits
Count
Doctor_Visits
Normal Quantiles


Distribution and Probability Plot for Dentist_Visits
Dentist_Visits
Count
Dentist_Visits
Normal Quantiles


Distribution and Probability Plot for Medication
Medication
Count
Medication
Normal Quantiles