. Q2(5),4(5),7(5), 11(5),12(5),17(10),23(10),25(a-f 30),28(a-h 40)=115

Q20 -3, Q25 -3,-2,-2,5,-2 =-17

Intro to SAS HW Chapter 9

85.2(98/115)

Ian Liu

HW Chapter 9 Q2

2. Which procedure will display the median by default?

a. TTEST
b. MEANS
c. UNIVARIATE
d. All of the above

C. UNIVARIATE

HW Chapter 9 Q4

4. Which procedure can compute a one-sample *t* test against any desired hypothesized value?

a. TTEST
b. MEANS
c. Both
d. Neither

a. TTEST

HW Chapter 9 Q7

7. Which procedure can produce descriptive statistics for both character and numeric data?

a. TTEST
b. MEANS
c. FREQ
d. None of the above

c. FREQ

HW Chapter 9 Q11

11. PROC CORR can produce which of the following plot types?

a. Fit plots
b. Density curves
c. Box plots
d. Scatter plots

d. Scatter plots

HW Chapter 9 Q12

12. Which MODEL statement will correctly specify a simple linear regression model with the variable Age predicting the variable Sales?

a. `MODEL Age = Sales;`
b. `MODEL Sales = Age;`
c. `MODEL Age * Sales;`
d. `MODEL Sales * Age;`

b. MODEL Sales = Age;

HW Chapter 9 Q17

17. You would like to carry out a paired *t* test to compare the difference in mean heart rate before and after treatment. The data set that you are given contains only one variable that represents the difference between the two measurements. Which procedures and accompanying statements could enable you to carry out this test with this data set? Explain your choices.

Instead of doing paired comparisons, we can test if the difference is significantly different from 0. For example, we can use PROC TTEST H0=; VAR difference;

HW Chapter 9 Q20

20. Explain why it makes sense that PROC UNIVARIATE will create a histogram and PROC FREQ will create bar charts.

PROC FREQ visualizes counts or frequencies of discrete data so it creates bar charts that have gaps in between the bars while PROC UNIVARIATE visualizes numeric data so it makes sense that it creates a histogram (no gaps between bars).

In fact, PROC can deal with both numeric and character data

-3

HW Chapter 9 Q25

25. A study was conducted to see whether taking vitamin E daily would reduce the levels of atherosclerotic disease in a random sample of 500 individuals. Clinical measurements, including thickness of plaque of the carotid artery (taken via ultrasound), were recorded at baseline and at two subsequent visits in a SAS data set called VITE. Patients were divided into two strata according to their baseline plaque measurement.

a. Examine this SAS data set including the variable labels and attributes. Is there evidence to suggest that the true mean systolic blood pressure at baseline is significantly greater than 140 mm/Mg? Carry out an appropriate hypothesis test for each strata, and compare the resulting p-values to alpha = 0.05.

b. You would like to test for differences in true mean plaque before treatment and at the second-year visit. Examine the layout of the data set, and in a comment describe why this data set structure would not work for the hypothesis test that you need to use.

c. Test for a difference in plaque as specified in part b) for the treatment group and also for the placebo group. Carry out appropriate hypothesis tests for each strata, and compare the resulting p-values to alpha = 0.05.

d. To complete the analysis, compare the differences that you saw in mean plaque before and at the second-year visit across the two treatment groups. Carry out appropriate hypothesis tests for each strata, and compare the resulting p-values to alpha = 0.05.

e. Verify the assumption of normality for the tests in parts c) and d). Include plots as well as p-values.

f. In a comment in your program, discuss the results from parts c) and d).

25A

CODE

```
/* 25A */
DATA VITE;
  set '/home/u62223361/Intro to SAS/HW9/vite.sas7bdat';
RUN;

PROC CONTENTS DATA = VITE;
RUN;

PROC PRINT DATA = VITE;
RUN;

PROC TTEST H0= 140 DATA = VITE ALPHA=0.05 SIDES= U;
        CLASS Strata;
        VAR SBP;
/* U specifies upper one-sided tests (in which the alternative hypothesis
indicates a parameter value greater than the null value) and
upper one-sided confidence intervals between the lower confidence limit and infinity. */
```

The CLASS statement is not allowed for a one sample test, however, using a BY statement will break analysis into strata groups

The question A is asking to use the baseline data only

LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 25A */
70         DATA VITE;
71           set '/home/u62223361/Intro to
SAS/HW9/vite.sas7bdat';
NOTE: Data file WC000001.VITE.DATA is in a format that is native
to another host, or the file encoding does not match the session
```

```
      encoding. Cross Environment Data Access will be used,
which might require additional CPU resources and might reduce
      performance.
72          RUN;
NOTE: There were 1500 observations read from the data set
/home/u62223361/Intro to SAS/HW9/vite.sas7bdat.
NOTE: The data set WORK.VITE has 1500 observations and 12
variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              1299.50k
      OS Memory           28332.00k
      Timestamp           05/02/2024 04:25:35 AM
      Step Count                        162  Switch Count  3
      Page Faults                       0
      Page Reclaims                     126
      Page Swaps                        0
      Voluntary Context Switches        24
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           520


73
74          PROC CONTENTS DATA = VITE;
75          RUN;
NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.02 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              1876.25k
```

```
      OS Memory           28332.00k
      Timestamp           05/02/2024 04:25:35 AM
      Step Count                        163  Switch Count  0
      Page Faults                       0
      Page Reclaims                     95
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8

76
77         PROC PRINT DATA = VITE;
78         RUN;
NOTE: There were 1500 observations read from the data set
WORK.VITE.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           1.55 seconds
      user cpu time       1.55 seconds
      system cpu time     0.00 seconds
      memory              825.00k
      OS Memory           28072.00k
      Timestamp           05/02/2024 04:25:36 AM
      Step Count                        164  Switch Count  0
      Page Faults                       0
      Page Reclaims                     64
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           1144
```

```
79
80          PROC TTEST H0= 140 DATA = VITE ALPHA=0.05 SIDES= U;
81          CLASS Strata;
82          VAR SBP;
83          /* U specifies upper one-sided tests (in which the
alternative hypothesis
84          indicates a parameter value greater than the null
value) and
85          upper one-sided confidence intervals between the
lower confidence limit and infinity. */
86
87          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
97
```

User: u62223361

Messages: 20

RESULTS

The CONTENTS Procedure

Data Set Name	WORK.VITE	**Observations**	1500
Member Type	DATA	**Variables**	12
Engine	V9	**Indexes**	0
Created	05/02/2024 00:25:35	**Observation Length**	96
Last Modified	05/02/2024 00:25:35	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	2
First Data Page	1
Max Obs per Page	1363
Obs in First Data Page	1325
Number of Data Set Repairs	0
Filename	/saswork/SAS_workF26B0000EDBF_odaws02-usw2-2.oda.sas.com/SAS_work7DB30000EDBF_odaws02-usw2-2.oda.sas.com/vite.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	2415921496
Access Permission	rw-r--r--
Owner Name	u62223361
File Size	384KB
File Size (bytes)	393216

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
11	Alcohol	Num	8	Number of alcoholic drinks per day
10	DBP	Num	8	Diastolic blood pressure (mm/Mg)
6	HDL	Num	8	HDL cholesterol (mg/DL)
1	ID	Num	8	Subject ID
7	LDL	Num	8	LDL cholesterol (mg/DL)
5	Plaque	Num	8	Plaque measurement (mm)
9	SBP	Num	8	Systolic blood pressure (mm/Mg)
12	Smoke	Num	8	Number of cigarettes smoked per day

12	Smoke	Num	8	Number of cigarettes smoked per day
3	Strata	Num	8	Strata 1=baseline plaque 0.60mm+ and 2=baseline plaque below 0.60mm
4	Treatment	Num	8	0=placebo and 1=vitamin E
8	Trig	Num	8	triglycerides mg/dL
2	Visit	Num	8	0=baseline, 1=first year, and 2=second year

Obs	ID	Visit	Strata	Treatment	Plaque	HDL	LDL	Trig	SBP	DBP	Alcohol	Smoke
1	1	0	1	0	0.8073	42	127	149	106	70	1	0
2	1	1	1	0	0.7580	44	143	49	131	109	2	0
3	1	2	1	0	0.8098	40	158	98	136	87	2	0
4	2	0	1	0	0.7576	39	138	211	157	100	1	6
5	2	1	1	0	0.6866	46	147	29	154	108	2	6
6	2	2	1	0	0.8231	53	161	177	65	70	3	6
7	3	0	1	0	0.7522	53	133	163	169	106	1	0
8	3	1	1	0	0.7857	36	146	198	172	91	2	0
9	3	2	1	0	0.8031	39	119	140	140	90	3	0
10	4	0	1	0	0.8163	46	139	247	142	103	0	0
11	4	1	1	0	0.6004	46	139	47	131	93	0	0
12	4	2	1	0	0.9694	40	150	153	164	96	0	0
13	5	0	1	0	0.7977	47	152	182	163	101	0	3
14	5	1	1	0	0.9573	45	145	90	73	89	0	5
15	5	2	1	0	0.7973	41	142	133	150	59	0	0
16	6	0	1	0	0.8180	46	147	326	145	101	0	0
17	6	1	1	0	0.8699	48	154	90	123	74	0	0
18	6	2	1	0	0.8396	48	138	250	135	88	0	0
19	7	0	1	0	0.8747	48	136	178	148	98	2	0
20	7	1	1	0	0.9285	46	150	213	141	108	3	0
21	7	2	1	0	0.7459	35	120	206	154	85	3	0
22	8	0	1	0	0.8679	55	142	105	136	87	0	0
23	8	1	1	0	0.7639	41	154	54	99	112	0	0
24	8	2	1	0	0.8181	46	159	163	185	83	0	0
25	9	0	1	0	0.8258	51	167	304	176	99	1	0
26	9	1	1	0	0.8427	48	135	313	146	91	1	0
27	9	2	1	0	0.7831	41	112	152	180	75	1	0
28	10	0	1	0	0.7480	55	126	202	90	81	0	6
29	10	1	1	0	0.8058	48	139	276	146	126	0	7

There is no results from the PROC TTEST

-3

25B

/* 25B */

/* The current data structure would not allow for testing for difference between true mean plaque before treatment and at the second visit because they are different variables and thus one cannot stratify via T-test. */

25C

CODE

/* 25C */

DATA VITE_0 (RENAME = (

```
        Visit = Visit_0
        Plaque = Plaque_0
        ));
        IF VISIT = 0;
        SET VITE (KEEP = Plaque Visit Treatment ID);
RUN;

DATA VITE_2 (RENAME = (
        Visit = Visit_2
        Plaque = Plaque_2
        ));
        IF VISIT = 2;
        SET VITE (KEEP = Plaque Visit Treatment ID);
RUN;

DATA VITE_VISIT_TEST;
        MERGE VITE_0 VITE_2;
        BY ID;
RUN;


PROC SORT DATA = VITE_VISIT_TEST;
        BY Treatment;
RUN;

PROC TTEST H0= 0 DATA = VITE_VISIT_TEST ALPHA=0.05 SIDES= 2;
        BY Treatment;        by strata treatment;
        PAIRED Plaque_0 * Plaque_2;
RUN;
```

Forget strata variable

```
PROC SORT DATA = ???;
BY Strata Treatment;
RUN;
```

LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 25C */
70         DATA VITE_0 (RENAME = (
71         Visit = Visit_0
72         Plaque = Plaque_0
73         ));
74         SET VITE (KEEP = Plaque Visit Treatment ID);
75         IF VISIT = 0;
76         RUN;
NOTE: There were 1500 observations read from the data set
WORK.VITE.
NOTE: The data set WORK.VITE_0 has 500 observations and 4
variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              1231.56k
      OS Memory           28332.00k
      Timestamp           05/02/2024 04:30:59 AM
      Step Count                        210  Switch Count  2
      Page Faults                       0
      Page Reclaims                     126
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264
```

```
77
78          DATA VITE_2 (RENAME = (
79          Visit = Visit_2
80          Plaque = Plaque_2
81          ));
82          SET VITE (KEEP = Plaque Visit Treatment ID);
83          IF VISIT = 2;
84          RUN;
NOTE: There were 1500 observations read from the data set
WORK.VITE.
NOTE: The data set WORK.VITE_2 has 500 observations and 4
variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1203.87k
      OS Memory           28332.00k
      Timestamp           05/02/2024 04:30:59 AM
      Step Count                        211  Switch Count  2
      Page Faults                       0
      Page Reclaims                     123
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264

85
86          DATA VITE_VISIT_TEST;
87          MERGE VITE_0 VITE_2;
88          BY ID;
```

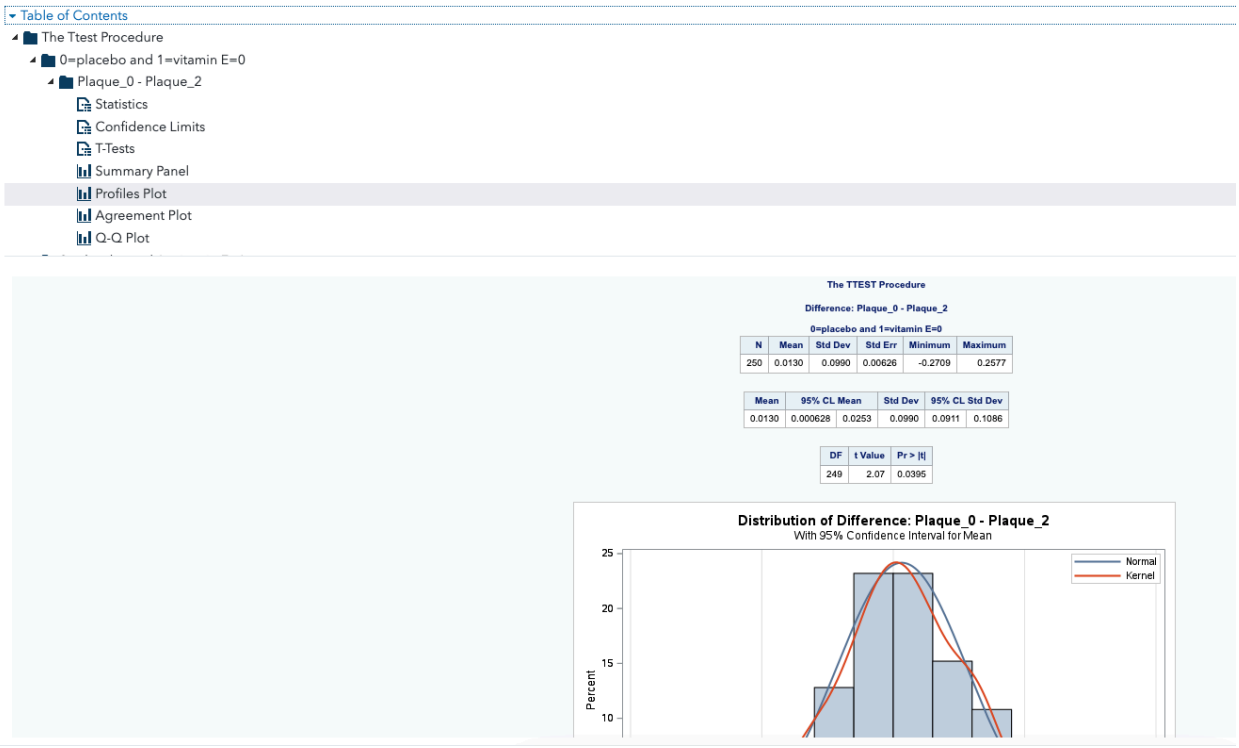
```
89          RUN;
NOTE: There were 500 observations read from the data set
WORK.VITE_0.
NOTE: There were 500 observations read from the data set
WORK.VITE_2.
NOTE: The data set WORK.VITE_VISIT_TEST has 500 observations and
6 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1381.59k
      OS Memory           28592.00k
      Timestamp           05/02/2024 04:30:59 AM
      Step Count                        212  Switch Count  2
      Page Faults                       0
      Page Reclaims                     167
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272

90
91
92          PROC SORT DATA = VITE_VISIT_TEST;
93          BY Treatment;
94          RUN;
NOTE: There were 500 observations read from the data set
WORK.VITE_VISIT_TEST.
NOTE: The data set WORK.VITE_VISIT_TEST has 500 observations and
6 variables.
```

```
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              929.65k
      OS Memory           28332.00k
      Timestamp           05/02/2024 04:30:59 AM
      Step Count                        213  Switch Count  2
      Page Faults                       0
      Page Reclaims                     112
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


95
96          PROC TTEST H0= 0 DATA = VITE_VISIT_TEST ALPHA=0.05
SIDES= 2;
97          BY Treatment;
98          PAIRED Plaque_0 * Plaque_2;
99          RUN;
NOTE: PROCEDURE TTEST used (Total process time):
      real time           0.76 seconds
      user cpu time       0.38 seconds
      system cpu time     0.12 seconds
      memory              26322.65k
      OS Memory           58816.00k
      Timestamp           05/02/2024 04:30:59 AM
      Step Count                        214  Switch Count  60
      Page Faults                       0
      Page Reclaims                     51906
```

```
Page Swaps                          0
Voluntary Context Switches          4877
Involuntary Context Switches        1
Block Input Operations              0
Block Output Operations             5152
```

```
100
101         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
111
```

RESULTS



The TTEST Procedure

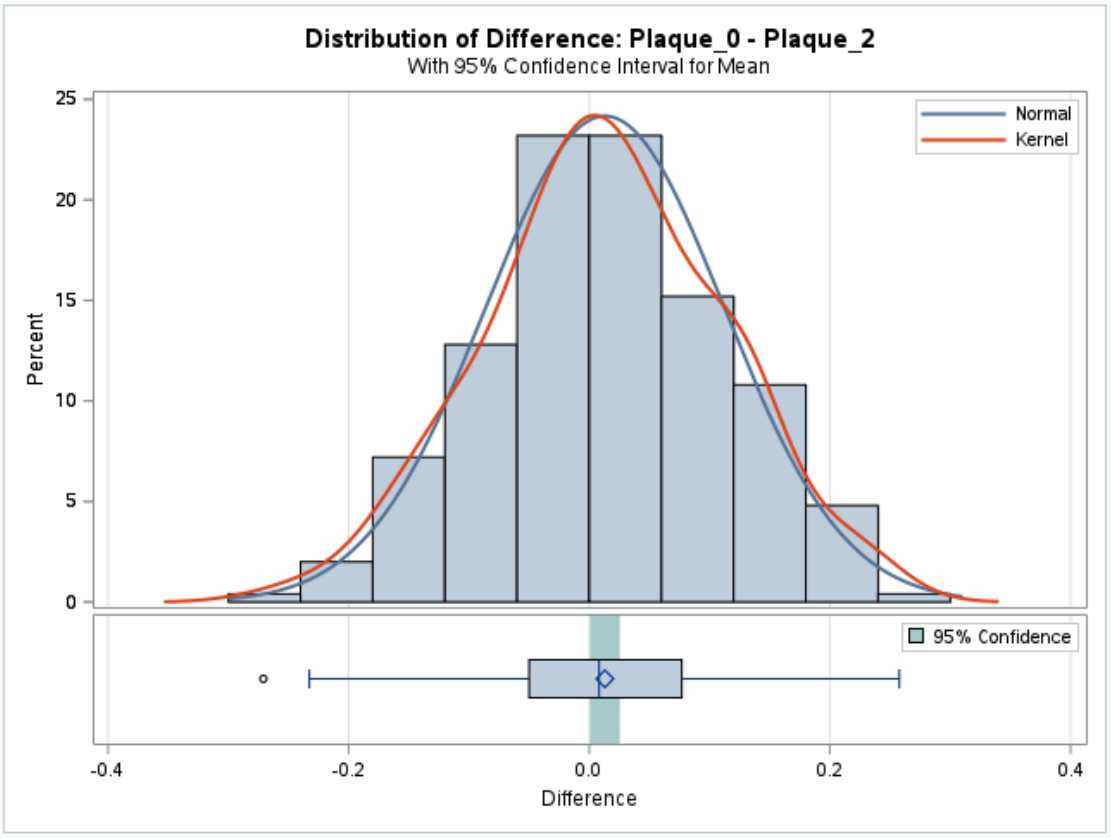
Difference: Plaque_0 - Plaque_2

0=placebo and 1=vitamin E=0

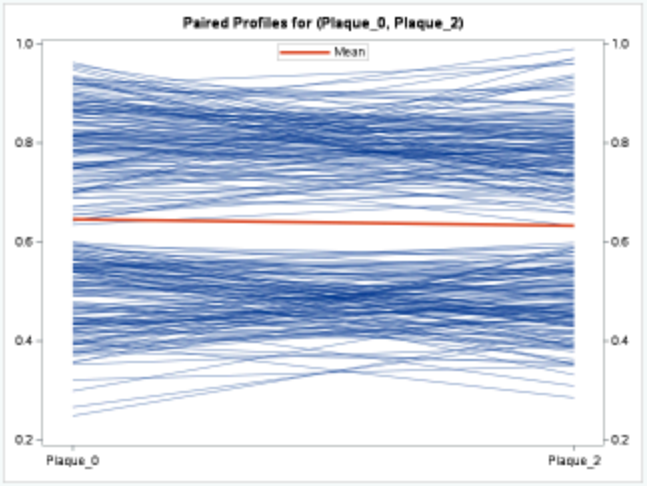
N	Mean	Std Dev	Std Err	Minimum	Maximum
250	0.0130	0.0990	0.00626	-0.2709	0.2577

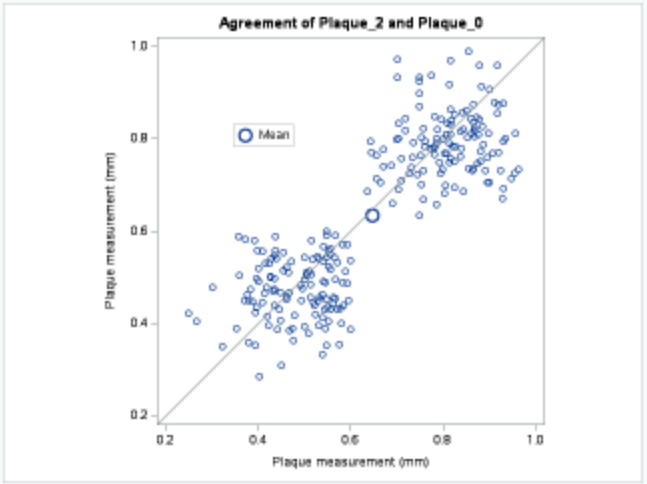
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.0130	0.000628	0.0253	0.0990	0.0911	0.1086

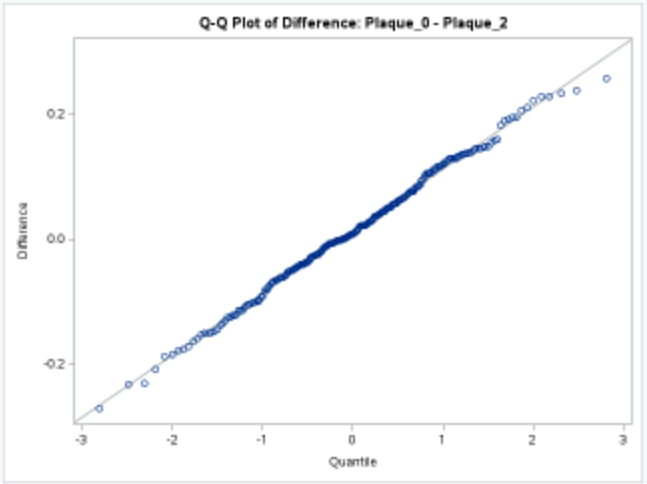
DF	t Value	Pr > |t|
249	2.07	0.0395




-0.4
-0.2
0.0
0.2
0.4
Difference


Paired Profiles for (Plaque_0, Plaque_2)
Mean
1.0
0.8
0.6
0.4
0.2
Plaque_0
Plaque_2


Agreement of Plaque_2 and Plaque_0
Mean
Plaque measurement (mm)
Plaque measurement (mm)
0.2
0.4
0.6
0.8
1.0


Q-Q Plot of Difference: Plaque_0 - Plaque_2
Difference
Quantile
0.2
0.0
-0.2
-3
-2
-1
0
1
2
3

The TTEST Procedure

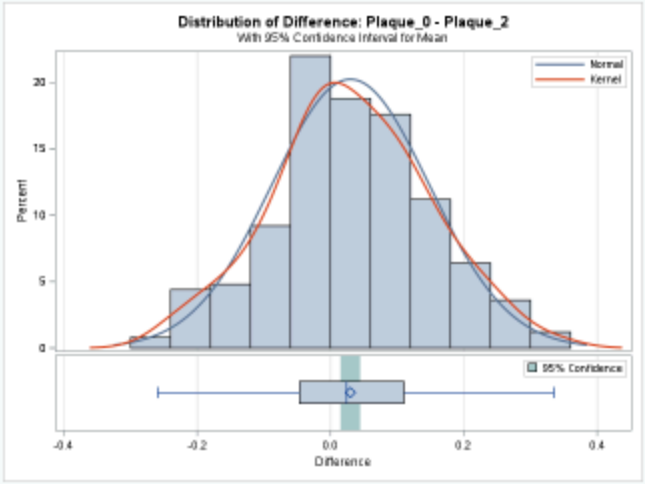
Difference: Plaque_0 - Plaque_2

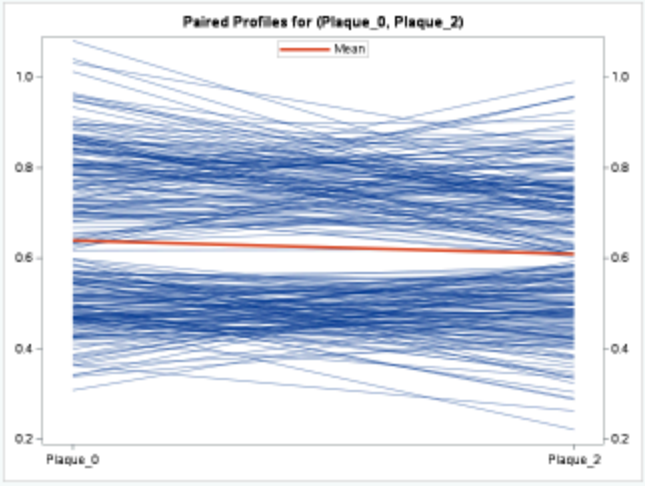
0=placebo and 1=vitamin E=1

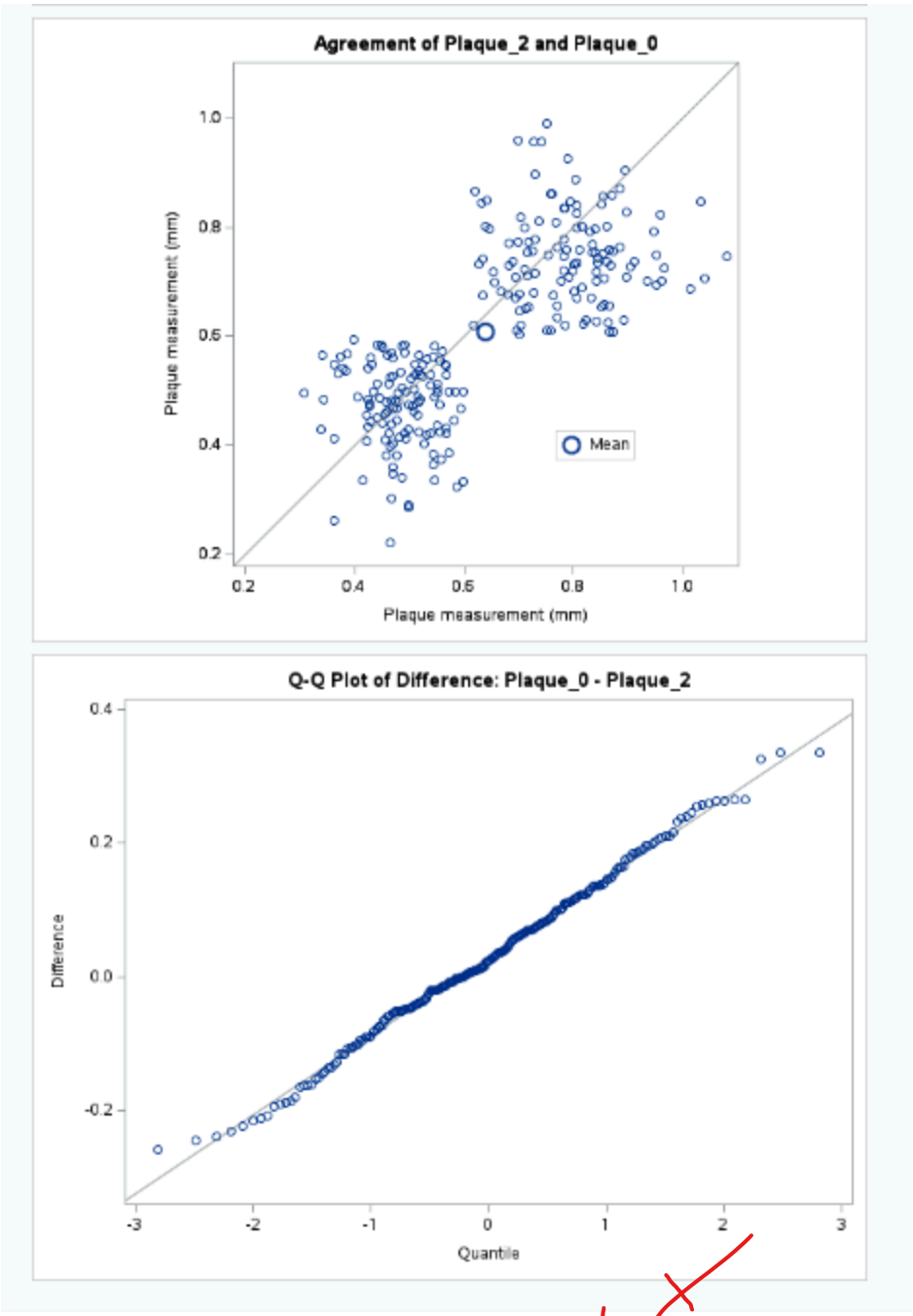
N	Mean	Std Dev	Std Err	Minimum	Maximum
250	0.0298	0.1162	0.00748	-0.2590	0.3351

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.0298	0.0150	0.0445	0.1162	0.1067	0.1296

DF	t Value	Pr > |t|
249	3.98	<.0001







-2

25D

CODE

/* 25D */

PROC SORT DATA = VITE_VISIT_TEST;

```
        BY Treatment;
RUN;
PROC TTEST H0= 0 DATA = VITE_VISIT_TEST ALPHA=0.05 SIDES= 2;
        PAIRED Plaque_0 * Plaque_2;
RUN;
```

25 D is to test the difference (Diff=Plaque_2-Plague_0) across the two treatment groups for by strata.

Here are the correct codes

```
PROC TTEST DATA = ???;
VAR Diff;
CLASS Treatment;
BY Strata;
```

-5

LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 25D */
70         PROC SORT DATA = VITE_VISIT_TEST;
71         BY Treatment;
72         RUN;
NOTE: Input data set is already sorted, no sorting done.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              536.75k
      OS Memory           28328.00k
      Timestamp           05/02/2024 04:33:18 AM
      Step Count                        227  Switch Count  0
      Page Faults                       0
      Page Reclaims                     50
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0
```

```
73          PROC TTEST H0= 0 DATA = VITE_VISIT_TEST ALPHA=0.05
SIDES= 2;
74          PAIRED Plaque_0 * Plaque_2;
75          RUN;
NOTE: PROCEDURE TTEST used (Total process time):
      real time           0.49 seconds
      user cpu time       0.25 seconds
      system cpu time     0.07 seconds
      memory              24172.25k
      OS Memory           56492.00k
      Timestamp           05/02/2024 04:33:18 AM
      Step Count                        228  Switch Count  30
      Page Faults                       0
      Page Reclaims                     29501
      Page Swaps                        0
      Voluntary Context Switches        3952
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           3144


76
77          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
87
```

RESULTS

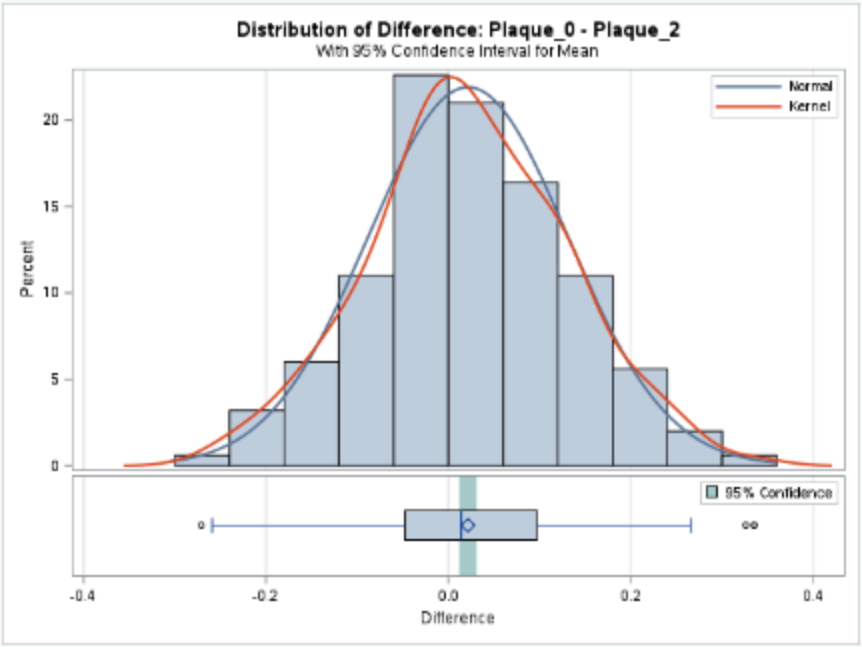
The TTEST Procedure

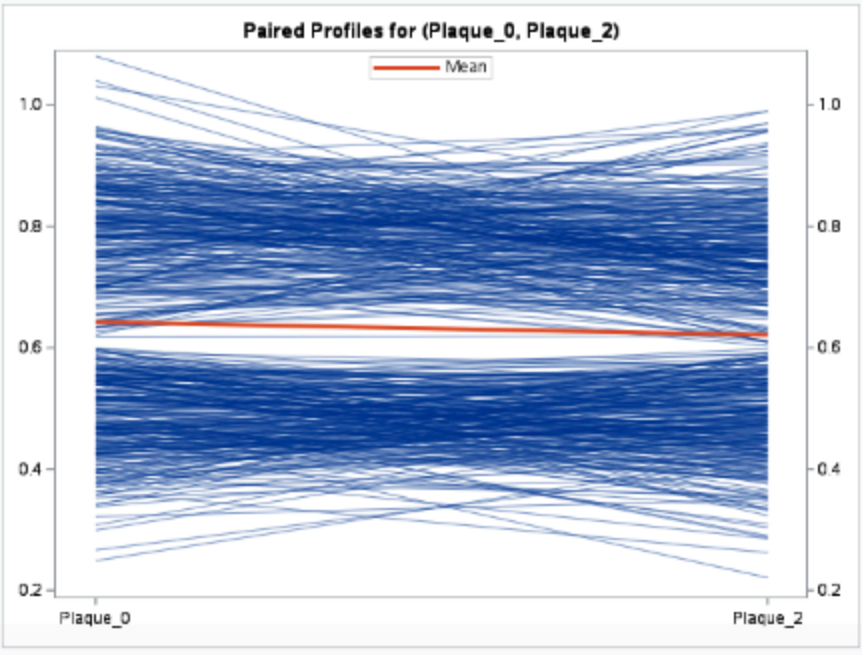
Difference: Plaque_0 - Plaque_2

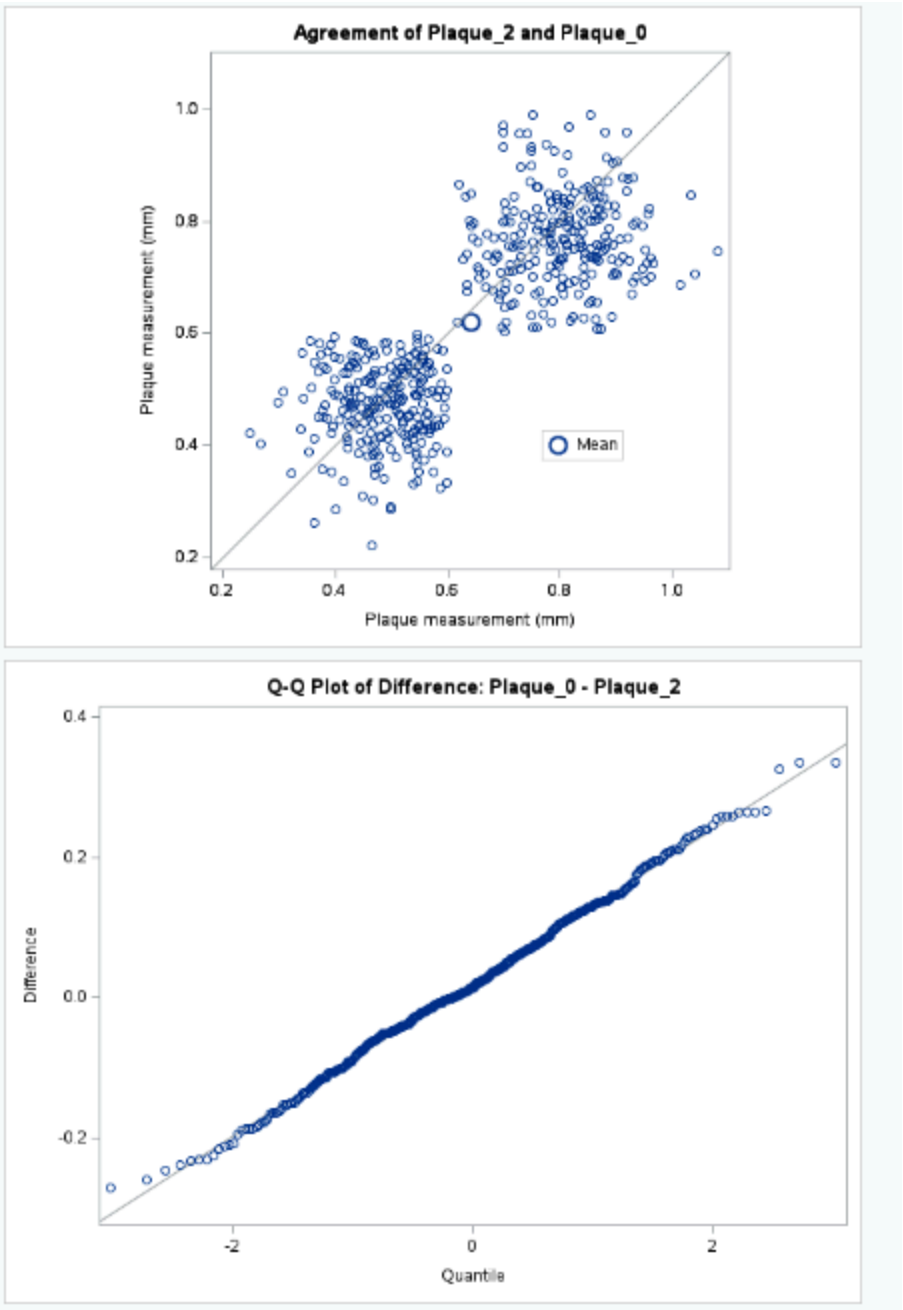
N	Mean	Std Dev	Std Err	Minimum	Maximum
500	0.0214	0.1093	0.00489	-0.2709	0.3351

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.0214	0.0118	0.0310	0.1093	0.1029	0.1165

DF	t Value	Pr > |t|
499	4.37	<.0001






Agreement of Plaque_2 and Plaque_0
Plaque measurement (mm)
1.0
0.8
0.5
0.4
0.2
Mean
0.2
0.4
0.5
0.8
1.0
Plaque measurement (mm)
Q-Q Plot of Difference: Plaque_0 - Plaque_2
Difference
0.4
0.2
0.0
-0.2
-2
0
2
Quantile

25E

CODE

/* 25E */

```
PROC SORT DATA = VITE_VISIT_TEST;
        BY Treatment;
RUN;

PROC TTEST PLOTS(ONLY) = (ALL) H0= 0 DATA = VITE_VISIT_TEST ALPHA=0.05
SIDES= 2;
        BY Treatment;
        PAIRED Plaque_0 * Plaque_2;
RUN;

PROC SORT DATA = VITE_VISIT_TEST;
        BY Treatment;
RUN;
PROC TTEST PLOTS(ONLY) = (ALL) H0= 0 DATA = VITE_VISIT_TEST ALPHA=0.05
SIDES= 2;
        PAIRED Plaque_0 * Plaque_2;
RUN;

PROC UNIVARIATE DATA = VITE_VISIT_TEST NORMALTEST NORMAL;
        VAR Plaque_0 Plaque_2 Treatment;
        CDFPLOT Treatment / NORMAL;
        HISTOGRAM Treatment / NORMAL;
        PPPLOT Treatment / NORMAL;
        PROBPLOT Treatment / NORMAL;
        QQPLOT Treatment / NORMAL;

        CDFPLOT Plaque_0 / NORMAL;
        HISTOGRAM Plaque_0 / NORMAL;
        PPPLOT Plaque_0 / NORMAL;
        PROBPLOT Plaque_0 / NORMAL;
        QQPLOT Plaque_0 / NORMAL;
```

There is no need to test distribution of treatment because it is a binary variable.

there is no need for all the plots listed here. All you need is the normal OPTION in PROC UNIVARIATE statement

```
        CDFPLOT Plaque_2 / NORMAL;
        HISTOGRAM Plaque_2 / NORMAL;
        PPPLOT Plaque_2 / NORMAL;
        PROBPLOT Plaque_2 / NORMAL;
        QQPLOT Plaque_2 / NORMAL;
RUN;
```

should also test normality for the difference

-2

LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 25E */
70
71         PROC SORT DATA = VITE_VISIT_TEST;
72         BY Treatment;
73         RUN;
NOTE: Input data set is already sorted, no sorting done.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              535.09k
      OS Memory           28584.00k
      Timestamp           05/02/2024 04:34:22 AM
      Step Count                        234  Switch Count  0
      Page Faults                       0
      Page Reclaims                     50
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
```

```
      Block Input Operations            0
      Block Output Operations           0


74
75          PROC TTEST PLOTS(ONLY) = (ALL) H0= 0 DATA =
VITE_VISIT_TEST ALPHA=0.05 SIDES= 2;
76          BY Treatment;
77          PAIRED Plaque_0 * Plaque_2;
78          RUN;
NOTE: PROCEDURE TTEST used (Total process time):
      real time           1.10 seconds
      user cpu time       0.52 seconds
      system cpu time     0.15 seconds
      memory              29425.03k
      OS Memory           62400.00k
      Timestamp           05/02/2024 04:34:23 AM
      Step Count                        235  Switch Count  60
      Page Faults                       0
      Page Reclaims                     54648
      Page Swaps                        0
      Voluntary Context Switches        5917
      Involuntary Context Switches      83
      Block Input Operations            0
      Block Output Operations           7584


79
80          PROC SORT DATA = VITE_VISIT_TEST;
81          BY Treatment;
82          RUN;
NOTE: Input data set is already sorted, no sorting done.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
```

```
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              553.53k
      OS Memory           36788.00k
      Timestamp           05/02/2024 04:34:23 AM
      Step Count                       236  Switch Count  0
      Page Faults                      0
      Page Reclaims                    51
      Page Swaps                       0
      Voluntary Context Switches       1
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          0

83         PROC TTEST PLOTS(ONLY) = (ALL) H0= 0 DATA =
VITE_VISIT_TEST ALPHA=0.05 SIDES= 2;
84         PAIRED Plaque_0 * Plaque_2;
85         RUN;
NOTE: PROCEDURE TTEST used (Total process time):
      real time           0.61 seconds
      user cpu time       0.30 seconds
      system cpu time     0.08 seconds
      memory              20650.31k
      OS Memory           61120.00k
      Timestamp           05/02/2024 04:34:24 AM
      Step Count                       237  Switch Count  30
      Page Faults                      0
      Page Reclaims                    28808
      Page Swaps                       0
      Voluntary Context Switches       4480
      Involuntary Context Switches     130
      Block Input Operations           0
```

```
      Block Output Operations           3840


86
87          PROC UNIVARIATE DATA = VITE_VISIT_TEST NORMALTEST
NORMAL;
88          VAR Plaque_0 Plaque_2 Treatment;
89          CDFPLOT Treatment / NORMAL;
90          HISTOGRAM Treatment / NORMAL;
91          PPPLOT Treatment / NORMAL;
92          PROBPLOT Treatment / NORMAL;
93          QQPLOT Treatment / NORMAL;
94
95          CDFPLOT Plaque_0 / NORMAL;
96          HISTOGRAM Plaque_0 / NORMAL;
97          PPPLOT Plaque_0 / NORMAL;
98          PROBPLOT Plaque_0 / NORMAL;
99          QQPLOT Plaque_0 / NORMAL;
100
101         CDFPLOT Plaque_2 / NORMAL;
102         HISTOGRAM Plaque_2 / NORMAL;
103         PPPLOT Plaque_2 / NORMAL;
104         PROBPLOT Plaque_2 / NORMAL;
105         QQPLOT Plaque_2 / NORMAL;
106         RUN;
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.99 seconds
      user cpu time       0.68 seconds
      system cpu time     0.07 seconds
      memory              25563.28k
      OS Memory           59928.00k
      Timestamp           05/02/2024 04:34:25 AM
      Step Count                        238  Switch Count  0
```

Page Faults	0
Page Reclaims	12513
Page Swaps	0
Voluntary Context Switches	3486
Involuntary Context Switches	193
Block Input Operations	0
Block Output Operations	5752

```
107
108         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
118
```

RESULTS

The TTEST Procedure

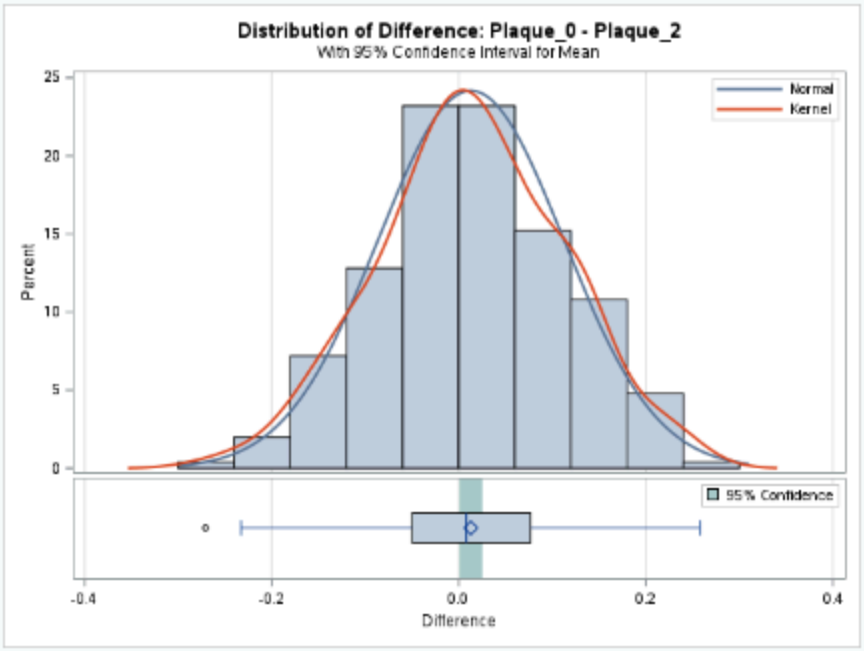
Difference: Plaque_0 - Plaque_2

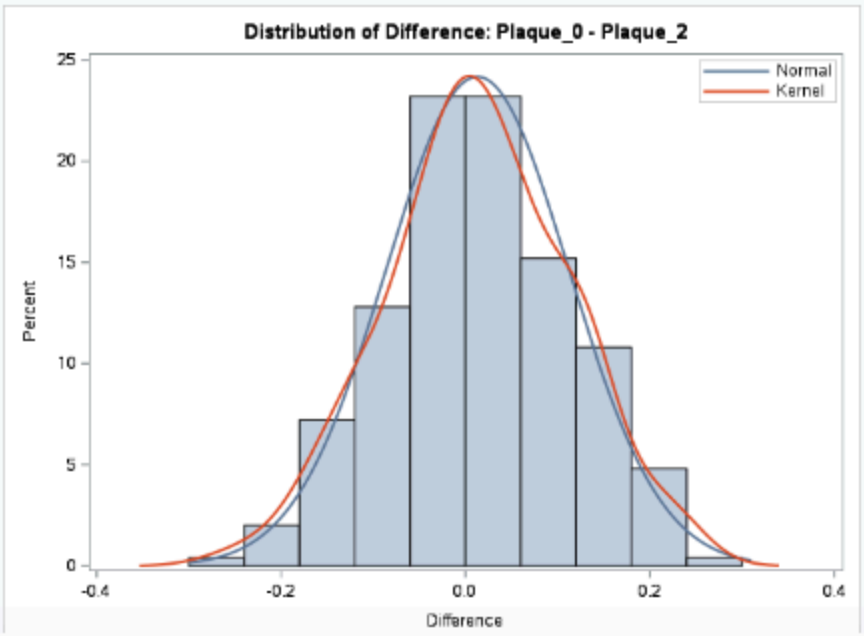
0=placebo and 1=vitamin E=0

N	Mean	Std Dev	Std Err	Minimum	Maximum
250	0.0130	0.0990	0.00626	-0.2709	0.2577

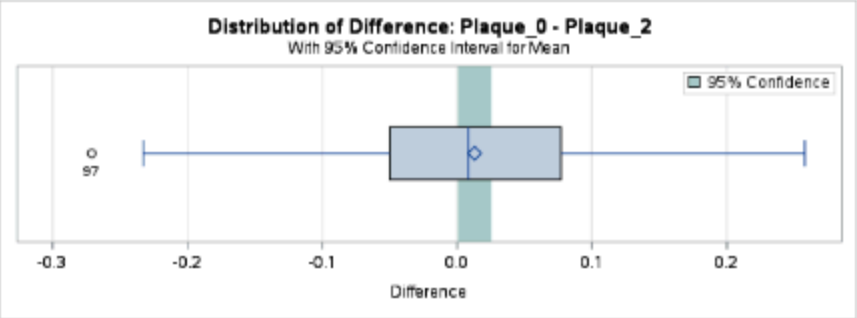
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.0130	0.000528	0.0253	0.0990	0.0911	0.1086

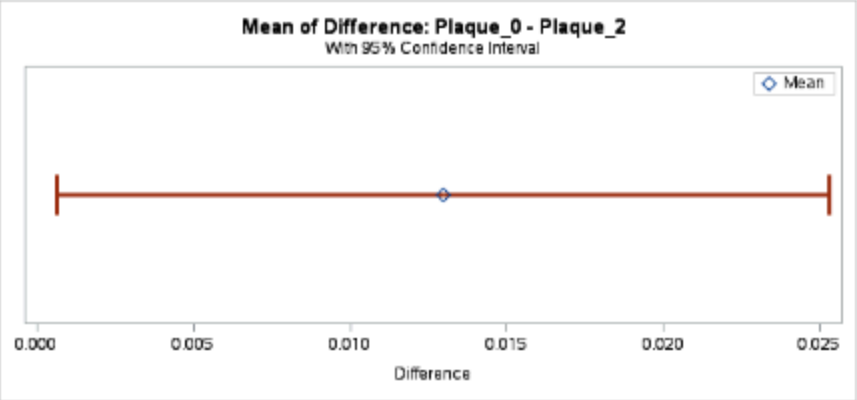
DF	t Value	Pr > |t|
249	2.07	0.0395

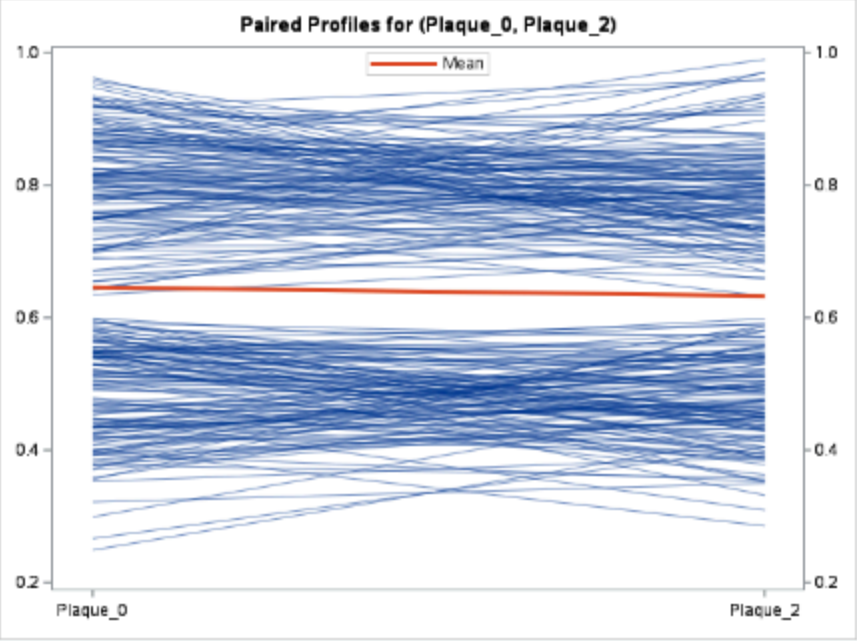


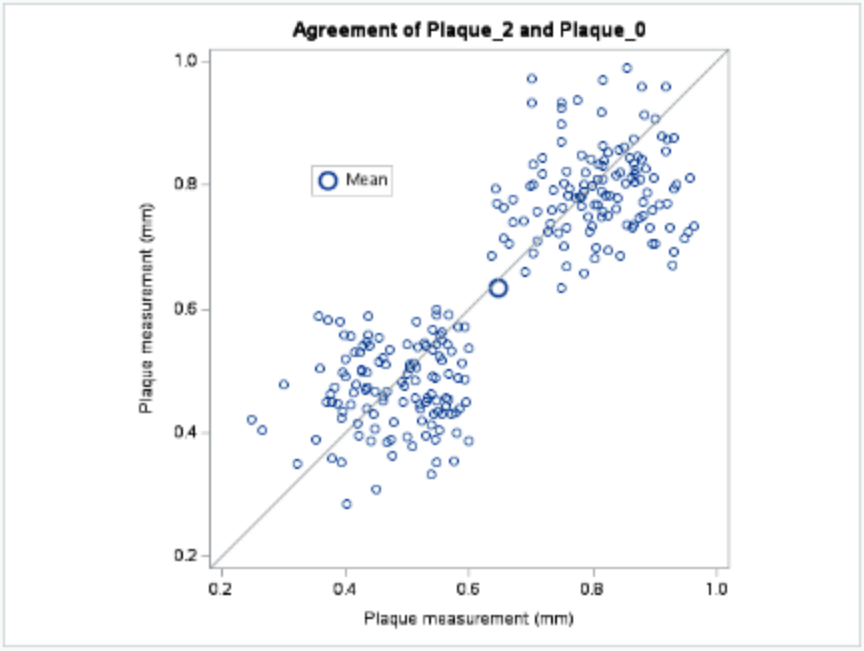


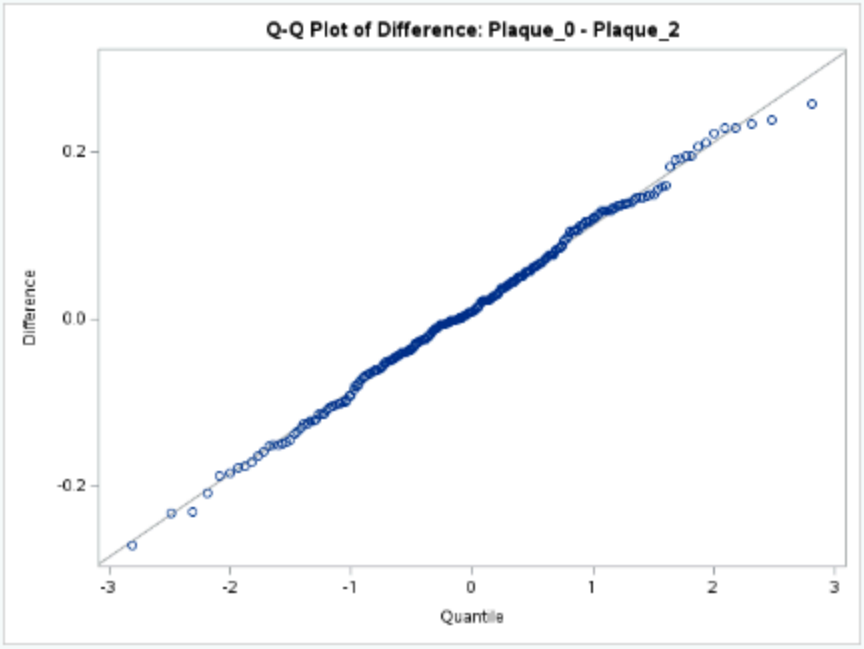
this is data distribution illustration rather than testing for normality. You need to use PROC UNIVARITE to test normality of the difference between Plaque _0 and _2.


Distribution of Difference: Plaque_0 - Plaque_2
With 95% Confidence Interval for Mean
95% Confidence
97
-0.3
-0.2
-0.1
0.0
0.1
0.2
Difference


Mean of Difference: Plaque_0 - Plaque_2
With 95% Confidence Interval
Mean
0.000
0.005
0.010
0.015
0.020
0.025
Difference


Paired Profiles for (Plaque_0, Plaque_2)
Mean
1.0
0.8
0.6
0.4
0.2
Plaque_0
Plaque_2


Agreement of Plaque_2 and Plaque_0
Mean
Plaque measurement (mm)
Plaque measurement (mm)
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0


Q-Q Plot of Difference: Plaque_0 - Plaque_2
Difference
Quantile
0.2
0.0
-0.2
-3
-2
-1
0
1
2
3

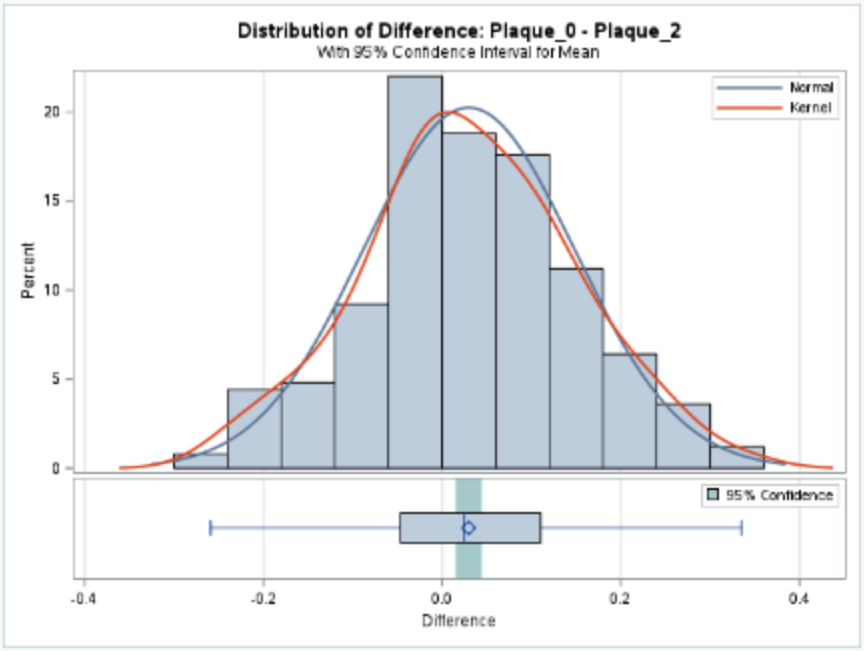
Difference: Plaque_0 - Plaque_2

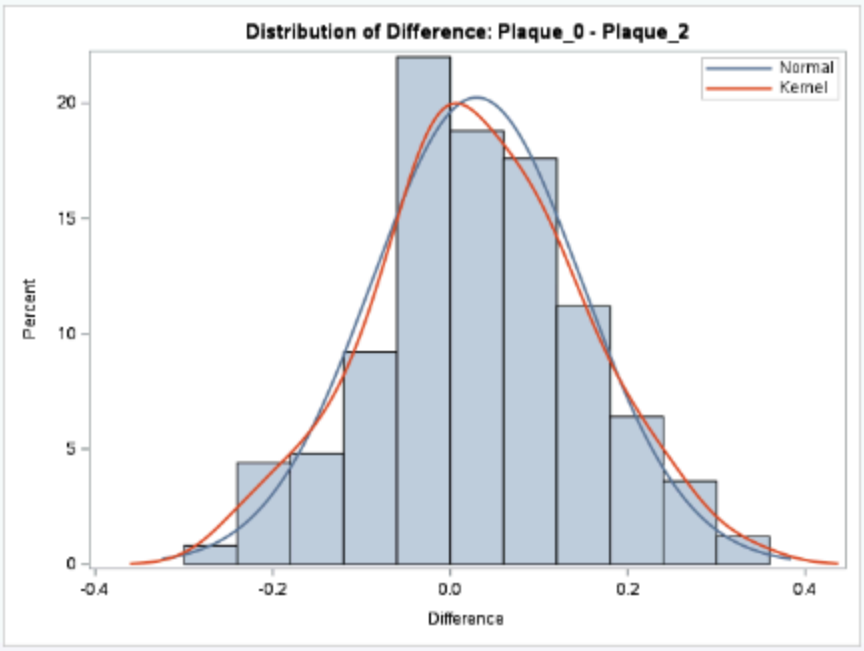
0=placebo and 1=vitamin E=1

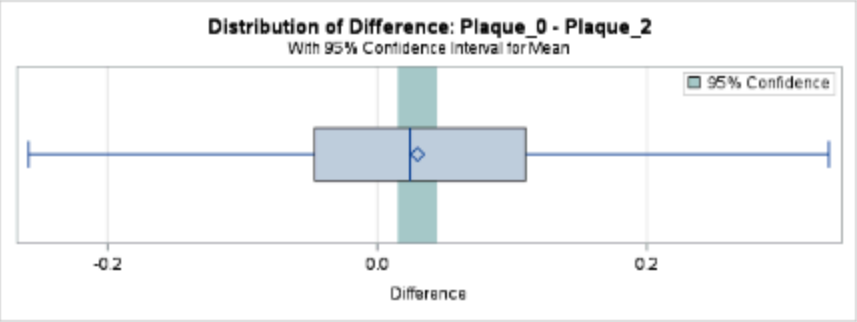
N	Mean	Std Dev	Std Err	Minimum	Maximum
250	0.0298	0.1182	0.00748	-0.2590	0.3351

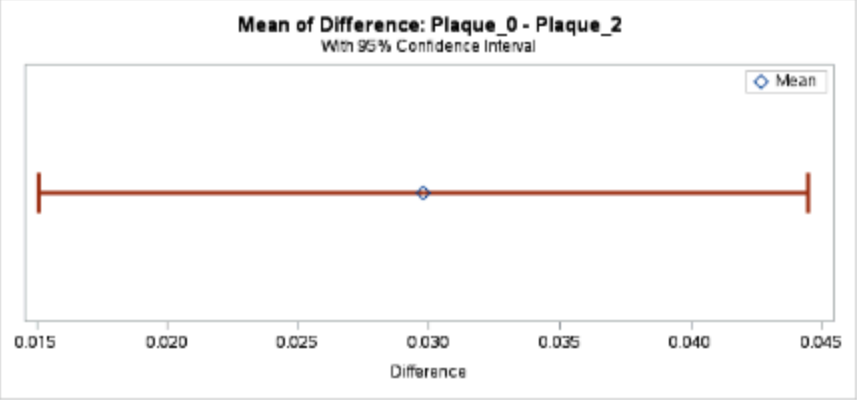
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.0298	0.0150	0.0445	0.1182	0.1087	0.1296

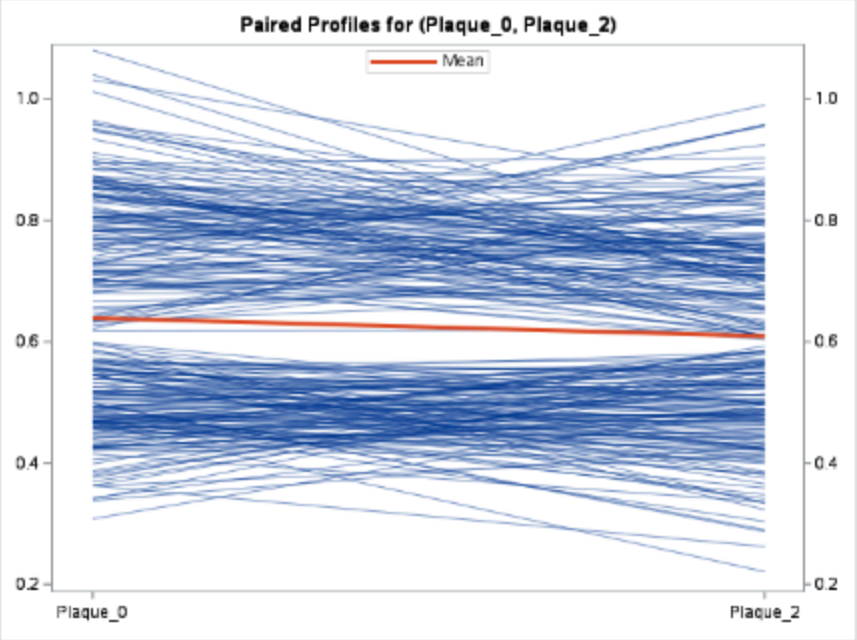
DF	t Value	Pr > |t|
249	3.98	<.0001

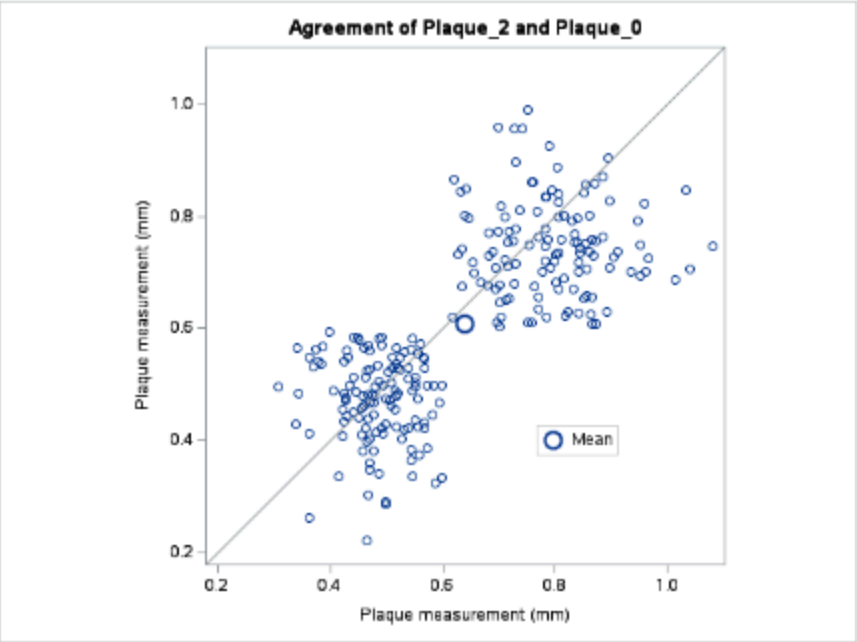


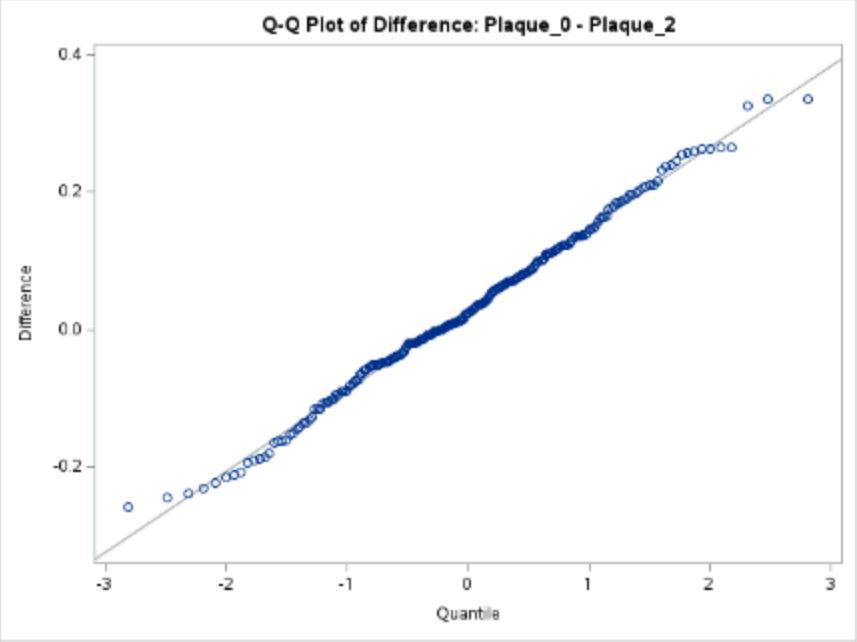



Distribution of Difference: Plaque_0 - Plaque_2
With 95% Confidence Interval for Mean
95% Confidence
-0.2
0.0
0.2
Difference


Mean of Difference: Plaque_0 - Plaque_2
With 95% Confidence Interval
Mean
0.015
0.020
0.025
0.030
0.035
0.040
0.045
Difference


Paired Profiles for (Plaque_0, Plaque_2)
Mean
1.0
0.8
0.6
0.4
0.2
Plaque_0
Plaque_2


Agreement of Plaque_2 and Plaque_0
Plaque measurement (mm)
1.0
0.8
0.6
0.4
0.2
Mean
0.2
0.4
0.6
0.8
1.0
Plaque measurement (mm)


Q-Q Plot of Difference: Plaque_0 - Plaque_2
Difference
0.4
0.2
0.0
-0.2
-3
-2
-1
0
1
2
3
Quantile

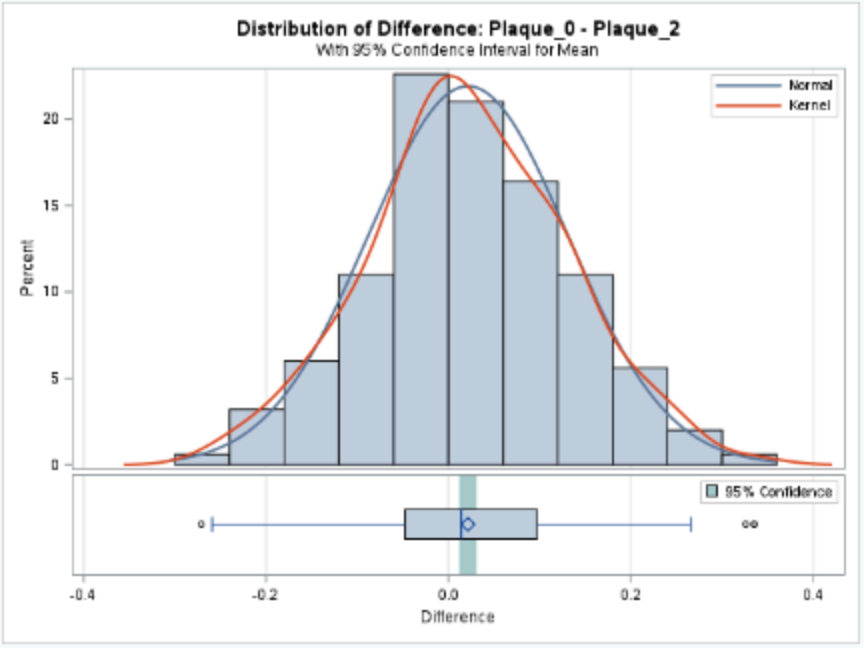
The TTEST Procedure

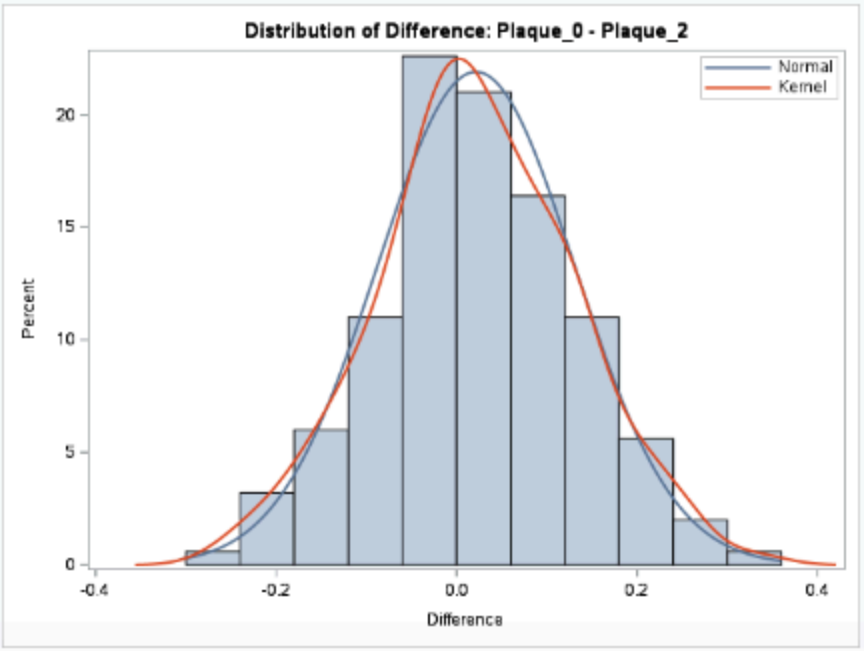
Difference: Plaque_0 - Plaque_2

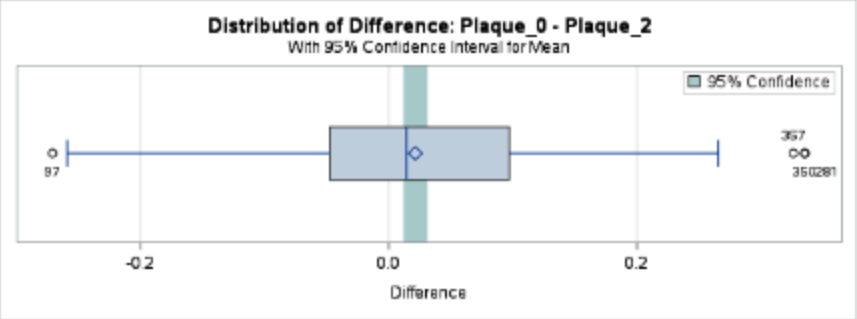
N	Mean	Std Dev	Std Err	Minimum	Maximum
500	0.0214	0.1093	0.00489	-0.2709	0.3351

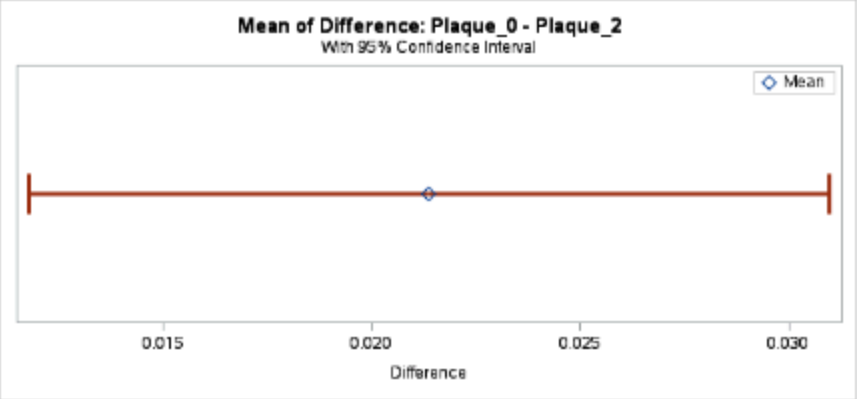
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.0214	0.0118	0.0310	0.1093	0.1029	0.1165

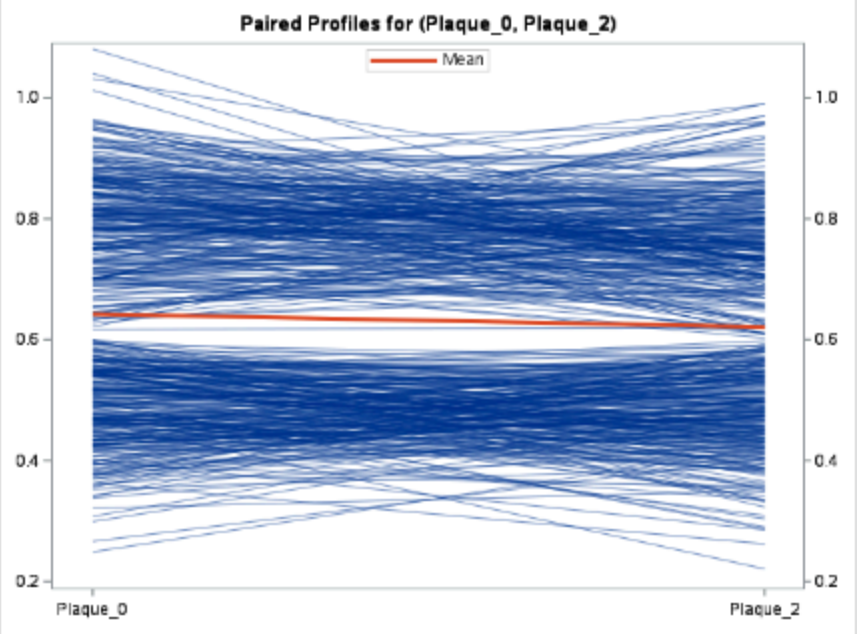
DF	t Value	Pr > |t|
499	4.37	<.0001

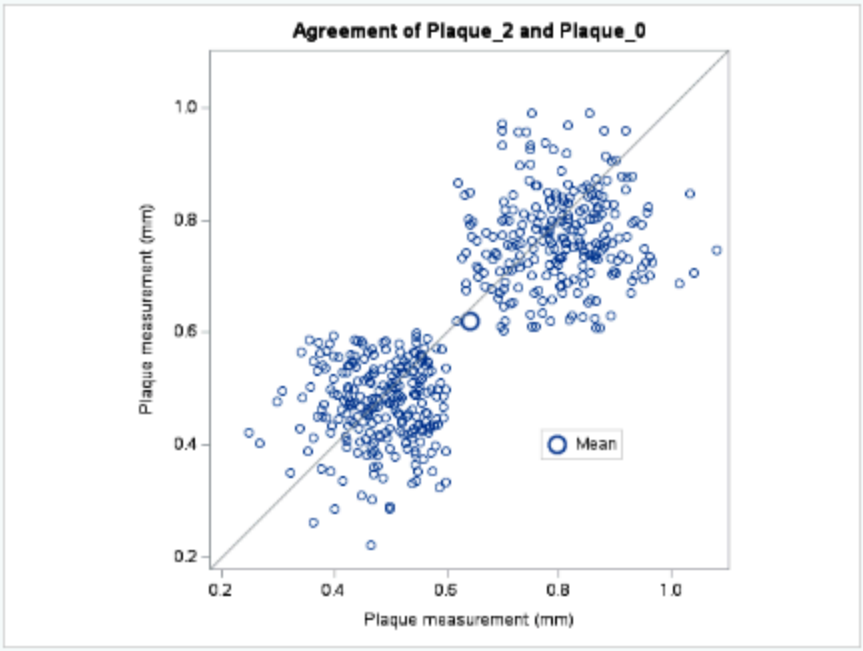


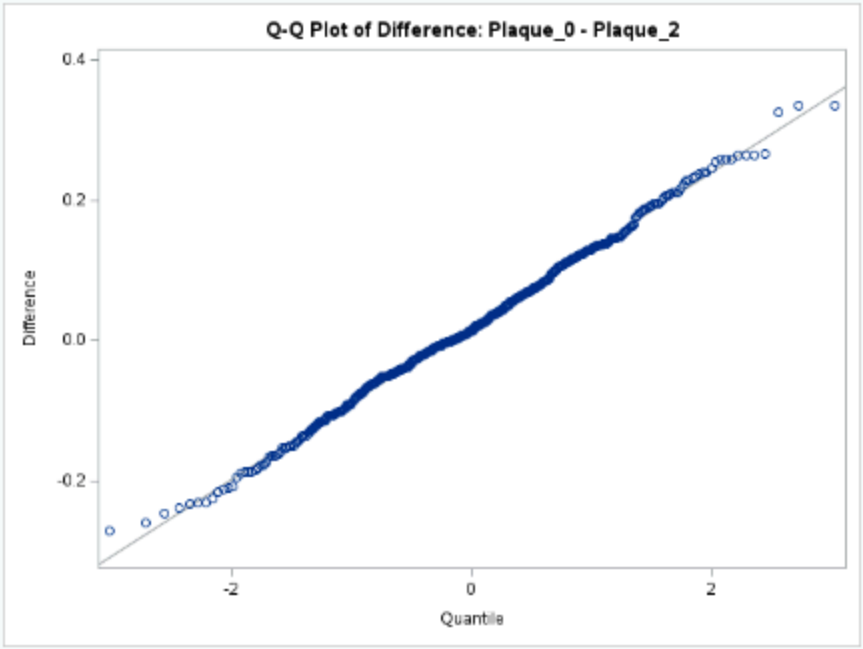



Distribution of Difference: Plaque_0 - Plaque_2
With 95% Confidence Interval for Mean
95% Confidence
357
97
350281
-0.2
0.0
0.2
Difference


Mean of Difference: Plaque_0 - Plaque_2
With 95% Confidence Interval
Mean
0.015
0.020
0.025
0.030
Difference


Paired Profiles for (Plaque_0, Plaque_2)
Mean
1.0
0.8
0.6
0.4
0.2
Plaque_0
Plaque_2


Agreement of Plaque_2 and Plaque_0
Plaque measurement (mm)
1.0
0.8
0.6
0.4
0.2
Mean
0.2
0.4
0.6
0.8
1.0
Plaque measurement (mm)


Q-Q Plot of Difference: Plaque_0 - Plaque_2
Difference
0.4
0.2
0.0
-0.2
-2
0
2
Quantile

The UNIVARIATE Procedure
Variable: Plaque_0 (Plaque measurement (mm))

Moments			
N	500	Sum Weights	500
Mean	0.6425344	Sum Observations	321.2672
Std Deviation	0.17873088	Variance	0.03194473
Skewness	0.11189188	Kurtosis	-1.2139348
Uncorrected SS	222.365647	Corrected SS	15.9404193
Coeff Variation	27.8165467	Std Error Mean	0.00799309

Basic Statistical Measures			
Location		Variability	
Mean	0.642534	Std Deviation	0.17873
Median	0.608300	Variance	0.03194
Mode	0.398800	Range	0.83180
		Interquartile Range	0.31850

Note: The mode displayed is the smallest of 23 modes with a count of 2.

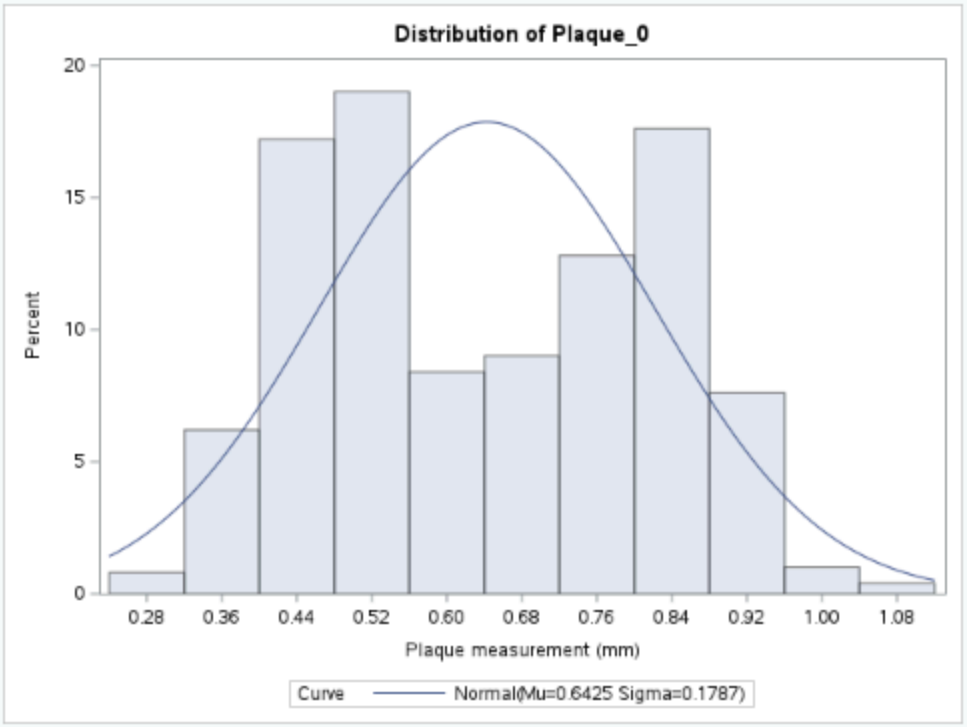
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	80.38625	Pr > |t|	<.0001
Sign	M	250	Pr >= |M|	<.0001
Signed Rank	S	62625	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.950574	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.116282	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.859763	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	10.15841	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	1.08080
99%	0.96350
95%	0.91450
90%	0.87425
75% Q3	0.80590
50% Median	0.60830
25% Q1	0.48740
10%	0.42515
5%	0.38635
1%	0.32985
0% Min	0.24900

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.2490	156	0.9648	316
0.2664	218	1.0127	357
0.2992	203	1.0317	354
0.3080	454	1.0405	281
0.3215	175	1.0808	350

The UNIVARIATE Procedure



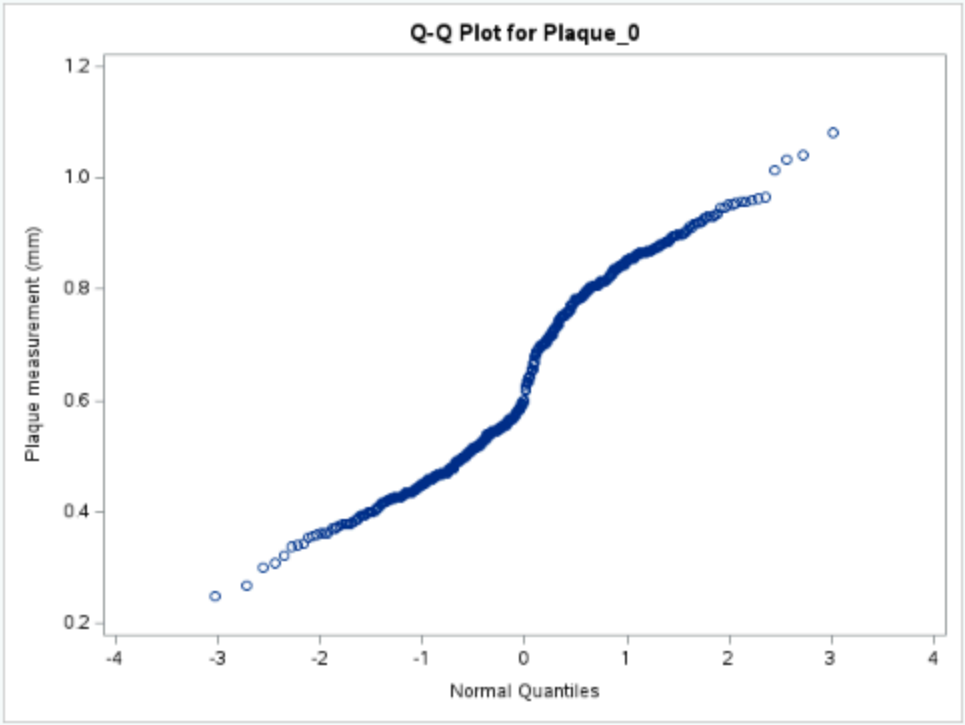
The UNIVARIATE Procedure
Fitted Normal Distribution for Plaque_0 (Plaque measurement (mm))

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	0.642534
Std Dev	Sigma	0.178731

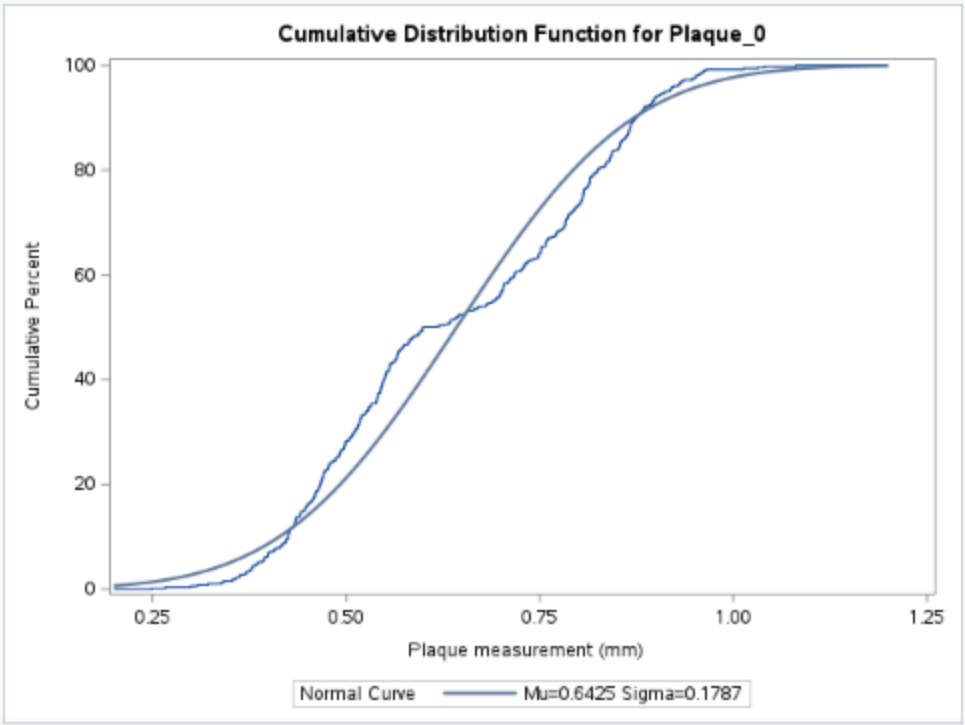
Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.1162819	Pr > D	<0.010
Cramer-von Mises	W-Sq	1.8597630	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	10.1584140	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	0.32985	0.22674
5.0	0.38635	0.34855
10.0	0.42515	0.41348
25.0	0.48740	0.52198
50.0	0.60830	0.64253
75.0	0.80590	0.76309
90.0	0.87425	0.87159
95.0	0.91450	0.93652
99.0	0.96350	1.05832

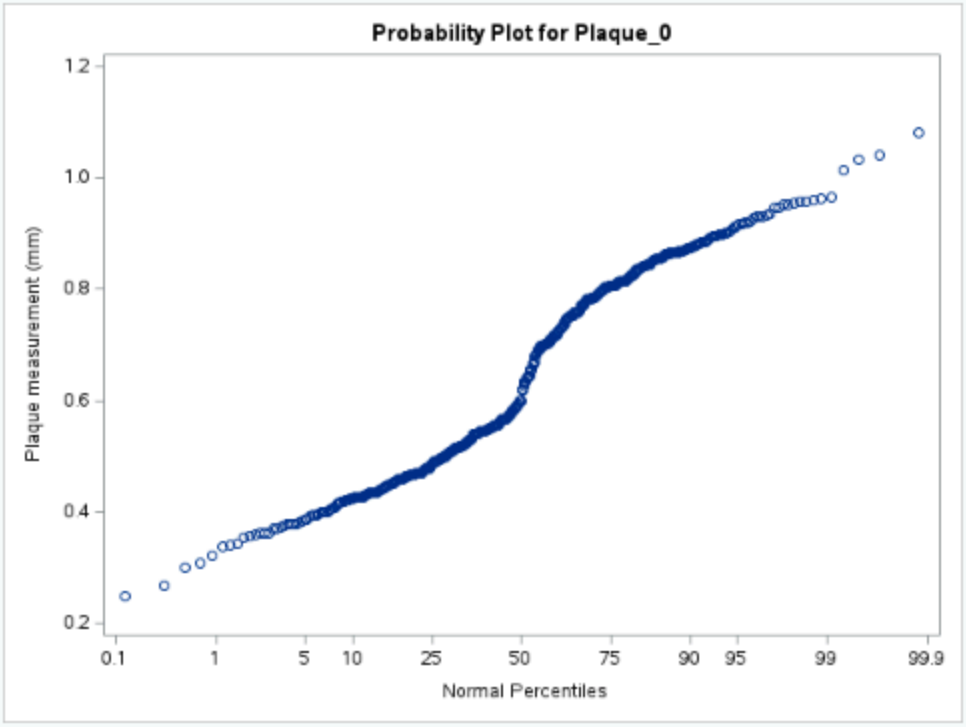
The UNIVARIATE Procedure



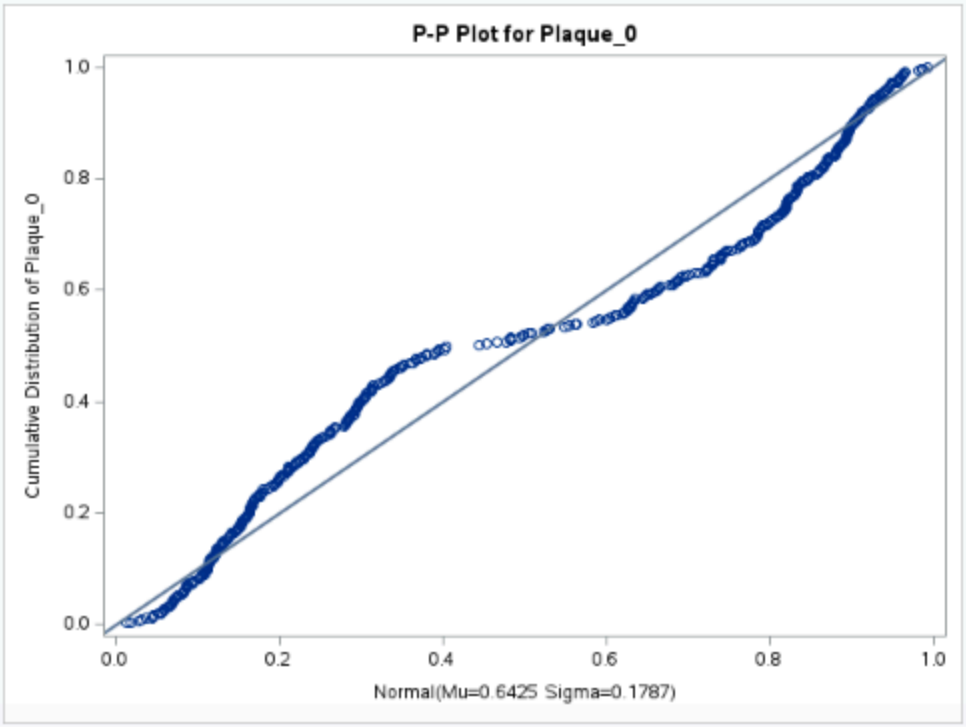
The UNIVARIATE Procedure



The UNIVARIATE Procedure



The UNIVARIATE Procedure



The UNIVARIATE Procedure
Variable: Plaque_2 (Plaque measurement (mm))

Moments			
N	500	Sum Weights	500
Mean	0.6211652	Sum Observations	310.5826
Std Deviation	0.16730151	Variance	0.0279898
Skewness	0.07149137	Kurtosis	-1.0802196
Uncorrected SS	206.890011	Corrected SS	13.9669086
Coeff Variation	26.9334977	Std Error Mean	0.00748195

Basic Statistical Measures			
Location		Variability	
Mean	0.621165	Std Deviation	0.16730
Median	0.600750	Variance	0.02799
Mode	0.352700	Range	0.76950
		Interquartile Range	0.28390

Note: The mode displayed is the smallest of 16 modes with a count of 2.

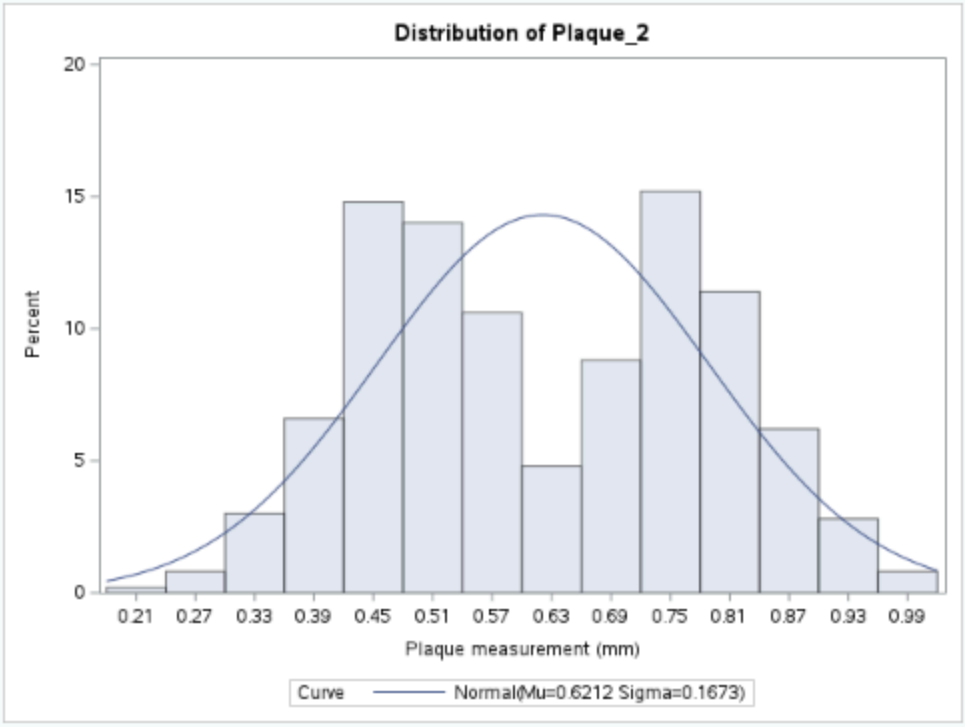
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	83.02182	Pr > |t|	<.0001
Sign	M	250	Pr >= |M|	<.0001
Signed Rank	S	62625	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.96361	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.091152	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.452517	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	7.627263	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	0.99040
99%	0.95950
95%	0.87160
90%	0.83750
75% Q3	0.76280
50% Median	0.60075
25% Q1	0.47890
10%	0.41500
5%	0.38175
1%	0.29705
0% Min	0.22090

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.2209	406	0.9596	105
0.2624	403	0.9694	4
0.2855	174	0.9709	97
0.2872	461	0.9897	47
0.2902	462	0.9904	273

The UNIVARIATE Procedure



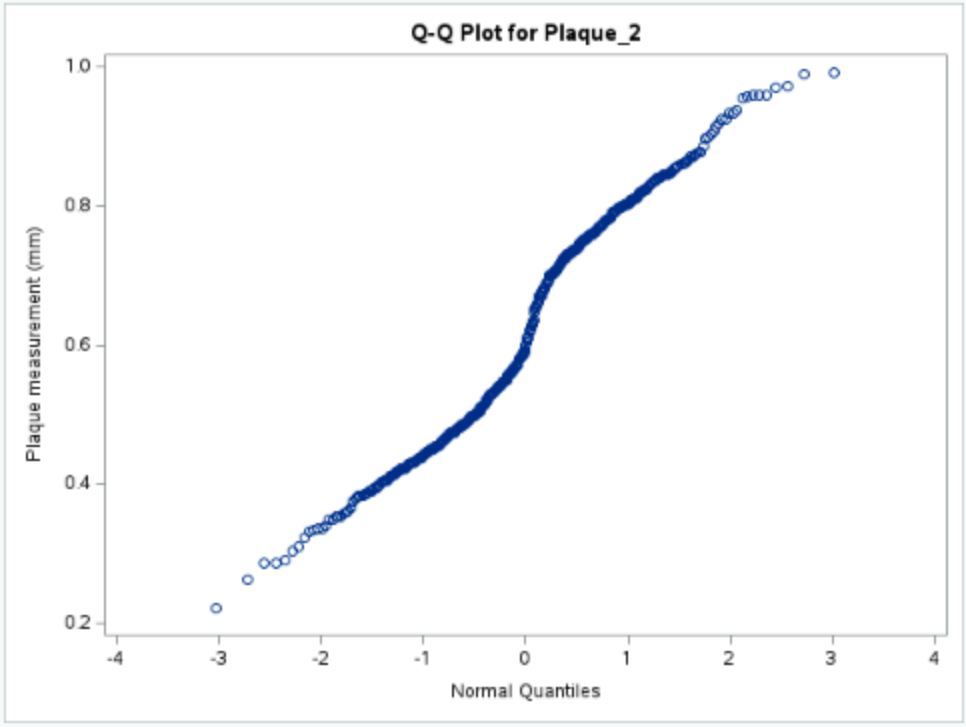
The UNIVARIATE Procedure
Fitted Normal Distribution for Plaque_2 (Plaque measurement (mm))

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	0.621165
Std Dev	Sigma	0.167302

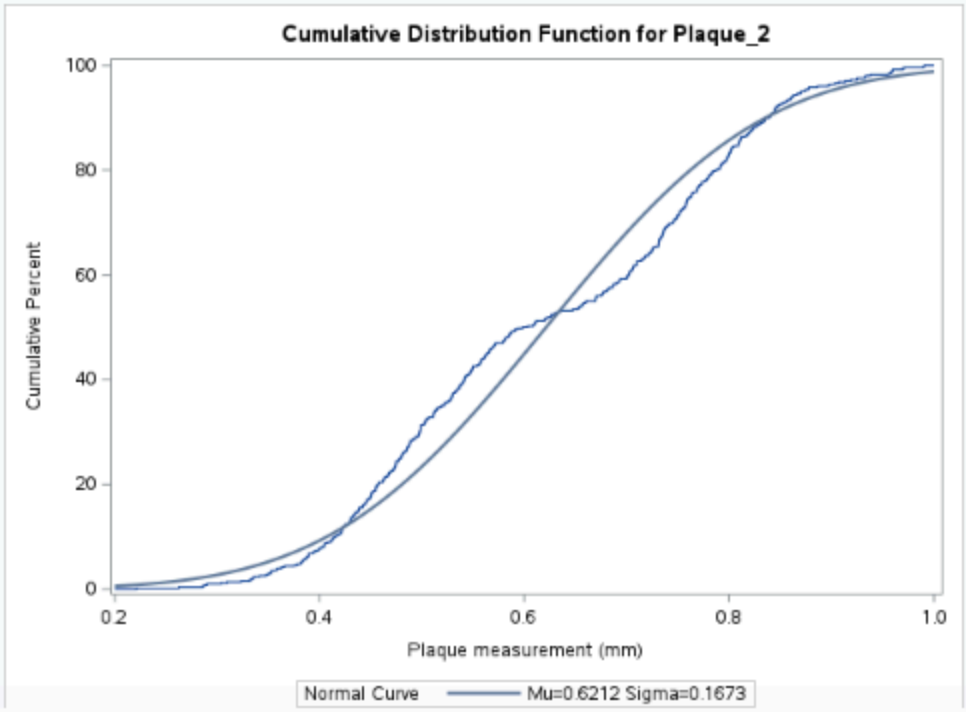
Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.09115240	Pr > D	<0.010
Cramer-von Mises	W-Sq	1.45251704	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	7.62726324	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	0.29705	0.23196
5.0	0.38175	0.34598
10.0	0.41500	0.40676
25.0	0.47890	0.50832
50.0	0.60075	0.62117
75.0	0.76280	0.73401
90.0	0.83750	0.83557
95.0	0.87160	0.89635
99.0	0.95950	1.01037

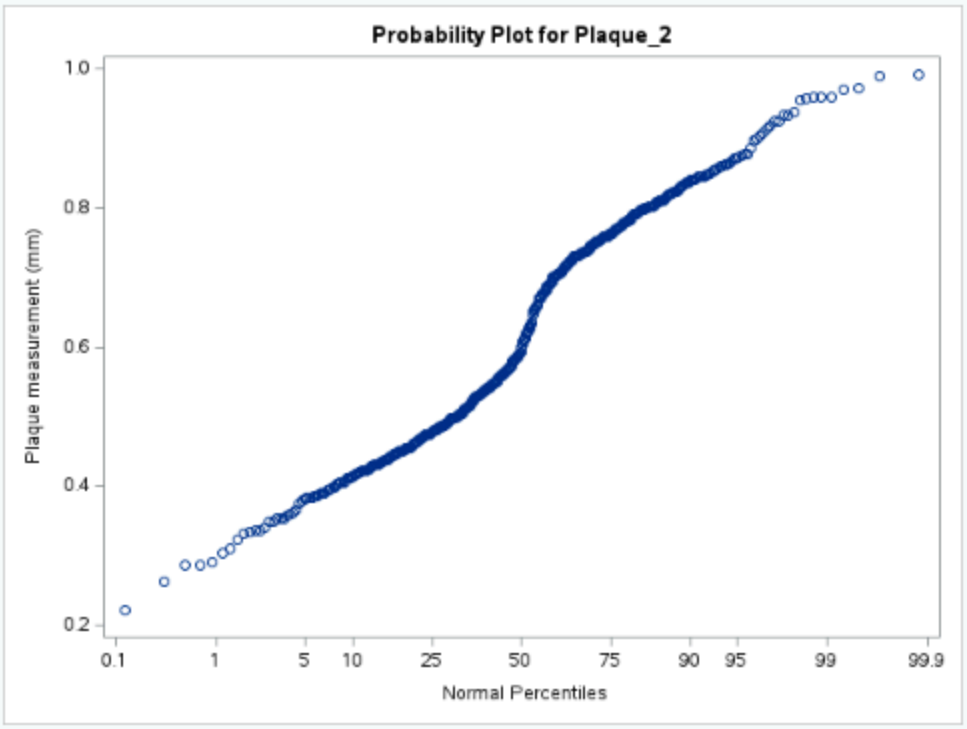
The UNIVARIATE Procedure



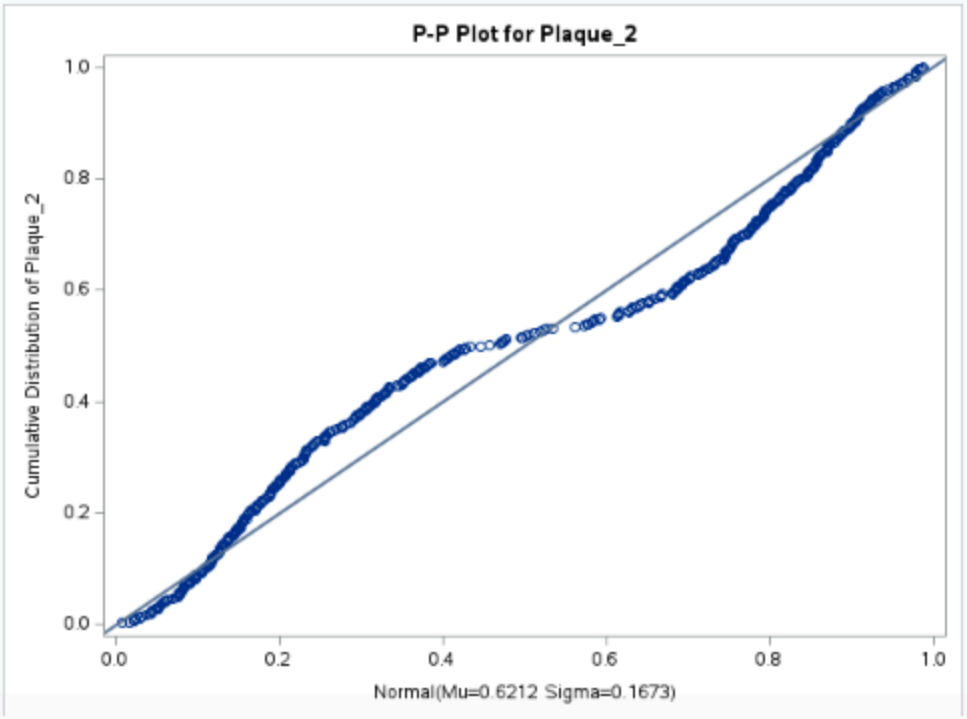
The UNIVARIATE Procedure



The UNIVARIATE Procedure



The UNIVARIATE Procedure



The UNIVARIATE Procedure
Variable: Treatment (0=placebo and 1=vitamin E)

Moments			
N	500	Sum Weights	500
Mean	0.5	Sum Observations	250
Std Deviation	0.50050075	Variance	0.250501
Skewness	0	Kurtosis	-2.0080483
Uncorrected SS	250	Corrected SS	125
Coeff Variation	100.10015	Std Error Mean	0.02238307

Basic Statistical Measures			
Location		Variability	
Mean	0.500000	Std Deviation	0.50050
Median	0.500000	Variance	0.25050
Mode	0.000000	Range	1.00000
		Interquartile Range	1.00000

Note: The mode displayed is the smallest of 2 modes with a count of 250.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	22.33831	Pr > |t|	<.0001
Sign	M	125	Pr >= |M|	<.0001
Signed Rank	S	15687.5	Pr >= |S|	<.0001

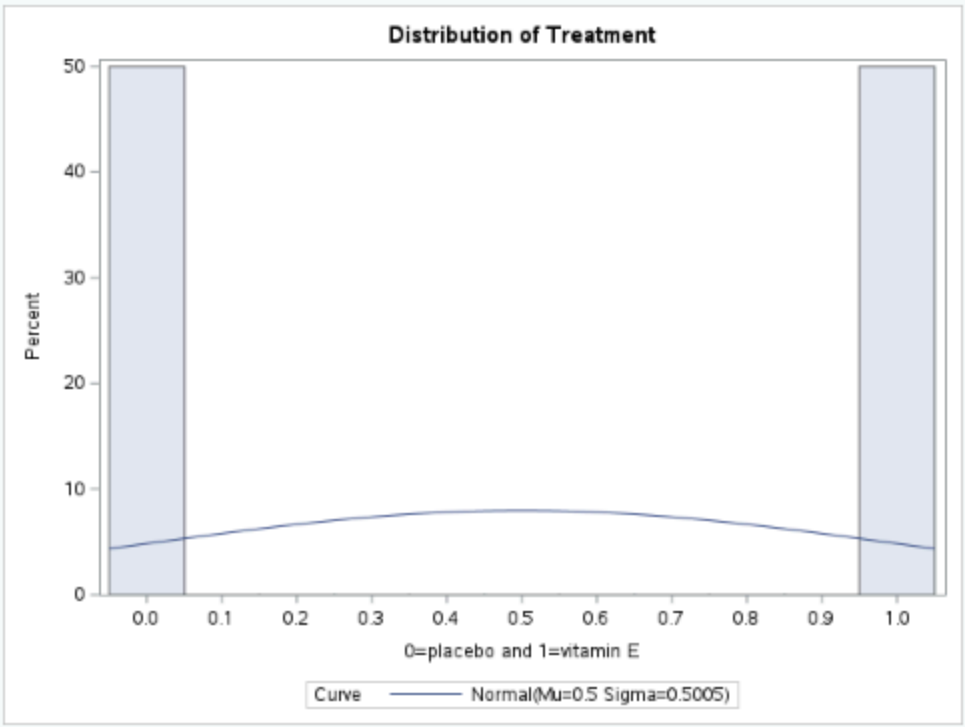
this is the only test you need

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.636543	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.341103	Pr > D	<0.0100
Cramer-von Mises	W-Sq	14.5665	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	89.65532	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	1.0
99%	1.0
95%	1.0
90%	1.0
75% Q3	1.0
50% Median	0.5
25% Q1	0.0
10%	0.0
5%	0.0
1%	0.0
0% Min	0.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	250	1	496
0	249	1	497
0	248	1	498
0	247	1	499
0	246	1	500

The UNIVARIATE Procedure

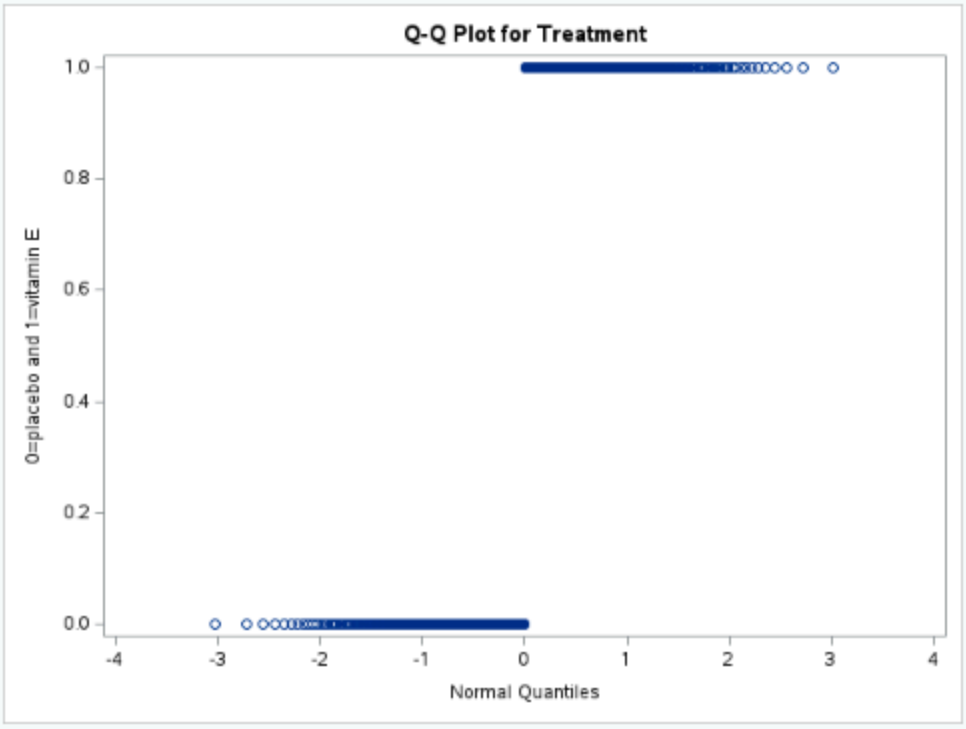


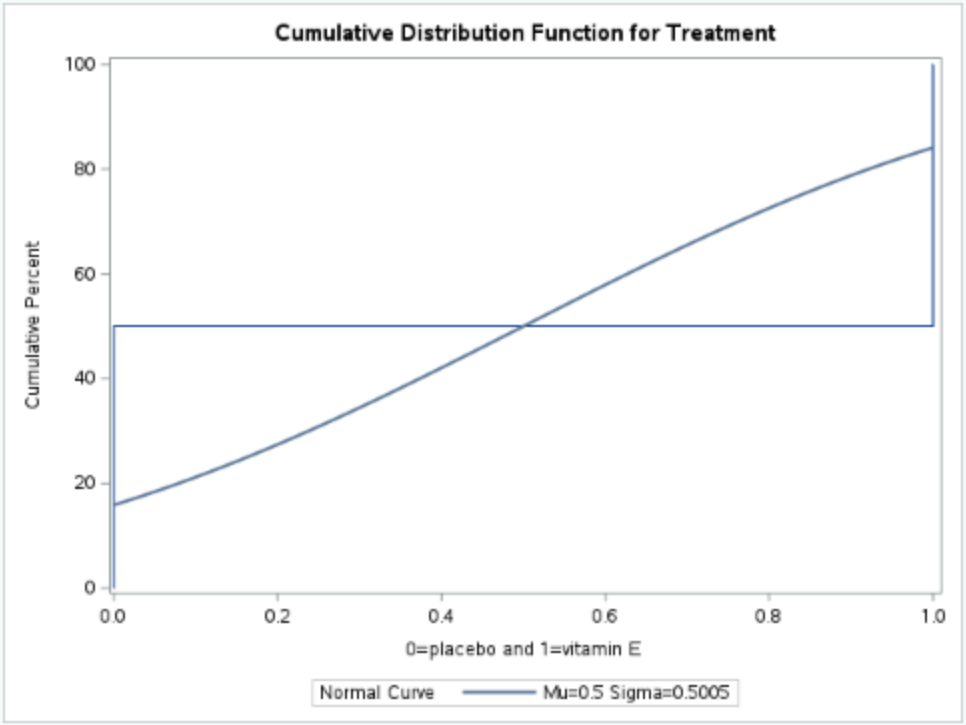
The UNIVARIATE Procedure
Fitted Normal Distribution for Treatment (0=placebo and 1=vitamin E)

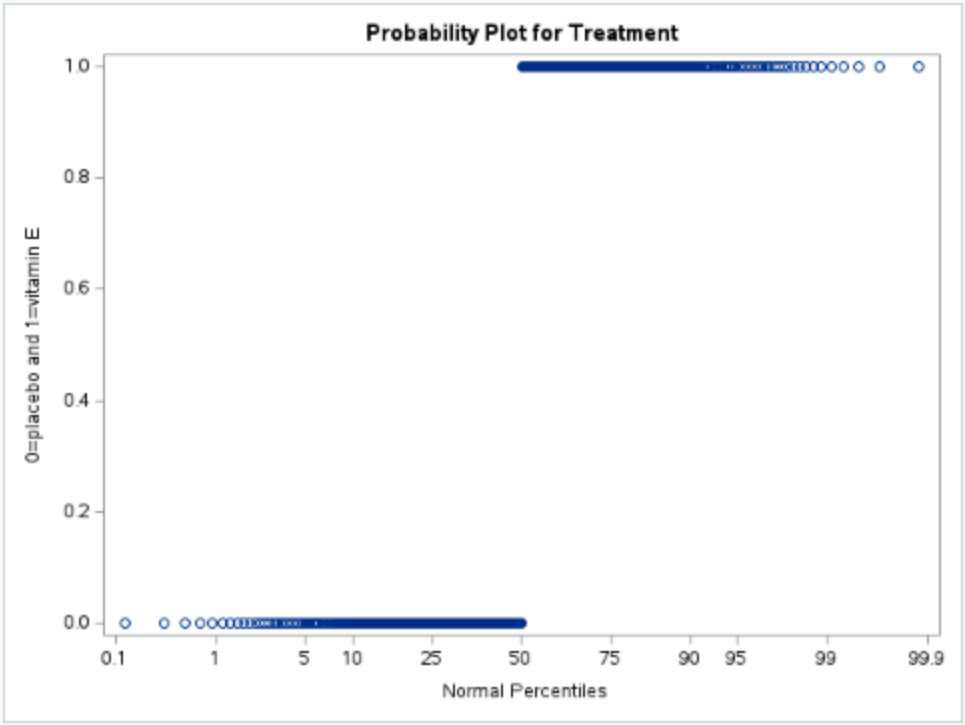
Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	0.5
Std Dev	Sigma	0.500501

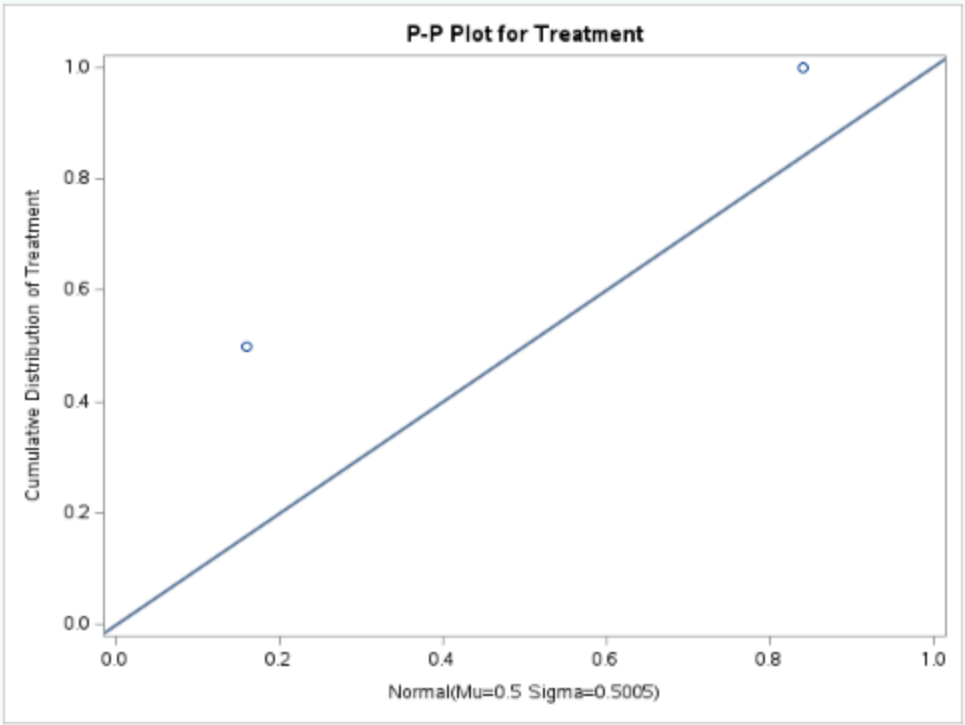
Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.3411025	Pr > D	<0.010
Cramer-von Mises	W-Sq	14.5665024	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	89.6553181	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	0.00000	-0.66434
5.0	0.00000	-0.32325
10.0	0.00000	-0.14142
25.0	0.00000	0.16242
50.0	0.50000	0.50000
75.0	1.00000	0.83758
90.0	1.00000	1.14142
95.0	1.00000	1.32325
99.0	1.00000	1.66434


Q-Q Plot for Treatment
0=placebo and 1=vitamin E
1.0
0.8
0.6
0.4
0.2
0.0
-4
-3
-2
-1
0
1
2
3
4
Normal Quantiles


Cumulative Distribution Function for Treatment
Cumulative Percent
100
80
60
40
20
0
0.0
0.2
0.4
0.6
0.8
1.0
0=placebo and 1=vitamin E
Normal Curve
Mu=0.5 Sigma=0.5005


Probability Plot for Treatment
0=placebo and 1=vitamin E
1.0
0.8
0.6
0.4
0.2
0.0
0.1
1
5
10
25
50
75
90
95
99
99.9
Normal Percentiles


P-P Plot for Treatment
Cumulative Distribution of Treatment
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Normal(Mu=0.5 Sigma=0.5005)

25F

/* 25F */

/* Part c): The results of the T-test show that the p-value is

0.0471 < 0.05, which means that we reject

the null hypothesis (difference of plaque_2 - plaque_0 = 0)

in favor of the alternative hypothesis that

the plaque before and after 2nd visit for both the

treatment and placebo group is significantly different.*/

/* Part d): The results of the T-test display that the p-value is

<0.0001 which is less than 0.05.

This means that we strongly reject the null hypothesis that

there is no difference between the plaque before vs after 2nd visit.

Instead, we favor the alternative hypothesis that the plaque

before and after the 2nd visit across the treatment groups is

significantly different. */

HW Chapter 9 Q28

28. The United States Patent and Trademark Office reports the number of utility patent grants (patents for inventions) per year. The American Community Survey reports selected social characteristics in the United States on geographic areas with a population of 65,000 or more. These data sources have been combined in a SAS data set called PATENTS that represents by county the number of patent grants, population estimate, and various demographic characteristics, for one year.

    a. Examine this SAS data set including the variable labels and attributes. Test to see whether there is a linear relationship between education and number of patents, comparing the resulting p-value to alpha = 0.05. Limit this test to counties with 100 or more patents.
    b. In a comment in your program, discuss the results and any concerns that you might have about the assumptions of normality and constant variance of the residuals for the test from part a). Identify any extreme observations.
    c. Calculate the base 10 log of the variables from part a), and repeat the test using the transformed values.
    d. In a comment in your program, discuss the results and any concerns that you might have about the assumptions of normality and constant variance of the residuals for the test from part c).
    e. Output the data used to create the residual plot for part c) to a data set, and use this to calculate a test of normality for the residuals comparing the resulting p-value to alpha = 0.05.
    f. Include a comment in your program about the results from part e).
    g. Convert your analysis to a simple macro using a parameter to pass in the independent variable. Add titles with the dependent and independent variable names, and name the output data set for the residuals differently for each independent variable. Call the macro twice using independent variables of your choice.
    h. Create a PDF file that contains the results of your entire analysis.

28A

CODE

```
/* 28A */
DATA PATENTS;
  set '/home/u62223361/Intro to SAS/HW9/patents.sas7bdat';
RUN;
```

```
PROC CONTENTS DATA = PATENTS;
RUN;

DATA PATENTS_100;
        SET PATENTS;
        IF Patents => 100;
RUN;

PROC REG DATA = PATENTS_100 ALPHA = 0.05
        CORR EDF ALL SIMPLE RSQUARE PLOTS(ONLY)=(FITPLOT RESIDUALS
DIAGNOSTICS COOKSD OBSERVEDBYPREDICTED QQPLOT
RESIDUALBYPREDICTED RESIDUALHISTOGRAM RFPLOT
RSTUDENTBYLEVERAGE RSTUDENTBYPREDICTED);
        MODEL Patents = Education;
RUN;
```

Those blocked codes are not necessary. The correct code is

```
PROC REG data=PATENTS_100;
Model Patents=education;
run;
```

LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 28A */
70         DATA PATENTS;
71           set '/home/u62223361/Intro to
SAS/HW9/patents.sas7bdat';
NOTE: Data file WC000001.PATENTS.DATA is in a format that is
native to another host, or the file encoding does not match the
      session encoding. Cross Environment Data Access will be
used, which might require additional CPU resources and might
reduce
      performance.
```

```
72          RUN;
NOTE: There were 808 observations read from the data set
/home/u62223361/Intro to SAS/HW9/patents.sas7bdat.
NOTE: The data set WORK.PATENTS has 808 observations and 14
variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1360.06k
      OS Memory           28844.00k
      Timestamp           05/02/2024 04:38:31 AM
      Step Count                        244  Switch Count  3
      Page Faults                       0
      Page Reclaims                     125
      Page Swaps                        0
      Voluntary Context Switches        25
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           520

73
74          PROC CONTENTS DATA = PATENTS;
75          RUN;
NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.02 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              2068.68k
      OS Memory           28844.00k
      Timestamp           05/02/2024 04:38:31 AM
      Step Count                        245  Switch Count  0
```

```
      Page Faults                          0
      Page Reclaims                        95
      Page Swaps                           0
      Voluntary Context Switches           2
      Involuntary Context Switches         0
      Block Input Operations               0
      Block Output Operations              16


76
77          DATA PATENTS_100;
78          SET PATENTS;
79          IF Patents => 100;
80          RUN;
NOTE: There were 808 observations read from the data set
WORK.PATENTS.
NOTE: The data set WORK.PATENTS_100 has 181 observations and 14
variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1149.93k
      OS Memory           28844.00k
      Timestamp           05/02/2024 04:38:31 AM
      Step Count                        246  Switch Count  2
      Page Faults                          0
      Page Reclaims                        118
      Page Swaps                           0
      Voluntary Context Switches           11
      Involuntary Context Switches         0
      Block Input Operations               0
      Block Output Operations              264
```

```
81
82          PROC REG DATA = PATENTS_100 ALPHA = 0.05
83          CORR EDF ALL SIMPLE RSQUARE PLOTS(ONLY)=(FITPLOT
RESIDUALS DIAGNOSTICS COOKSD OBSERVEDBYPREDICTED QQPLOT
83        ! RESIDUALBYPREDICTED RESIDUALHISTOGRAM RFPLOT
RSTUDENTBYLEVERAGE RSTUDENTBYPREDICTED);
84          MODEL Patents = Education;
85          RUN;
86
87          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
97
```

RESULTS

The CONTENTS Procedure

Data Set Name	WORK.PATENTS	**Observations**	808
Member Type	DATA	**Variables**	14
Engine	V9	**Indexes**	0
Created	05/02/2024 00:38:31	**Observation Length**	160
Last Modified	05/02/2024 00:38:31	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	2
First Data Page	1
Max Obs per Page	818
Obs in First Data Page	794
Number of Data Set Repairs	0
Filename	/saswork/SAS_workF26B0000EDBF_odaws02-usw2-2.oda.sas.com/SAS_work7DB30000EDBF_odaws02-usw2-2.oda.sas.com/patents.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	2415921510
Access Permission	rw-r--r--
Owner Name	u62223361
File Size	384KB
File Size (bytes)	393216

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
13	AfricanAmerican	Num	8	African American (%)
6	Age	Num	8	Median Age
14	Asian	Num	8	Asian (%)
11	Caucasian	Num	8	Caucasian (%)
2	County	Char	33	US county name
7	Education	Num	8	Education level of bachelor's degree or more (%)
1	FIPS	Num	8	Federal information processing standard identifier
12	Hispanic	Num	8	Hispanic/Latino (%)
9	Housing	Num	8	Median housing value
8	Income	Num	8	Median household income
4	Patents	Num	8	Number of patents
5	Population	Num	8	Population estimate
3	State	Char	24	State
10	Unemployed	Num	8	Unemployment (%)

The REG Procedure

Number of Observations Read	181
Number of Observations Used	181

Descriptive Statistics

Variable	Sum	Mean	Uncorrected SS	Variance	Standard Deviation	Label
Intercept	181.00000	1.00000	181.00000	0	0	Intercept
Education	6532.50000	36.09116	250583	82.31837	9.07295	Education level of bachelor's degree or more (%)
Patents	89030	491.87845	191217308	819030	905.00267	Number of patents

Uncorrected Sums of Squares and Crossproducts

Variable	Label	Intercept	Education	Patents
Intercept	Intercept	181	6532.5	89030
Education	Education level of bachelor's degree or more (%)	6532.5	250582.81	3497863.8
Patents	Number of patents	89030	3497863.8	191217308

Correlation

Variable	Label	Education	Patents
Education	Education level of bachelor's degree or more (%)	1.0000	0.1926
Patents	Number of patents	0.1926	1.0000

The REG Procedure
Model: MODEL1

Model Crossproducts X'X X'Y Y'Y

Variable	Label	Intercept	Education	Patents
Intercept	Intercept	181	6532.5	89030
Education	Education level of bachelor's degree or more (%)	6532.5	250582.81	3497863.8
Patents	Number of patents	89030	3497863.8	191217308

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Number of Observations Read	181
Number of Observations Used	181

X'X Inverse, Parameter Estimates, and SSE

Variable	Label	Intercept	Education	Patents
Intercept	Intercept	0.09343368	-0.002435744	-201.4993768
Education	Education level of bachelor's degree or more (%)	-0.002435744	0.0000674887	19.211846491
Patents	Number of patents	-201.4993768	19.211846491	141956375.15

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5468994	5468994	6.90	0.0094
Error	179	141956375	793052		
Corrected Total	180	147425369			

Root MSE	890.53488	R-Square	0.0371
Dependent Mean	491.87845	Adj R-Sq	0.0317
Coeff Var	181.04775		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent Standard Error	Heteroscedasticity Consistent t Value	Heteroscedasticity Consistent Pr > |t|	Type I SS	Type II SS	Standardized Estimate	Squared Semi-partial Corr Type I	Squared Partial Corr Type I	Squared Semi-partial Corr Type II	Squared Partial Corr Type II	Tolerance	Variance Inflation	95% Confidence Limits	Heteroscedasticity Consistent 95% Confidence Limits
Intercept	Intercept	1	-201.49938	272.20911	-0.74	0.4601	225.12922	-0.90	0.3720	43791939	434554	0	.	.	.	.	.	0	-738.65110	335.65235	-645.74809	242.74933
Education	Education level of bachelor's degree or more (%)	1	19.21185	7.31588	2.63	0.0094	7.61027	2.52	0.0125	5468994	5468994	0.19261	0.03710	0.03710	0.03710	0.03710	1.00000	1.00000	4.77539	33.64830	4.19447	34.22923

Covariance of Estimates

Variable	Label	Intercept	Education
Intercept	Intercept	74097.801867	-1931.672371
Education	Education level of bachelor's degree or more (%)	-1931.672371	53.522035841

Correlation of Estimates			
Variable	Label	Intercept	Education
Intercept	Intercept	1.0000	-0.9700
Education	Education level of bachelor's degree or more (%)	-0.9700	1.0000

Sequential Parameter Estimates	
Intercept	Education
491.878453	0
-201.499377	19.211846

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Heteroscedasticity Consistent Covariance of Estimates			
Variable	Label	Intercept	Education
Intercept	Intercept	50683.165366	-1687.255573
Education	Education level of bachelor's degree or more (%)	-1687.255573	57.916150836

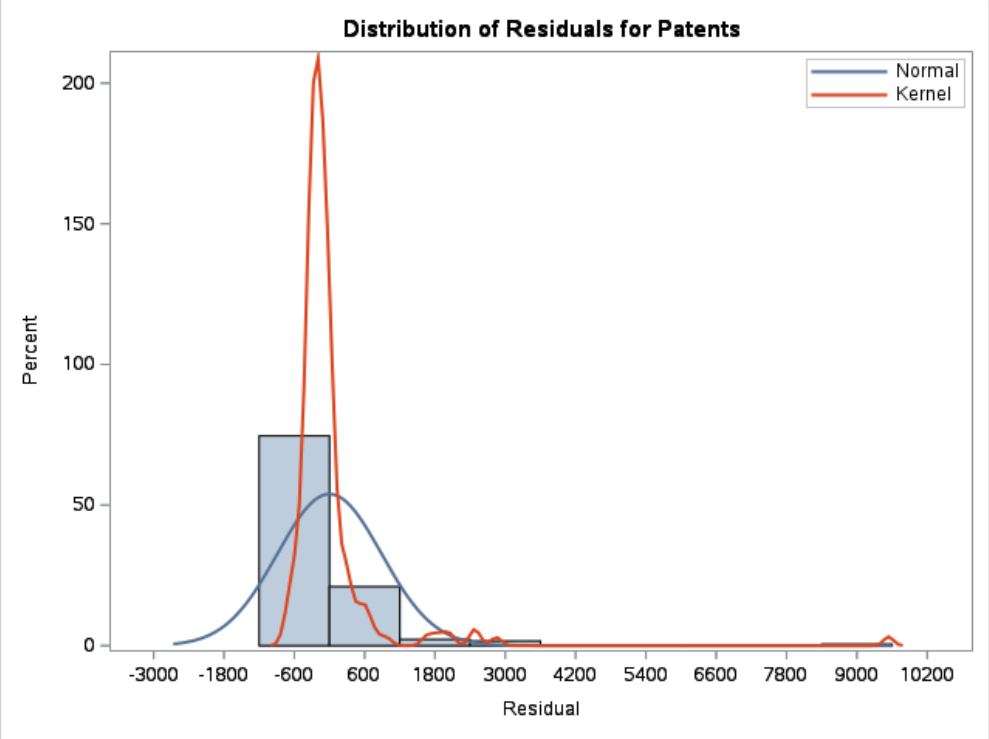
Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
2	1.68	0.4321

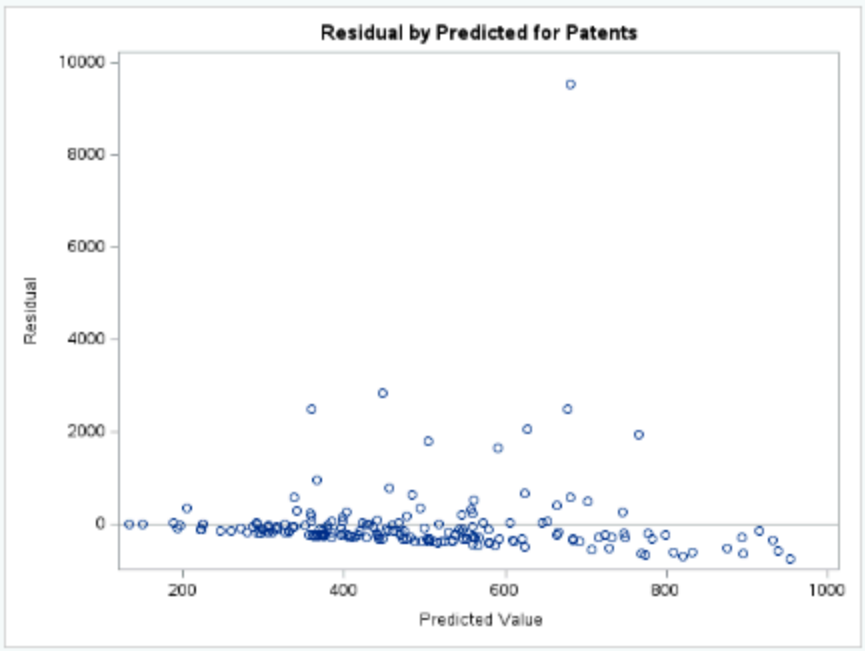
The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

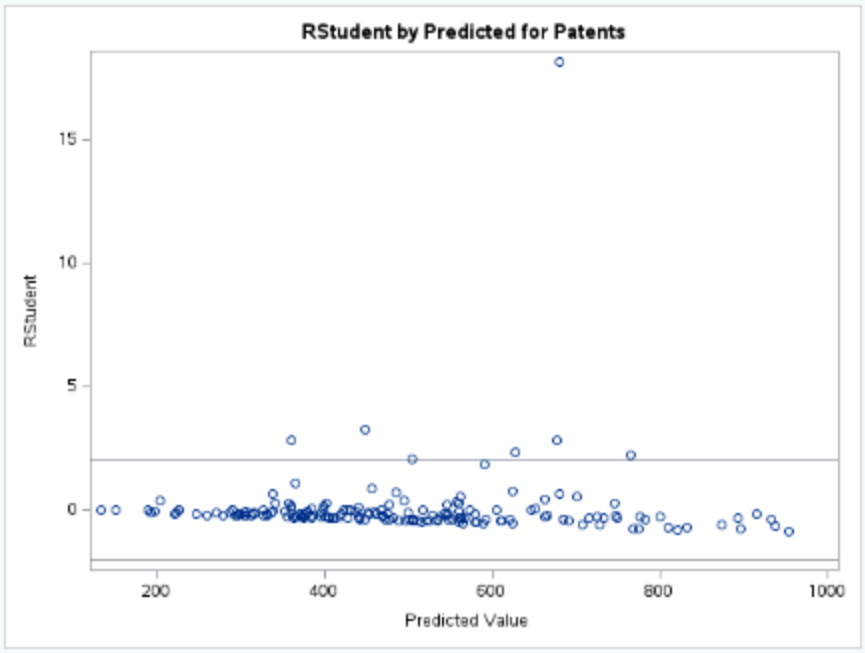
Output Statistics												
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D
1	122	488.2059	66.2077	357.5579	618.8539	-1274	2250	-366.2059	888.1	-0.412	| | |	0.000
2	1334	365.2501	81.8943	203.6475	526.8526	-1399	2130	968.7499	886.8	1.092	| ** |	0.005
3	601	357.5654	83.6509	192.4967	522.6340	-1407	2123	243.4346	886.6	0.275	| | |	0.000
4	2243	590.0287	76.0161	440.0258	740.0316	-1174	2354	1653	887.3	1.863	| *** |	0.013
5	745	545.8415	69.3092	409.0731	682.6098	-1217	2308	199.1585	887.8	0.224	| | |	0.000
6	103	363.3289	82.3272	200.8721	525.7857	-1401	2128	-260.3289	886.7	-0.294	| | |	0.000
7	2844	359.4865	83.2056	195.2966	523.6765	-1405	2124	2485	886.6	2.802	| ***** |	0.035
8	204	809.0437	137.7262	537.2680	1081	-969.1452	2587	-605.0437	879.8	-0.688	| *| |	0.006
9	2310	503.5754	66.3426	372.6611	634.4897	-1259	2266	1806	888.1	2.034	| | **** |	0.012
10	125	503.5754	66.3426	372.6611	634.4897	-1259	2266	-378.5754	888.1	-0.426	| | |	0.001
11	219	188.5011	133.1458	-74.2362	451.2384	-1588	1965	30.4989	880.5	0.035	| | |	0.000
12	143	326.8264	91.2790	146.7050	506.9478	-1440	2093	-183.8264	885.8	-0.208	| | |	0.000
13	157	150.0774	146.0226	-138.0697	438.2245	-1631	1931	6.9226	878.5	0.008	| | |	0.000
14	3293	447.8610	68.2822	313.1193	582.6027	-1315	2210	2845	887.9	3.204	| | ****** |	0.030
15	581	799.4378	134.5299	533.9693	1065	-977.7981	2577	-218.4378	880.3	-0.248	| | |	0.001
16	2695	626.5312	83.7300	461.3063	791.7561	-1139	2392	2068	886.6	2.333	| | **** |	0.024
17	274	419.0433	71.7689	277.4214	560.6652	-1344	2182	-145.0433	887.6	-0.163	| | |	0.000
18	10221	680.3244	97.6270	487.6764	872.9723	-1088	2448	9541	885.2	10.778	| | ****** |	0.707
19	530	517.0237	66.8819	385.0452	649.0021	-1245	2279	12.9763	888.0	0.015	| | |	0.000
20	198	417.1221	72.0548	274.9360	559.3081	-1346	2180	-219.1221	887.6	-0.247	| | |	0.000
21	535	397.9102	75.2458	249.4272	546.3933	-1366	2161	137.0898	887.4	0.154	| | |	0.000
22	100	515.1025	66.7811	383.3230	646.8820	-1247	2277	-415.1025	888.0	-0.467	| | |	0.001
23	155	534.3143	68.1369	399.8594	668.7693	-1228	2297	-379.3143	887.9	-0.427	| | |	0.001
24	595	931.9996	180.1962	576.4174	1288	-860.9122	2725	-336.9996	872.1	-0.386	| | |	0.003
25	130	820.5709	141.5912	541.1682	1100	-958.7997	2600	-690.5709	879.2	-0.785	| *| |	0.008
26	188	482.4424	66.2904	351.6312	613.2536	-1280	2245	-294.4424	888.1	-0.332	| | |	0.000
27	264	559.2897	70.9962	419.1925	699.3870	-1204	2322	-295.2897	887.7	-0.333	| | |	0.000
28	289	620.7677	82.4042	458.1590	783.3763	-1144	2386	-331.7677	886.7	-0.374	| | |	0.001
29	191	271.1120	106.9996	59.9691	482.2550	-1499	2041	-80.1120	884.1	-0.091	| | |	0.000
30	682	645.7431	88.3996	471.3036	820.1825	-1120	2412	36.2569	886.1	0.041	| | |	0.000
31	419	459.3881	67.3393	326.5072	592.2691	-1303	2222	-40.3881	888.0	-0.045	| | |	0.000

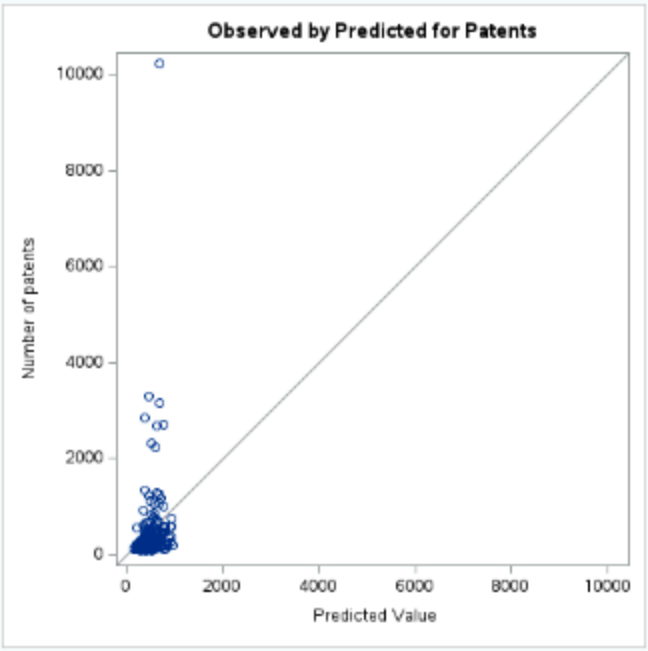
179	104	296.0874	99.7010	99.3469	492.8280	-1472	2064	-192.0874	884.9	-0.217	|	|	|	0.000
180	277	566.9745	72.1060	424.6874	709.2616	-1196	2330	-289.9745	887.6	-0.327	|	|	|	0.000
181	129	303.7722	97.5320	111.3117	496.2326	-1464	2072	-174.7722	885.2	-0.197	|	|	|	0.000

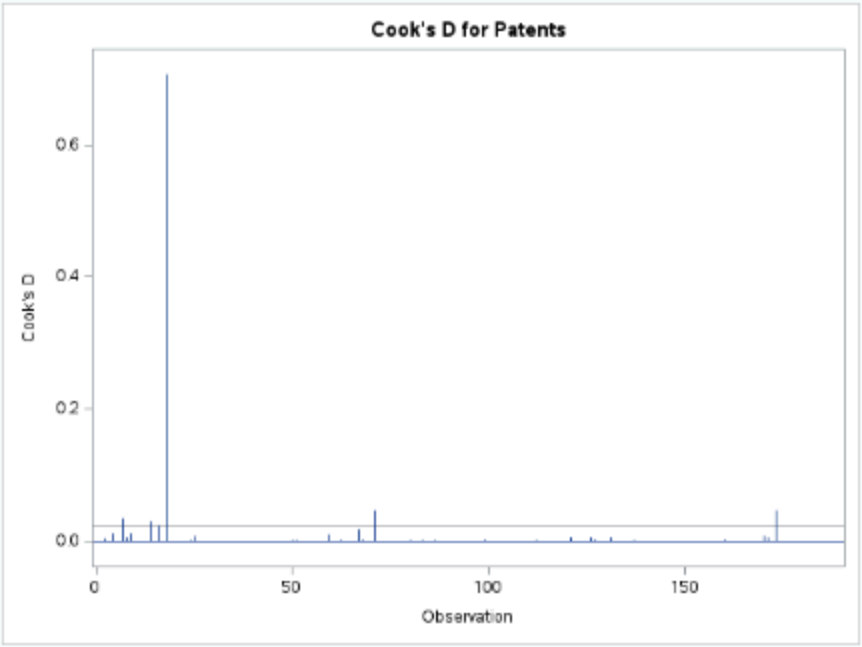
Sum of Residuals	0
Sum of Squared Residuals	141956375
Predicted Residual SS (PRESS)	145307933

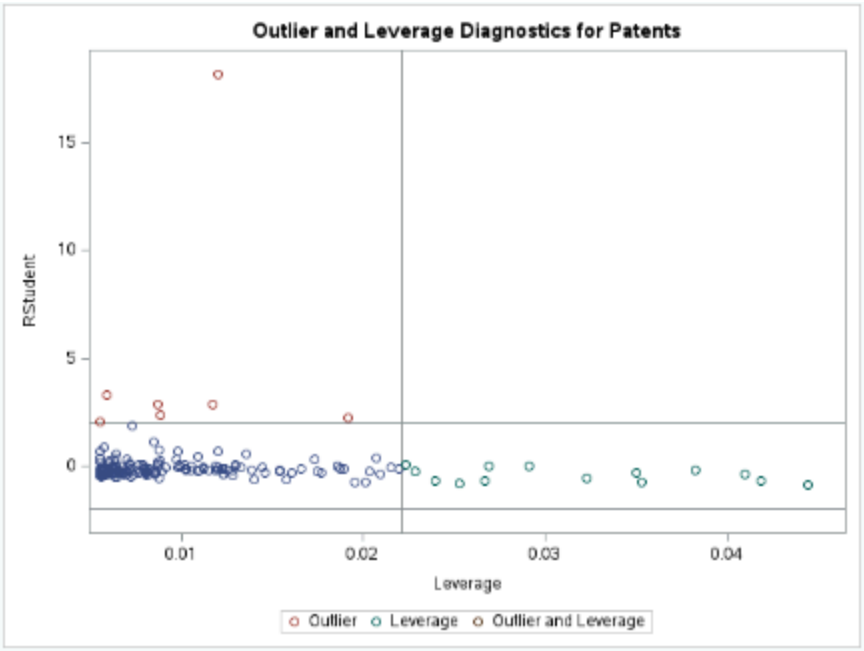


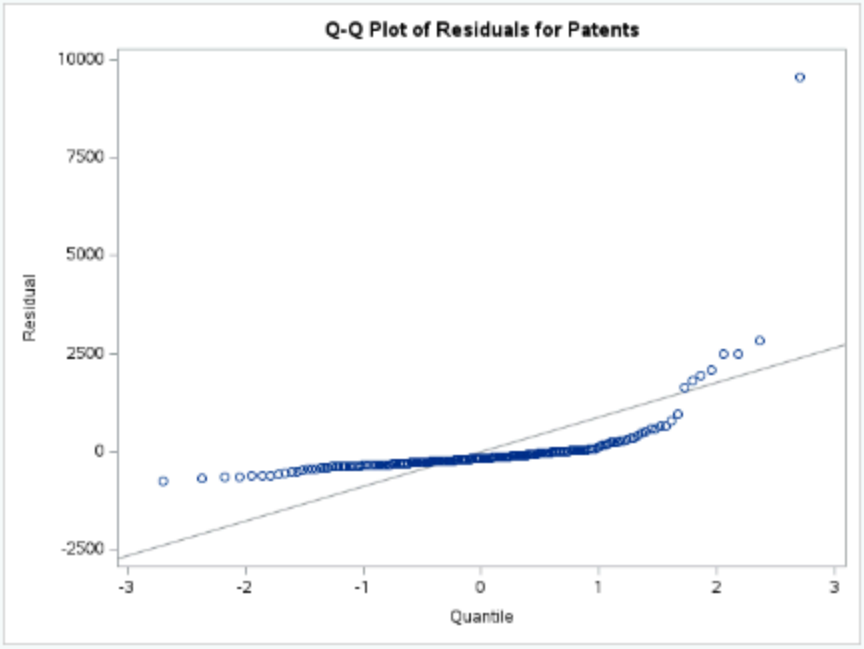

Residual by Predicted for Patents
Residual
10000
8000
6000
4000
2000
0
200
400
600
800
1000
Predicted Value

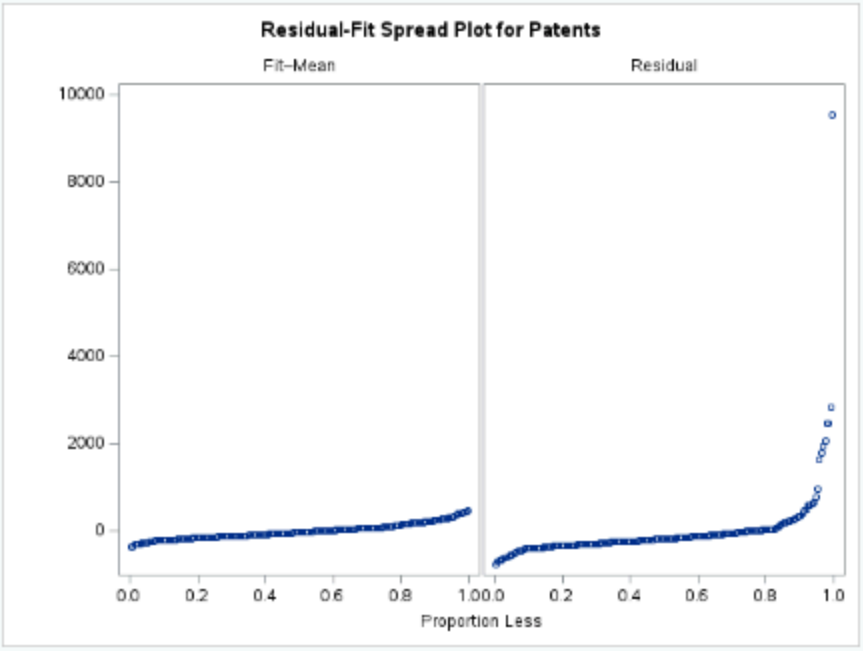

RStudent by Predicted for Patents
RStudent
15
10
5
0
200
400
600
800
1000
Predicted Value

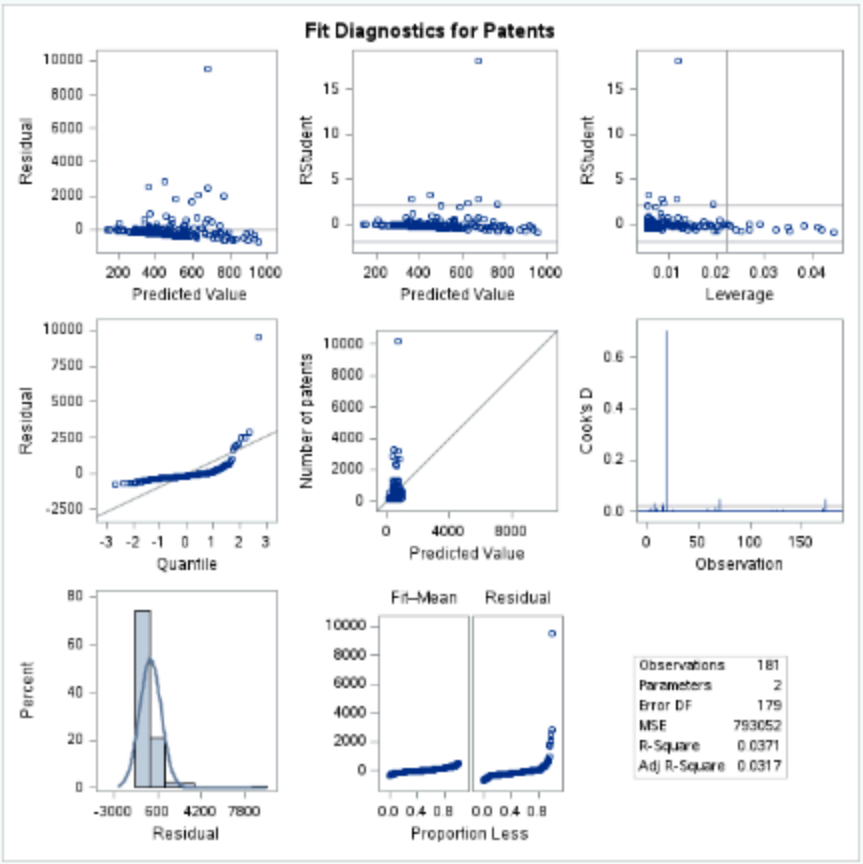

Observed by Predicted for Patents
Number of patents
Predicted Value
0
2000
4000
6000
8000
10000

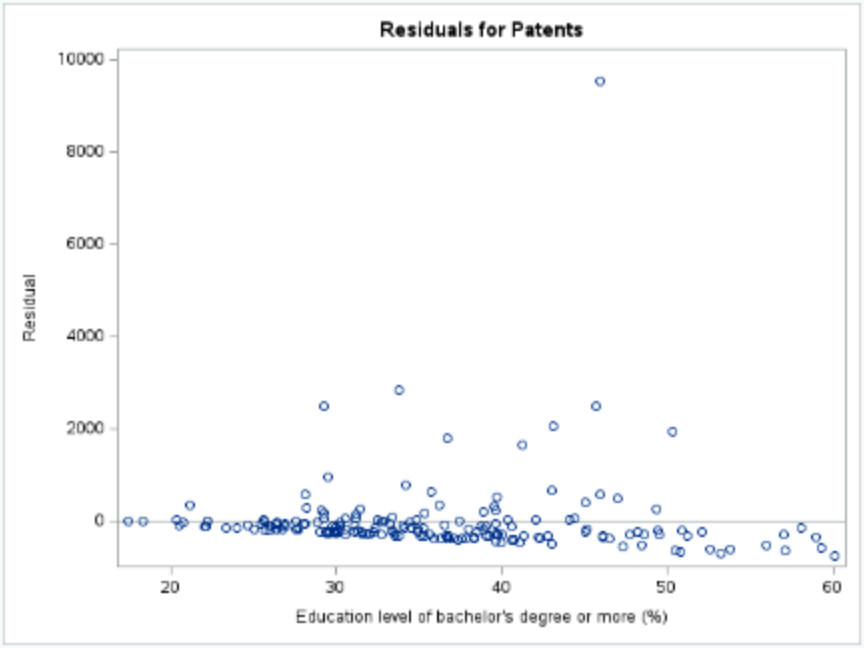

Cook's D for Patents
Cook's D
Observation
0.0
0.2
0.4
0.6
0
50
100
150

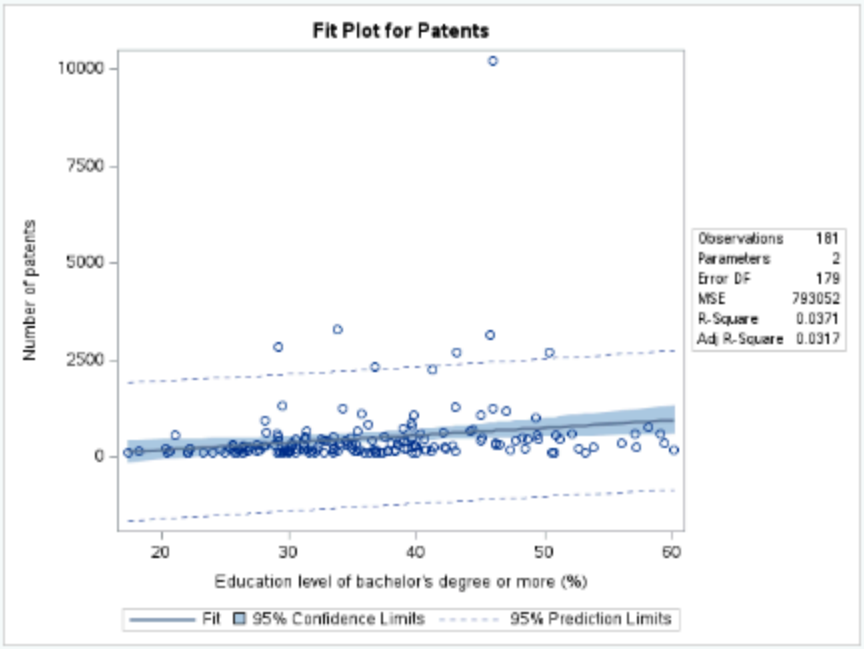

Outlier and Leverage Diagnostics for Patents
RStudent
15
10
5
0
0.01
0.02
0.03
0.04
Leverage
Outlier
Leverage
Outlier and Leverage


Q-Q Plot of Residuals for Patents
Residual
10000
7500
5000
2500
0
-2500
-3
-2
-1
0
1
2
3
Quantile


Residual-Fit Spread Plot for Patents
Fit–Mean
Residual
10000
8000
6000
4000
2000
0
0.0
0.2
0.4
0.6
0.8
1.0
Proportion Less


Fit Diagnostics for Patents
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Number of patents
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 181
Parameters 2
Error DF 179
MSE 793052
R-Square 0.0371
Adj R-Square 0.0317


Residuals for Patents
Residual
10000
8000
6000
4000
2000
0
20
30
40
50
60
Education level of bachelor's degree or more (%)


Fit Plot for Patents
Number of patents
10000
7500
5000
2500
0
20
30
40
50
60
Education level of bachelor's degree or more (%)
Fit
95% Confidence Limits
95% Prediction Limits
Observations 181
Parameters 2
Error DF 179
MSE 793052
R-Square 0.0371
Adj R-Square 0.0317

28B

/* 28B */

/* For 1% increase in %of people with education level of Bachelor's or more,

the number of patents is expected to increase by 19.21185.

The regression coefficient of Education is statistically significant (p-value = 0.0094),

however, the correlation between Education and Patents is only 0.1926, which is low.

The test of first and second moment specification (White test) has a Chi-square p-value of

0.4321, which is > 0.05, indicating that we fail to reject the null hypothesis of homoscedasticity

(we do not have enough evidence to conclude that the error terms' variance are not constant).

In other words, the test suggests constant variance.

However, I still have concerns over the constant variance assumption because

1) Both Residual vs Predicted & Observed vs Predicted plot does not showcase uniformly spreaded residual points

2) Residual for Patents also does not display uniformly spreaded residual points

The adjusted R-square is 0.0317, which means the model does not really fully explain the change in patents.

There are also some outlier and high leverage points.

The QQ-plot shows that the errors are not exactly normally distributed, which may be of concern.

Based on the ANOVA table, the p-value for the F-test is 0.0094 which means that

we can reject the null hypothesis (no statistically significant relationship between independent variable and dependent variable)

in favor of the alternative hypothesis (there is statistically significant relationship between Patents and Education).

- It seems taht you have explained more as needed

Since there is only 1 variable, it makes sense that it only has VIF = 1 (no multicollinearity).

*/

28C

CODE

```
/* 28C */
DATA PATENTS_LOG10;
        SET PATENTS_100;
        IF Patents => 100;
        Patents = LOG10(Patents);
        Education = LOG10(Education);
RUN;

PROC REG DATA = PATENTS_LOG10 ALPHA = 0.05
        CORR EDF ALL SIMPLE RSQUARE PLOTS(ONLY)=(FITPLOT RESIDUALS
DIAGNOSTICS COOKSD OBSERVEDBYPREDICTED QQPLOT
RESIDUALBYPREDICTED RESIDUALHISTOGRAM RFPLOT
RSTUDENTBYLEVERAGE RSTUDENTBYPREDICTED);
        MODEL Patents = Education;
RUN;
```

as good practice, consider use difference variable name for log transformation. for an example, logpatents=log(patients);

LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 28C */
70         DATA PATENTS_LOG10;
71         SET PATENTS_100;
72         IF Patents => 100;
73         Patents = LOG10(Patents);
74         Education = LOG10(Education);
```

```
75          RUN;
NOTE: There were 181 observations read from the data set
WORK.PATENTS_100.
NOTE: The data set WORK.PATENTS_LOG10 has 181 observations and
14 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1010.34k
      OS Memory           29356.00k
      Timestamp           05/02/2024 04:41:17 AM
      Step Count                        253  Switch Count  2
      Page Faults                       0
      Page Reclaims                     120
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264

76
77          PROC REG DATA = PATENTS_LOG10 ALPHA = 0.05
78          CORR EDF ALL SIMPLE RSQUARE PLOTS(ONLY)=(FITPLOT
RESIDUALS DIAGNOSTICS COOKSD OBSERVEDBYPREDICTED QQPLOT
78        ! RESIDUALBYPREDICTED RESIDUALHISTOGRAM RFPLOT
RSTUDENTBYLEVERAGE RSTUDENTBYPREDICTED);
79          MODEL Patents = Education;
80          RUN;
81
82          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
92
```

RESULTS

The REG Procedure

Number of Observations Read	181
Number of Observations Used	181

Descriptive Statistics

Variable	Sum	Mean	Uncorrected SS	Variance	Standard Deviation	Label
Intercept	181.00000	1.00000	181.00000	0	0	Intercept
Education	279.43237	1.54383	433.54901	0.01197	0.10940	Education level of bachelor's degree or more (%)
Patents	448.88202	2.48001	1137.06704	0.13241	0.36389	Number of patents

Uncorrected Sums of Squares and Crossproducts

Variable	Label	Intercept	Education	Patents
Intercept	Intercept	181	279.43236576	448.88201978
Education	Education level of bachelor's degree or more (%)	279.43236576	433.54900732	695.50840577
Patents	Number of patents	448.88201978	695.50840577	1137.0670398

Correlation

Variable	Label	Education	Patents
Education	Education level of bachelor's degree or more (%)	1.0000	0.3507
Patents	Number of patents	0.3507	1.0000

The REG Procedure
Model: MODEL1

Model Crossproducts X'X X'Y Y'Y

Variable	Label	Intercept	Education	Patents
Intercept	Intercept	181	279.43236576	448.88201978
Education	Education level of bachelor's degree or more (%)	279.43236576	433.54900732	695.50840577
Patents	Number of patents	448.88201978	695.50840577	1137.0670398

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Number of Observations Read	181
Number of Observations Used	181

X'X Inverse, Parameter Estimates, and SSE

Variable	Label	Intercept	Education	Patents
Intercept	Intercept	1.1118828229	-0.716634204	0.679094722
Education	Education level of bachelor's degree or more (%)	-0.716634204	0.4641938686	1.1665287026
Patents	Number of patents	0.679094722	1.1665287026	20.903111112

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.93151	2.93151	25.10	<.0001
Error	179	20.90311	0.11678		
Corrected Total	180	23.83462			

Root MSE	0.34173	R-Square	0.1230
Dependent Mean	2.48001	Adj R-Sq	0.1181
Coeff Var	13.77924		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent Standard Error	Heteroscedasticity Consistent t Value	Heteroscedasticity Consistent Pr > |t|	Type I SS	Type II SS	Standardized Estimate	Squared Semi-partial Corr Type I	Squared Partial Corr Type I	Squared Semi-partial Corr Type II	Squared Partial Corr Type II	Tolerance	Variance Inflation	95% Confidence Limits		Heteroscedasticity Consistent 95% Confidence Limits	
Intercept	Intercept	1	0.67909	0.36034	1.88	0.0611	0.31858	2.13	0.0344	1113.23242	0.41476	0	.	.	.	.	.	0	-0.03196	1.39015	0.05043	1.30776
Education	Education level of bachelor's degree or more (%)	1	1.16653	0.23282	5.01	<.0001	0.21099	5.53	<.0001	2.93151	2.93151	0.35070	0.12299	0.12299	0.12299	0.12299	1.00000	1.00000	0.70709	1.62596	0.75018	1.58287

Covariance of Estimates

Variable	Label	Intercept	Education
Intercept	Intercept	0.129842515	-0.083686505
Education	Education level of bachelor's degree or more (%)	-0.083686505	0.0542072403

Correlation of Estimates

Variable	Label	Intercept	Education
Intercept	Intercept	1.0000	-0.9975
Education	Education level of bachelor's degree or more (%)	-0.9975	1.0000

Sequential Parameter Estimates

Intercept	Education
2.480011	0
0.679095	1.166529

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Heteroscedasticity Consistent Covariance of Estimates			
Variable	Label	Intercept	Education
Intercept	Intercept	0.1014950267	-0.067027355
Education	Education level of bachelor's degree or more (%)	-0.067027355	0.0445163336

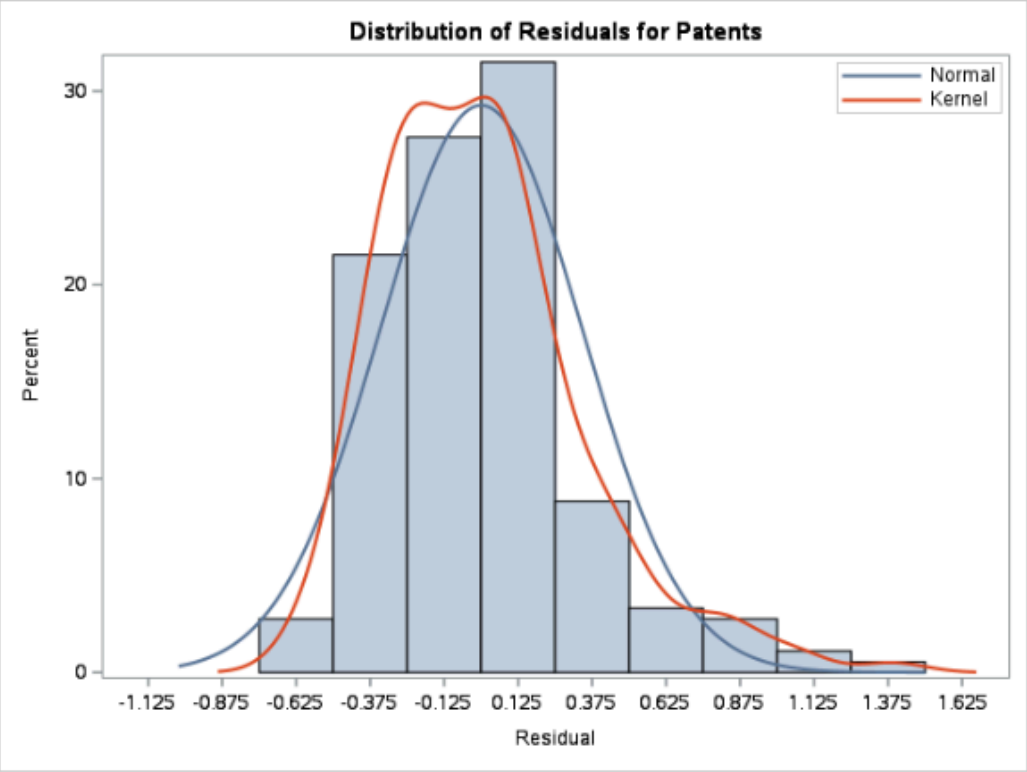
Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
2	8.53	0.0140

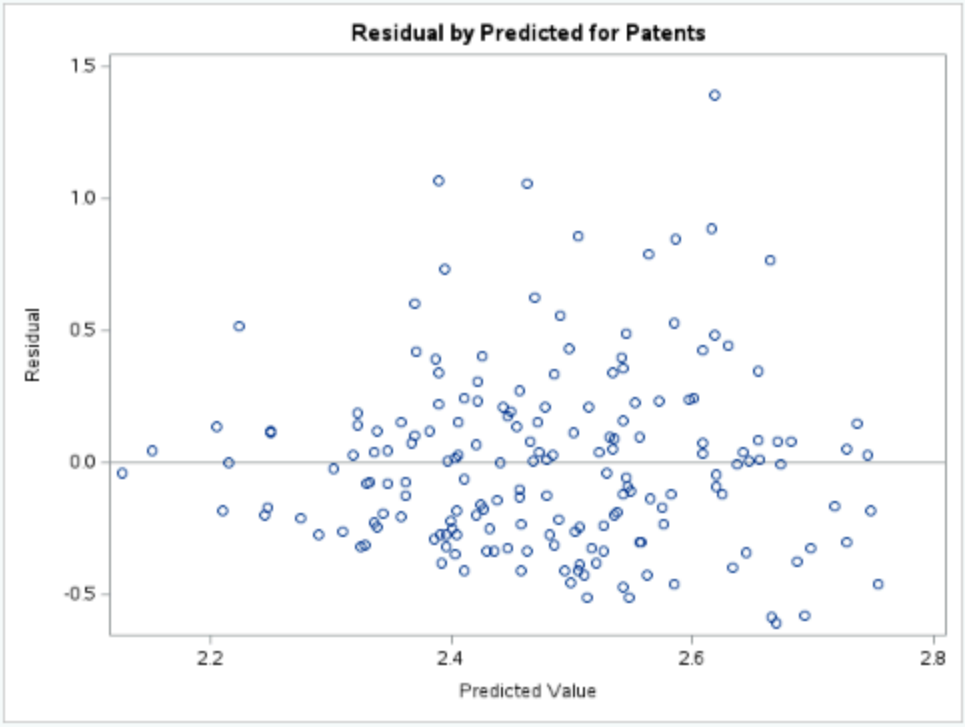
The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

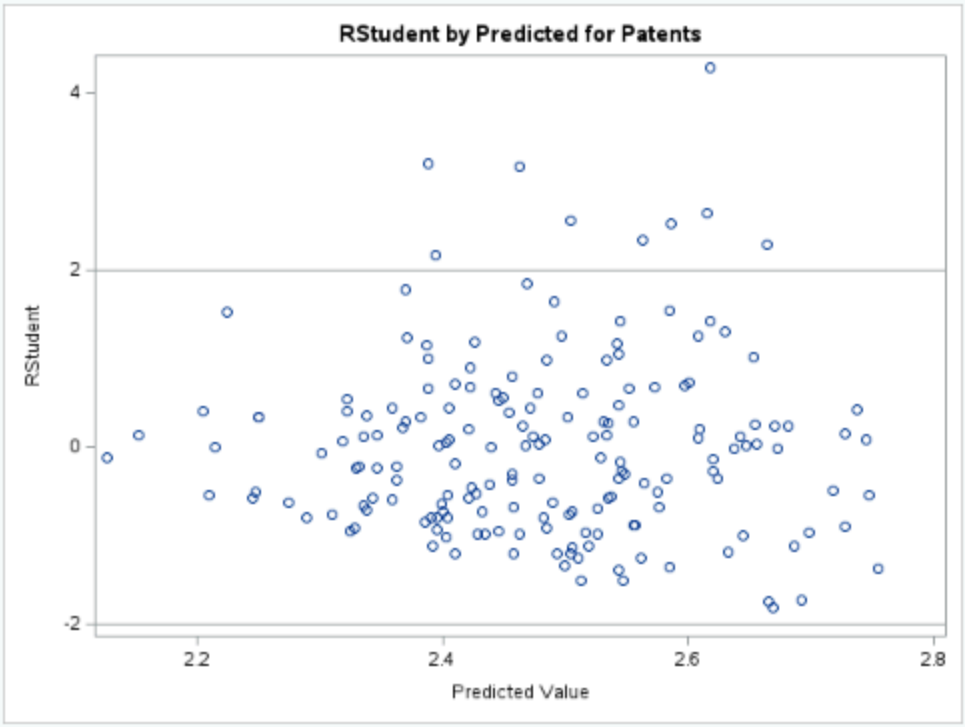
Output Statistics											
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D
1	2.09	2.4932	0.0255	2.4428	2.5435	1.8169	3.1694	-0.4068	0.341	-1.194	| **| |	0.004
2	3.13	2.3937	0.0307	2.3331	2.4543	1.7166	3.0707	0.7315	0.340	2.149	| |**** |	0.019
3	2.78	2.3868	0.0315	2.3246	2.4489	1.7096	3.0640	0.3921	0.340	1.152	| |** |	0.006
4	3.35	2.5629	0.0303	2.5031	2.6227	1.8859	3.2399	0.7879	0.340	2.315	| |**** |	0.021
5	2.87	2.5338	0.0276	2.4794	2.5882	1.8573	3.2103	0.3383	0.341	0.993	| |* |	0.003
6	2.01	2.3920	0.0309	2.3310	2.4529	1.7149	3.0690	-0.3791	0.340	-1.114	| **| |	0.005
7	3.45	2.3885	0.0313	2.3268	2.4502	1.7114	3.0657	1.0654	0.340	3.131	| |******|	0.041
8	2.31	2.6867	0.0484	2.5911	2.7823	2.0056	3.3677	-0.3770	0.338	-1.115	| **| |	0.013
9	3.36	2.5043	0.0259	2.4533	2.5554	1.8281	3.1806	0.8593	0.341	2.522	| |***** |	0.018
10	2.10	2.5043	0.0259	2.4533	2.5554	1.8281	3.1806	-0.4074	0.341	-1.196	| **| |	0.004
11	2.34	2.2043	0.0606	2.0847	2.3239	1.5195	2.8892	0.1361	0.336	0.405	| | |	0.003
12	2.16	2.3581	0.0352	2.2887	2.4275	1.6802	3.0360	-0.2028	0.340	-0.597	| *| |	0.002
13	2.20	2.1518	0.0703	2.0131	2.2904	1.4633	2.8402	0.0441	0.334	0.132	| | |	0.000
14	3.52	2.4626	0.0256	2.4120	2.5132	1.7864	3.1388	1.0550	0.341	3.096	| |******|	0.027
15	2.76	2.6818	0.0476	2.5879	2.7758	2.0010	3.3627	0.0823	0.338	0.243	| | |	0.001
16	3.43	2.5858	0.0330	2.5206	2.6509	1.9083	3.2632	0.8448	0.340	2.484	| |**** |	0.029
17	2.44	2.4396	0.0266	2.3870	2.4922	1.7632	3.1160	-0.001872	0.341	-0.005	| | |	0.000
18	4.01	2.6176	0.0374	2.5438	2.6915	1.9393	3.2960	1.3918	0.340	4.098	| |******|	0.102
19	2.72	2.5139	0.0263	2.4620	2.5658	1.8376	3.1902	0.2104	0.341	0.617	| |* |	0.001
20	2.30	2.4381	0.0267	2.3853	2.4908	1.7617	3.1144	-0.1414	0.341	-0.415	| | |	0.001
21	2.73	2.4221	0.0279	2.3670	2.4771	1.7455	3.0986	0.3063	0.341	0.899	| |* |	0.003
22	2.00	2.5125	0.0262	2.4608	2.5643	1.8362	3.1889	-0.5125	0.341	-1.504	| ***| |	0.007
23	2.19	2.5259	0.0270	2.4727	2.5792	1.8495	3.2024	-0.3356	0.341	-0.985	| *| |	0.003

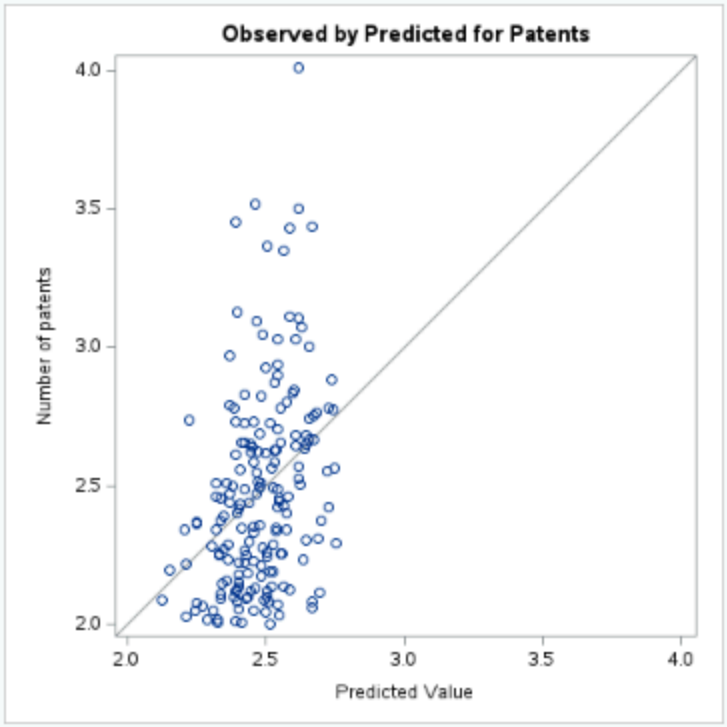
179	2.02	2.3277	0.0396	2.2496	2.4059	1.6489	3.0066	-0.3107	0.339	-0.915	|	* |	|	0.006
180	2.44	2.5479	0.0288	2.4911	2.6048	1.8712	3.2247	-0.1055	0.341	-0.310	|	|	|	0.000
181	2.11	2.3355	0.0384	2.2597	2.4113	1.6569	3.0141	-0.2249	0.340	-0.662	|	* |	|	0.003

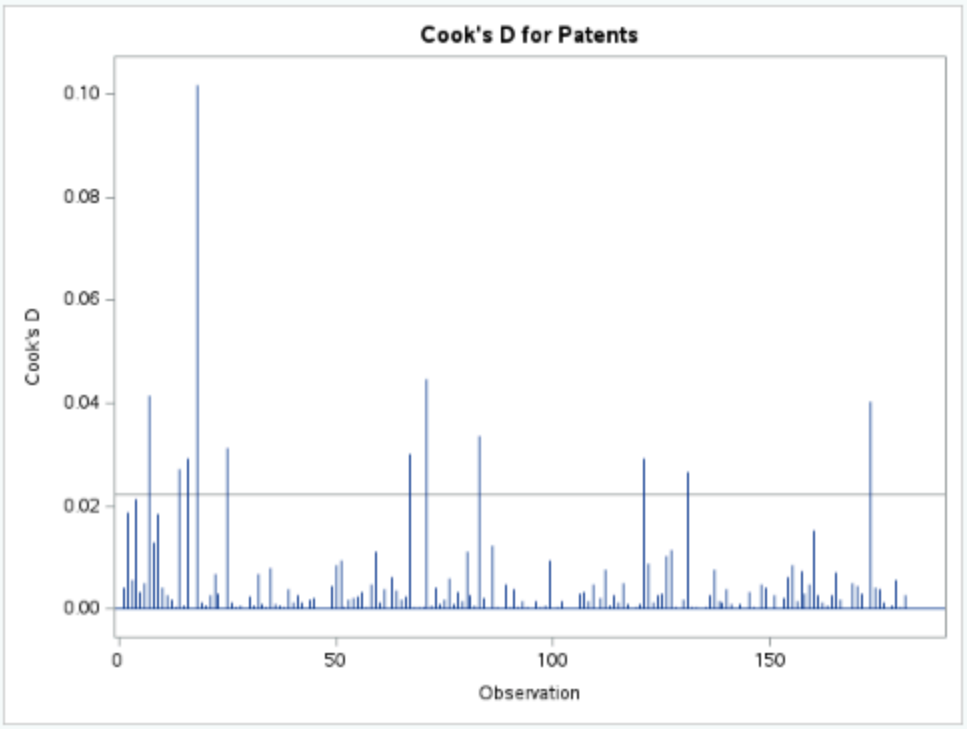
Sum of Residuals	0
Sum of Squared Residuals	20.90311
Predicted Residual SS (PRESS)	21.33428

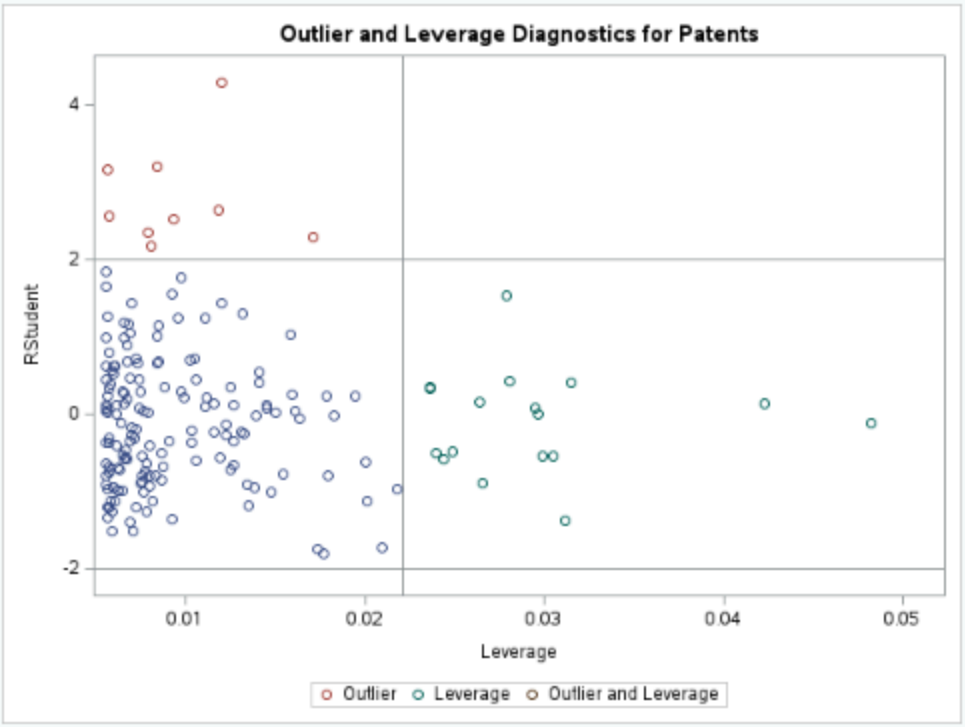


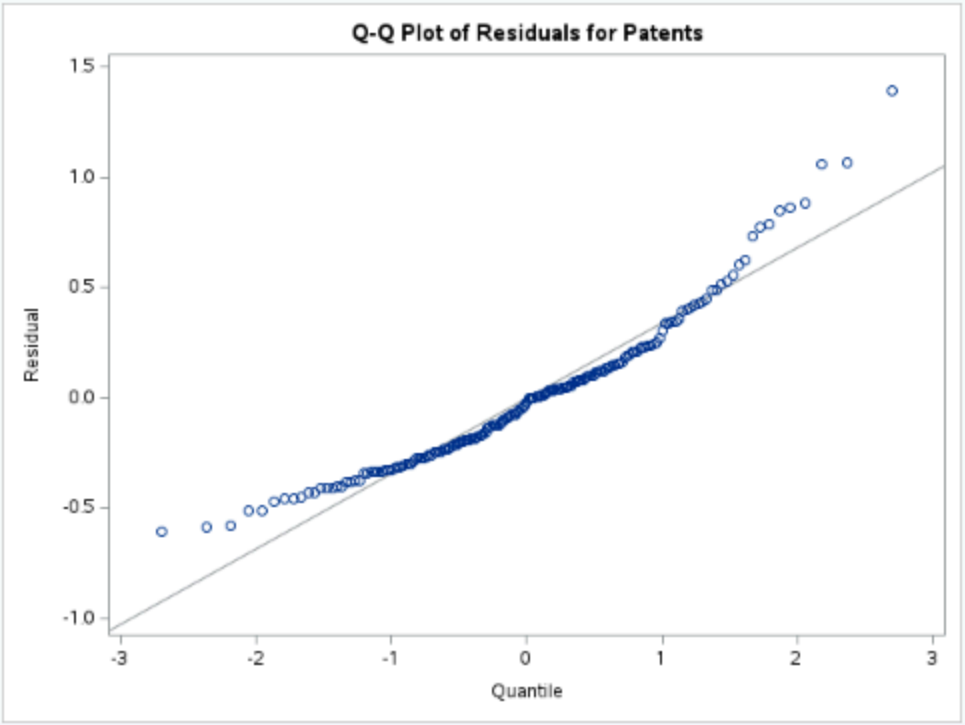

Residual by Predicted for Patents
Residual
1.5
1.0
0.5
0.0
-0.5
2.2
2.4
2.6
2.8
Predicted Value

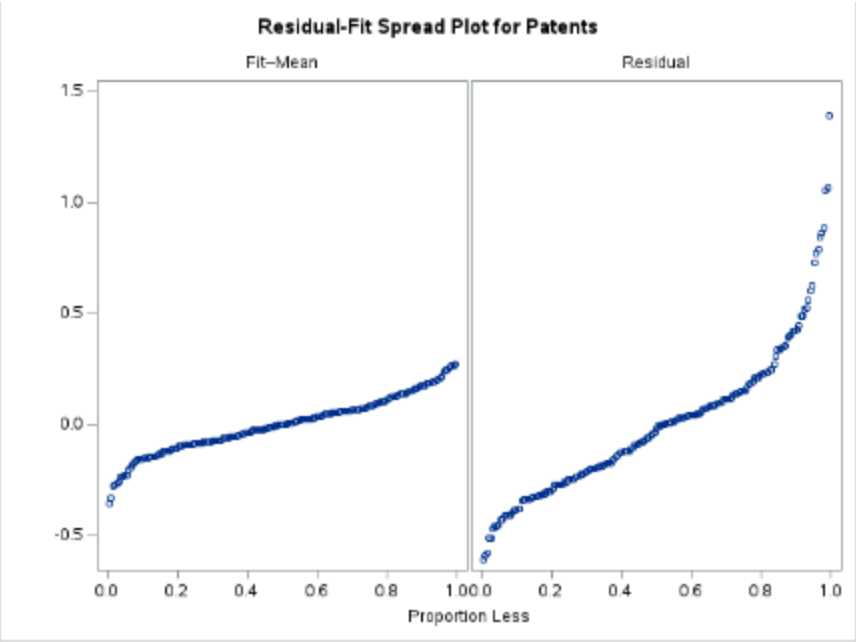

RStudent by Predicted for Patents
RStudent
4
2
0
-2
2.2
2.4
2.6
2.8
Predicted Value

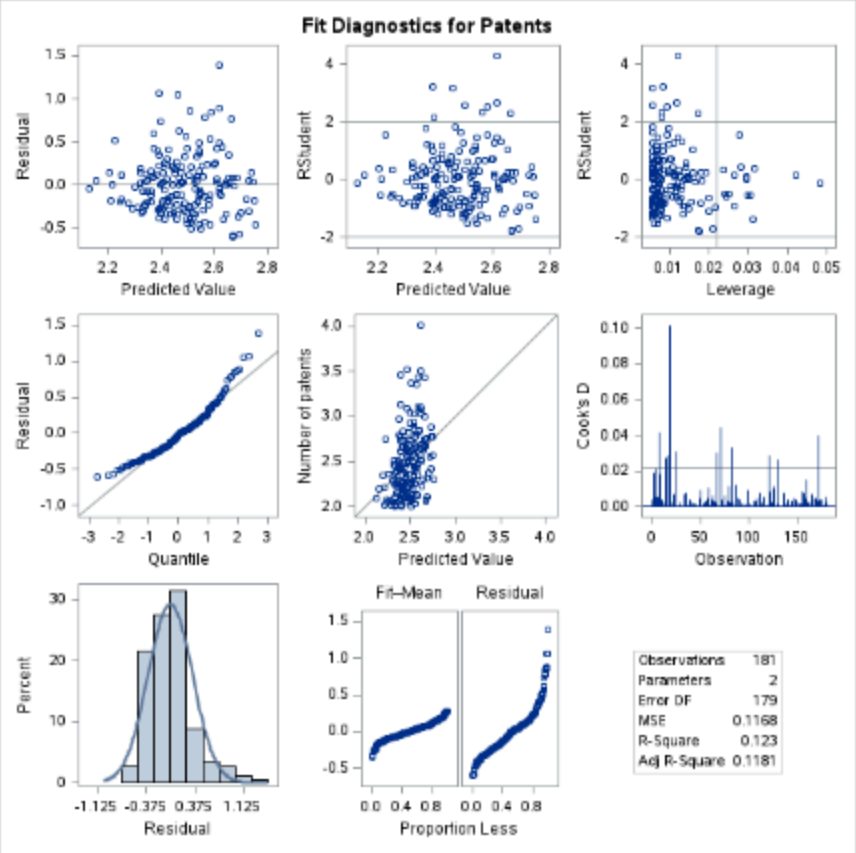

Observed by Predicted for Patents
Number of patents
Predicted Value
4.0
3.5
3.0
2.5
2.0
2.0
2.5
3.0
3.5
4.0

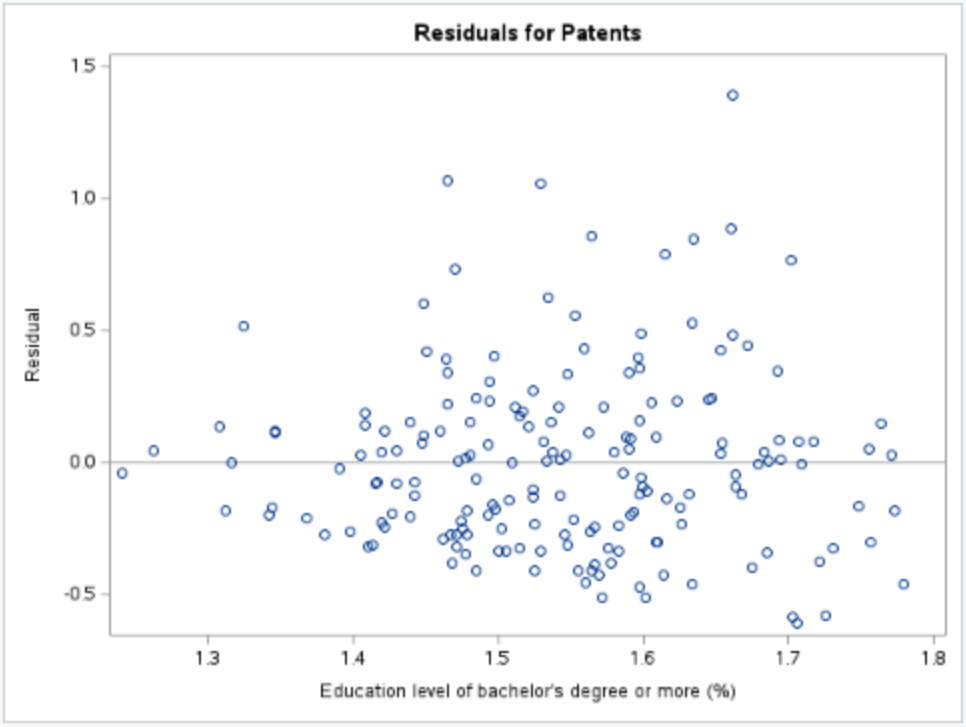

Cook's D for Patents
Cook's D
Observation
0.10
0.08
0.06
0.04
0.02
0.00
0
50
100
150

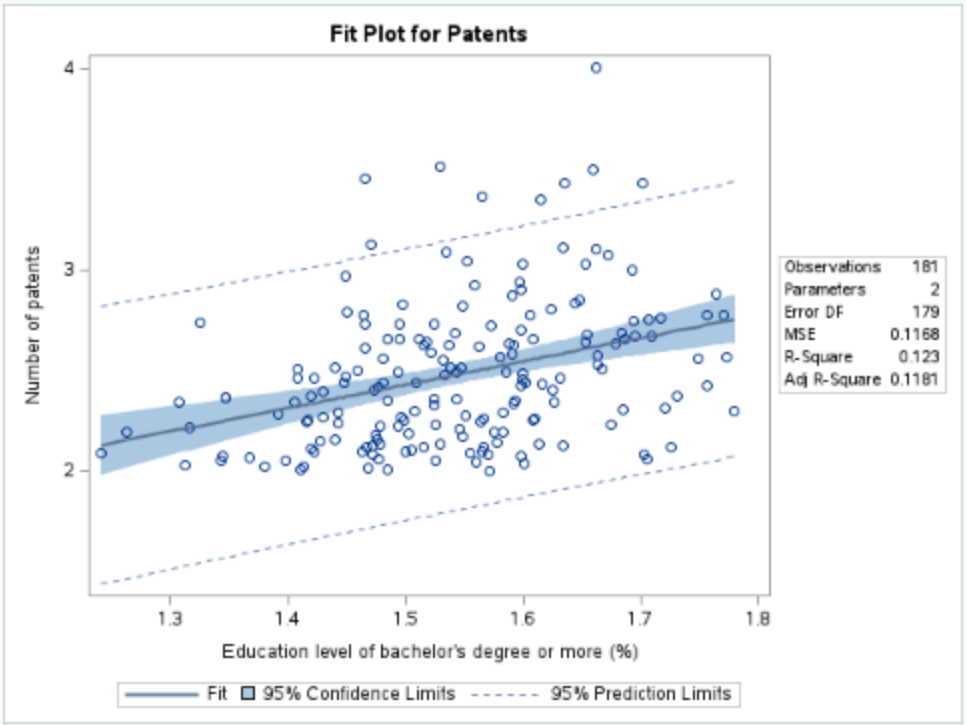

Outlier and Leverage Diagnostics for Patents
RStudent
4
2
0
-2
0.01
0.02
0.03
0.04
0.05
Leverage
Outlier
Leverage
Outlier and Leverage


Q-Q Plot of Residuals for Patents
Residual
1.5
1.0
0.5
0.0
-0.5
-1.0
-3
-2
-1
0
1
2
3
Quantile


Residual-Fit Spread Plot for Patents
Fit–Mean
Residual
1.5
1.0
0.5
0.0
-0.5
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0
Proportion Less


Fit Diagnostics for Patents
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Number of patents
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 181
Parameters 2
Error DF 179
MSE 0.1168
R-Square 0.123
Adj R-Square 0.1181


Residuals for Patents
Residual
1.5
1.0
0.5
0.0
-0.5
1.3
1.4
1.5
1.6
1.7
1.8
Education level of bachelor's degree or more (%)


Fit Plot for Patents
Number of patents
4
3
2
1.3
1.4
1.5
1.6
1.7
1.8
Education level of bachelor's degree or more (%)
Observations 181
Parameters 2
Error DF 179
MSE 0.1168
R-Square 0.123
Adj R-Square 0.1181
Fit
95% Confidence Limits
95% Prediction Limits

28D

/* 28D */

/* For 1% increase in %of people with education level of Bachelor's or more,

the number of patents is expected to increase by 1.16653.

The regression coefficient of Education is statistically significant <.0001 (but the intercept is not p-val: 0.0611) ,

however, the correlation between Education and Patents is only 0.3507, which is still low but higher than before.

The concerns over the constant variance assumption seem to have exacerbated.

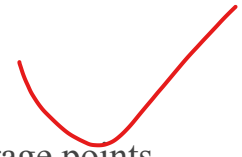
1) The test of first and second moment specification (White test) has a Chi-square p-value of

0.0140 < 0.05, indicating that we reject the null hypothesis of homoscedasticity.

(we have enough evidence to conclude that the variance is not constant).

2) Residuals vs Predicted seem to have more uniformly spreaded points, but Observed vs Predicted shows a pattern.

The adjusted R-square is 0.1181, which means the model has improved in its ability to explain change in patents.



There are still some outlier and high leverage points.

The QQ-plot shows that the errors are not exactly normally distributed, but by my eye it seems to have improved a bit.

Based on the ANOVA table, the p-value for the F-test is <.0001 which means that

we have even stronger evidence to reject the null hypothesis

(no statistically significant relationship between independent variable and dependent variable)

in favor of the alternative hypothesis (there is statistically significant relationship between Patents and Education).

Since there is only 1 variable, it makes sense that it only has VIF = 1 (no multicollinearity).

*/

28E

CODE

```
/* 28E */
PROC REG DATA = PATENTS_LOG10 ALPHA = 0.05
        CORR EDF ALL SIMPLE RSQUARE PLOTS(ONLY)=(FITPLOT RESIDUALS
DIAGNOSTICS COOKSD OBSERVEDBYPREDICTED QQPLOT
RESIDUALBYPREDICTED RESIDUALHISTOGRAM RFPLOT
RSTUDENTBYLEVERAGE RSTUDENTBYPREDICTED);
        MODEL Patents = Education;
        OUTPUT OUT= PATENT_REG_10_RESID
                RESIDUAL = patent_resid;
RUN;

PROC UNIVARIATE NORMAL NORMALTEST DATA = PATENT_REG_10_RESID
ALPHA= 0.05;
        VAR patent_resid;
        CDFPLOT patent_resid / NORMAL;
        HISTOGRAM patent_resid / NORMAL;
        PPPLOT patent_resid / NORMAL;
        PROBPLOT patent_resid / NORMAL;
        QQPLOT patent_resid / NORMAL;
RUN;
```

LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 28E */
70         PROC REG DATA = PATENTS_LOG10 ALPHA = 0.05
```

```
71          CORR EDF ALL SIMPLE RSQUARE PLOTS(ONLY)=(FITPLOT
RESIDUALS DIAGNOSTICS COOKSD OBSERVEDBYPREDICTED QQPLOT
71        ! RESIDUALBYPREDICTED RESIDUALHISTOGRAM RFPLOT
RSTUDENTBYLEVERAGE RSTUDENTBYPREDICTED);
72          MODEL Patents = Education;
73          OUTPUT OUT= PATENT_REG_10_RESID
74          RESIDUAL = patent_resid;
75          RUN;
76
NOTE: The data set WORK.PATENT_REG_10_RESID has 181 observations
and 15 variables.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.95 seconds
      user cpu time       0.65 seconds
      system cpu time     0.06 seconds
      memory              20675.50k
      OS Memory           43236.00k
      Timestamp           05/02/2024 04:43:33 AM
      Step Count                        260  Switch Count  51
      Page Faults                       0
      Page Reclaims                     30269
      Page Swaps                        0
      Voluntary Context Switches        2292
      Involuntary Context Switches      28
      Block Input Operations            0
      Block Output Operations           4352

77          PROC UNIVARIATE NORMAL NORMALTEST DATA =
PATENT_REG_10_RESID ALPHA= 0.05;
78          VAR patent_resid;
79          CDFPLOT patent_resid / NORMAL;
80          HISTOGRAM patent_resid / NORMAL;
```

```
81          PPPLOT patent_resid / NORMAL;
82          PROBPLOT patent_resid / NORMAL;
83          QQPLOT patent_resid / NORMAL;
84          RUN;
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.37 seconds
      user cpu time       0.24 seconds
      system cpu time     0.04 seconds
      memory              25105.12k
      OS Memory           59648.00k
      Timestamp           05/02/2024 04:43:33 AM
      Step Count                        261  Switch Count  0
      Page Faults                       0
      Page Reclaims                     8456
      Page Swaps                        0
      Voluntary Context Switches        1185
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1568

85
86          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
96
```

RESULTS

The REG Procedure

Number of Observations Read	181
Number of Observations Used	181

Descriptive Statistics

Variable	Sum	Mean	Uncorrected SS	Variance	Standard Deviation	Label
Intercept	181.00000	1.00000	181.00000	0	0	Intercept
Education	279.43237	1.54383	433.54901	0.01197	0.10940	Education level of bachelor's degree or more (%)
Patents	448.88202	2.48001	1137.06704	0.13241	0.36389	Number of patents

Uncorrected Sums of Squares and Crossproducts

Variable	Label	Intercept	Education	Patents
Intercept	Intercept	181	279.43236576	448.88201978
Education	Education level of bachelor's degree or more (%)	279.43236576	433.54900732	695.50840577
Patents	Number of patents	448.88201978	695.50840577	1137.0670398

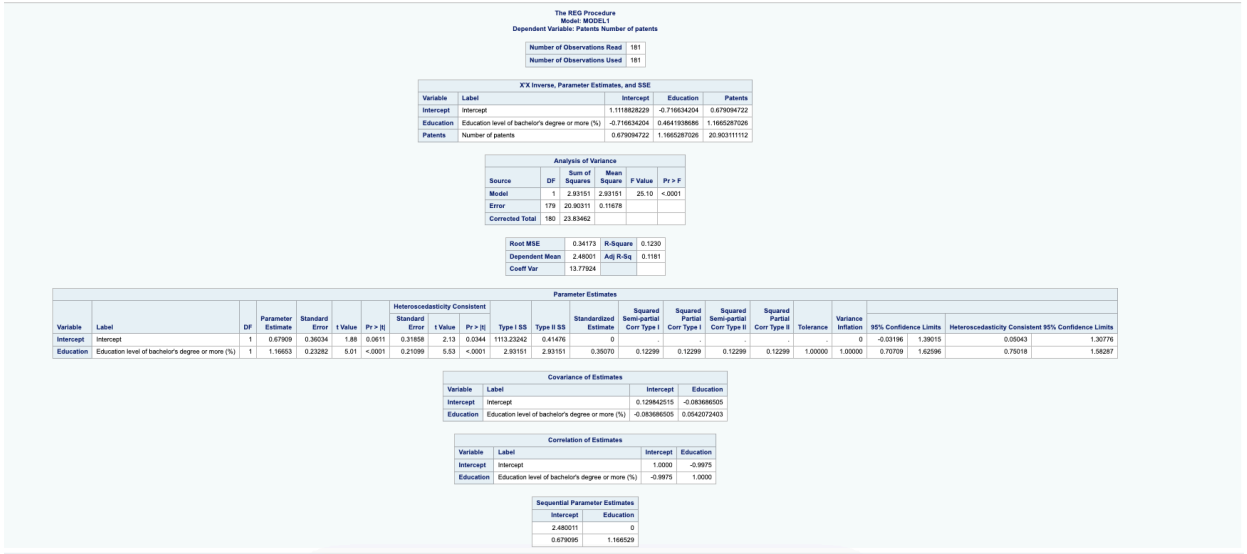
Correlation

Variable	Label	Education	Patents
Education	Education level of bachelor's degree or more (%)	1.0000	0.3507
Patents	Number of patents	0.3507	1.0000

The REG Procedure
Model: MODEL1

Model Crossproducts X'X X'Y Y'Y

Variable	Label	Intercept	Education	Patents
Intercept	Intercept	181	279.43236576	448.88201978
Education	Education level of bachelor's degree or more (%)	279.43236576	433.54900732	695.50840577
Patents	Number of patents	448.88201978	695.50840577	1137.0670398



The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Number of Observations Read	181
Number of Observations Used	181

X'X Inverse, Parameter Estimates, and SSE

Variable	Label	Intercept	Education	Patents
Intercept	Intercept	1.1118828229	-0.716634204	0.679094722
Education	Education level of bachelor's degree or more (%)	-0.716634204	0.4641938686	1.1665287026
Patents	Number of patents	0.679094722	1.1665287026	20.903111112

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.93151	2.93151	25.10	<.0001
Error	179	20.90311	0.11678		
Corrected Total	180	23.83462			

Root MSE	0.34173	R-Square	0.1230
Dependent Mean	2.48001	Adj R-Sq	0.1181
Coeff Var	13.77924		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent Standard Error	Heteroscedasticity Consistent t Value	Heteroscedasticity Consistent Pr > |t|	Type I SS	Type II SS	Standardized Estimate	Squared Semi-partial Corr Type I	Squared Partial Corr Type I	Squared Semi-partial Corr Type II	Squared Partial Corr Type II	Tolerance	Variance Inflation	95% Confidence Limits		Heteroscedasticity Consistent 95% Confidence Limits	
Intercept	Intercept	1	0.67909	0.36034	1.88	0.0611	0.31858	2.13	0.0344	1113.23242	0.41476	0						0	-0.03196	1.39015	0.05043	1.30776
Education	Education level of bachelor's degree or more (%)	1	1.16653	0.23282	5.01	<.0001	0.21099	5.53	<.0001	2.93151	2.93151	0.35070	0.12299	0.12299	0.12299	0.12299	1.00000	1.00000	0.70709	1.62596	0.75018	1.58287

Covariance of Estimates

Variable	Label	Intercept	Education
Intercept	Intercept	0.129842515	-0.083686505
Education	Education level of bachelor's degree or more (%)	-0.083686505	0.0542072403

Correlation of Estimates

Variable	Label	Intercept	Education
Intercept	Intercept	1.0000	-0.9975
Education	Education level of bachelor's degree or more (%)	-0.9975	1.0000

Sequential Parameter Estimates

Intercept	Education
2.480011	0
0.679095	1.166529

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Heteroscedasticity Consistent Covariance of Estimates			
Variable	Label	Intercept	Education
Intercept	Intercept	0.1014950267	-0.067027355
Education	Education level of bachelor's degree or more (%)	-0.067027355	0.0445163336

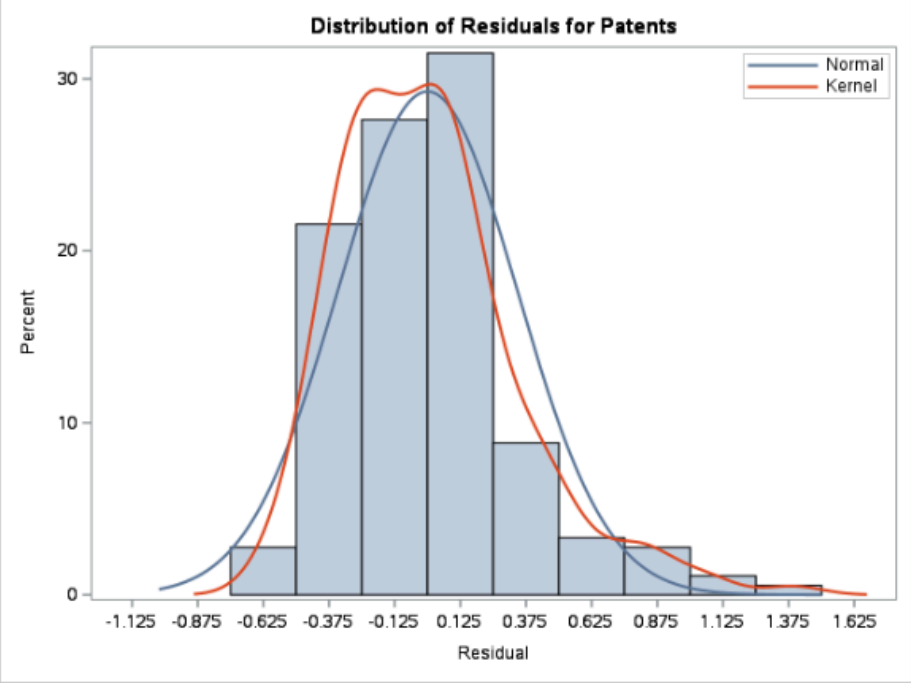
Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
2	8.53	0.0140

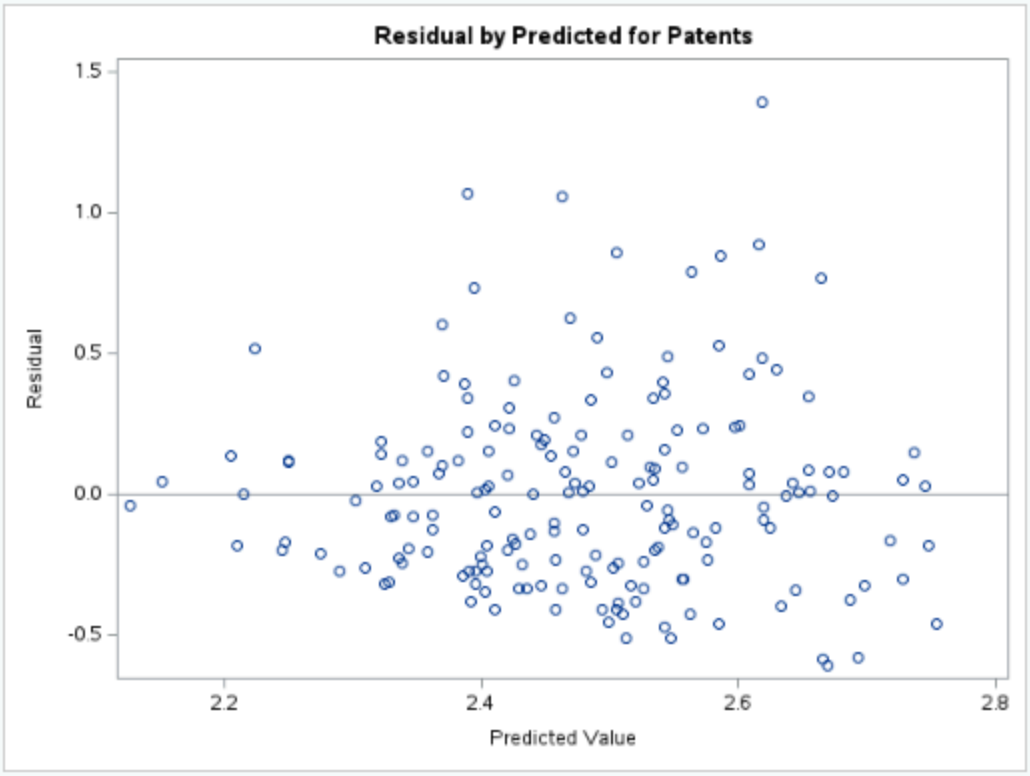
The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

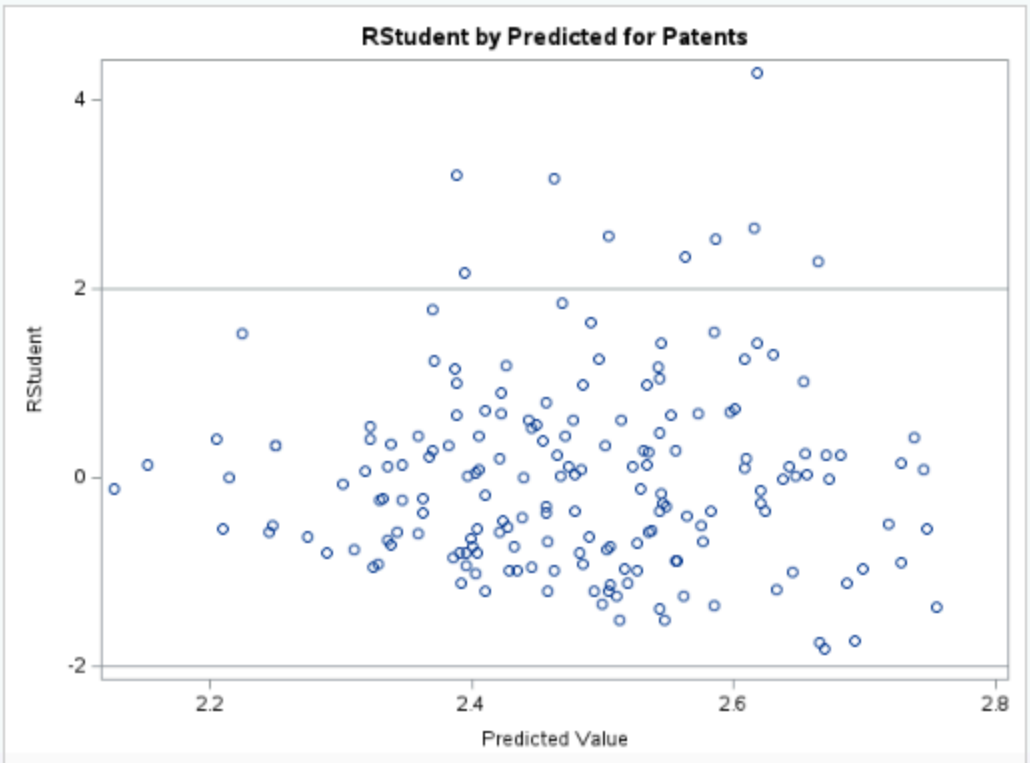
Output Statistics												
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D
1	2.09	2.4932	0.0255	2.4428	2.5435	1.8169	3.1694	-0.4068	0.341	-1.194	| **| |	0.004
2	3.13	2.3937	0.0307	2.3331	2.4543	1.7166	3.0707	0.7315	0.340	2.149	| |**** |	0.019
3	2.78	2.3868	0.0315	2.3246	2.4489	1.7096	3.0640	0.3921	0.340	1.152	| |** |	0.006
4	3.35	2.5629	0.0303	2.5031	2.6227	1.8859	3.2399	0.7879	0.340	2.315	| |**** |	0.021
5	2.87	2.5338	0.0276	2.4794	2.5882	1.8573	3.2103	0.3383	0.341	0.993	| |* |	0.003
6	2.01	2.3920	0.0309	2.3310	2.4529	1.7149	3.0690	-0.3791	0.340	-1.114	| **| |	0.005
7	3.45	2.3885	0.0313	2.3268	2.4502	1.7114	3.0657	1.0654	0.340	3.131	| |****** |	0.041
8	2.31	2.6867	0.0484	2.5911	2.7823	2.0056	3.3677	-0.3770	0.338	-1.115	| **| |	0.013
9	3.36	2.5043	0.0259	2.4533	2.5554	1.8281	3.1806	0.8593	0.341	2.522	| |***** |	0.018
10	2.10	2.5043	0.0259	2.4533	2.5554	1.8281	3.1806	-0.4074	0.341	-1.196	| **| |	0.004
11	2.34	2.2043	0.0606	2.0847	2.3239	1.5195	2.8892	0.1361	0.336	0.405	| | |	0.003
12	2.16	2.3581	0.0352	2.2887	2.4275	1.6802	3.0360	-0.2028	0.340	-0.597	| *| |	0.002
13	2.20	2.1518	0.0703	2.0131	2.2904	1.4633	2.8402	0.0441	0.334	0.132	| | |	0.000
14	3.52	2.4626	0.0256	2.4120	2.5132	1.7864	3.1388	1.0550	0.341	3.096	| |****** |	0.027
15	2.76	2.6818	0.0476	2.5879	2.7758	2.0010	3.3627	0.0823	0.338	0.243	| | |	0.001
16	3.43	2.5858	0.0330	2.5206	2.6509	1.9083	3.2632	0.8448	0.340	2.484	| |**** |	0.029
17	2.44	2.4396	0.0266	2.3870	2.4922	1.7632	3.1160	-0.001872	0.341	-0.005	| | |	0.000
18	4.01	2.6176	0.0374	2.5438	2.6915	1.9393	3.2960	1.3918	0.340	4.098	| |****** |	0.102
19	2.72	2.5139	0.0263	2.4620	2.5658	1.8376	3.1902	0.2104	0.341	0.617	| |* |	0.001
20	2.30	2.4381	0.0267	2.3853	2.4908	1.7617	3.1144	-0.1414	0.341	-0.415	| | |	0.001
21	2.73	2.4221	0.0279	2.3670	2.4771	1.7455	3.0986	0.3063	0.341	0.899	| |* |	0.003
22	2.00	2.5125	0.0262	2.4608	2.5643	1.8362	3.1889	-0.5125	0.341	-1.504	| ***| |	0.007
23	2.19	2.5259	0.0270	2.4727	2.5792	1.8495	3.2024	-0.3356	0.341	-0.985	| *| |	0.003

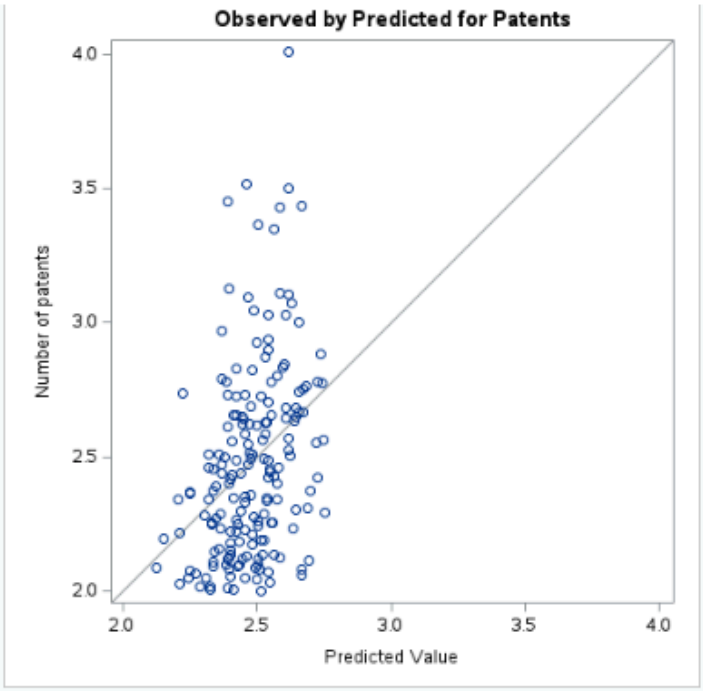
169	2.05	2.4988	0.0257	2.4481	2.5494	1.8225	3.1750	-0.4534	0.341	-1.331	**	0.005
170	2.56	2.7474	0.0591	2.6308	2.8640	2.0631	3.4318	-0.1827	0.337	-0.543	*	0.005
171	2.56	2.7184	0.0539	2.6120	2.8248	2.0357	3.4011	-0.1633	0.337	-0.484		0.003
172	2.39	2.3469	0.0367	2.2744	2.4195	1.6687	3.0252	0.0458	0.340	0.135		0.000
173	3.50	2.6154	0.0371	2.5422	2.6886	1.9371	3.2937	0.8838	0.340	2.602	*****	0.040
174	2.73	2.3885	0.0313	2.3268	2.4502	1.7114	3.0657	0.3415	0.340	1.003	**	0.004
175	2.06	2.2742	0.0483	2.1788	2.3695	1.5931	2.9552	-0.2097	0.338	-0.620	*	0.004
176	2.51	2.3581	0.0352	2.2887	2.4275	1.6802	3.0360	0.1524	0.340	0.448		0.001
177	2.57	2.6198	0.0377	2.5454	2.6943	1.9414	3.2983	-0.0481	0.340	-0.142		0.000
178	2.24	2.3618	0.0347	2.2934	2.4302	1.6840	3.0396	-0.1263	0.340	-0.371		0.001
179	2.02	2.3277	0.0396	2.2496	2.4059	1.6489	3.0066	-0.3107	0.339	-0.915	*	0.006
180	2.44	2.5479	0.0288	2.4911	2.6048	1.8712	3.2247	-0.1055	0.341	-0.310		0.000
181	2.11	2.3355	0.0384	2.2597	2.4113	1.6569	3.0141	-0.2249	0.340	-0.662	*	0.003

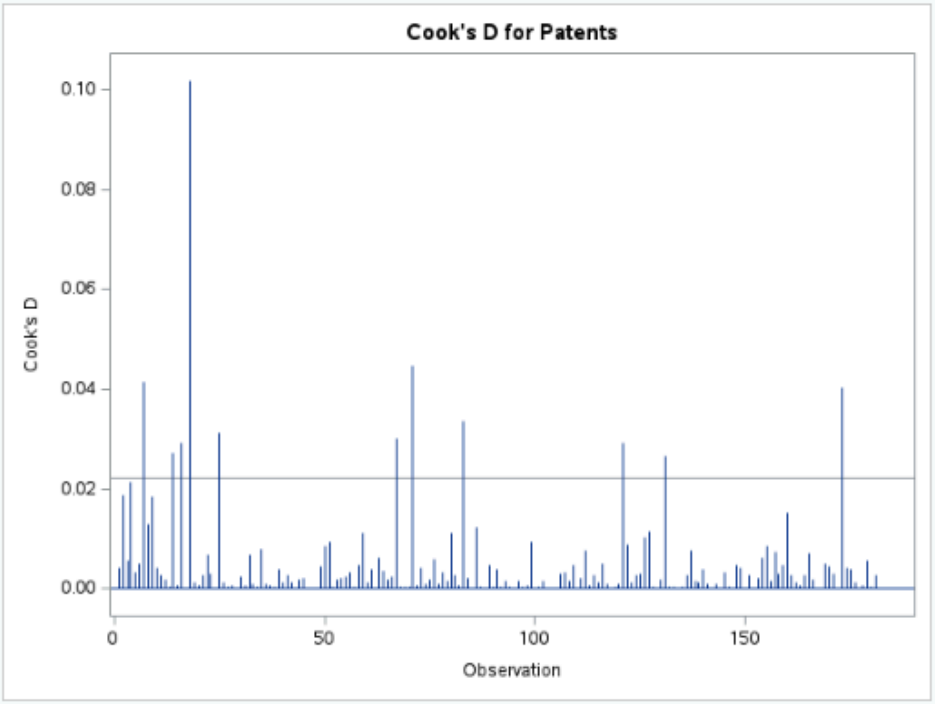
Sum of Residuals	0
Sum of Squared Residuals	20.90311
Predicted Residual SS (PRESS)	21.33428

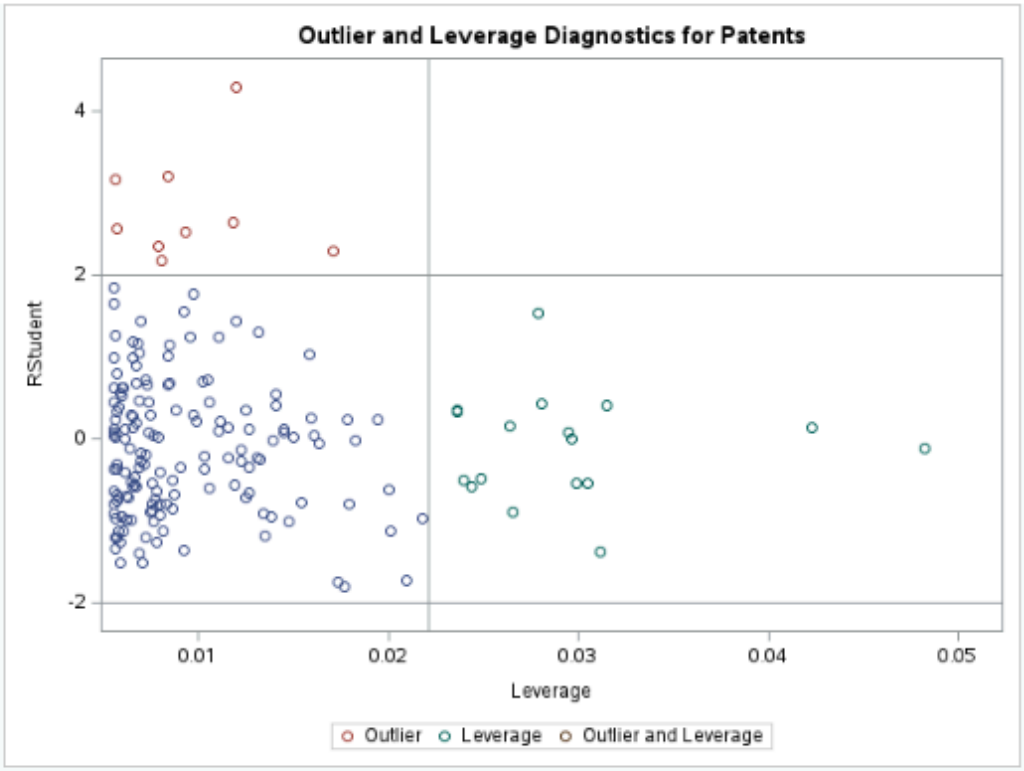


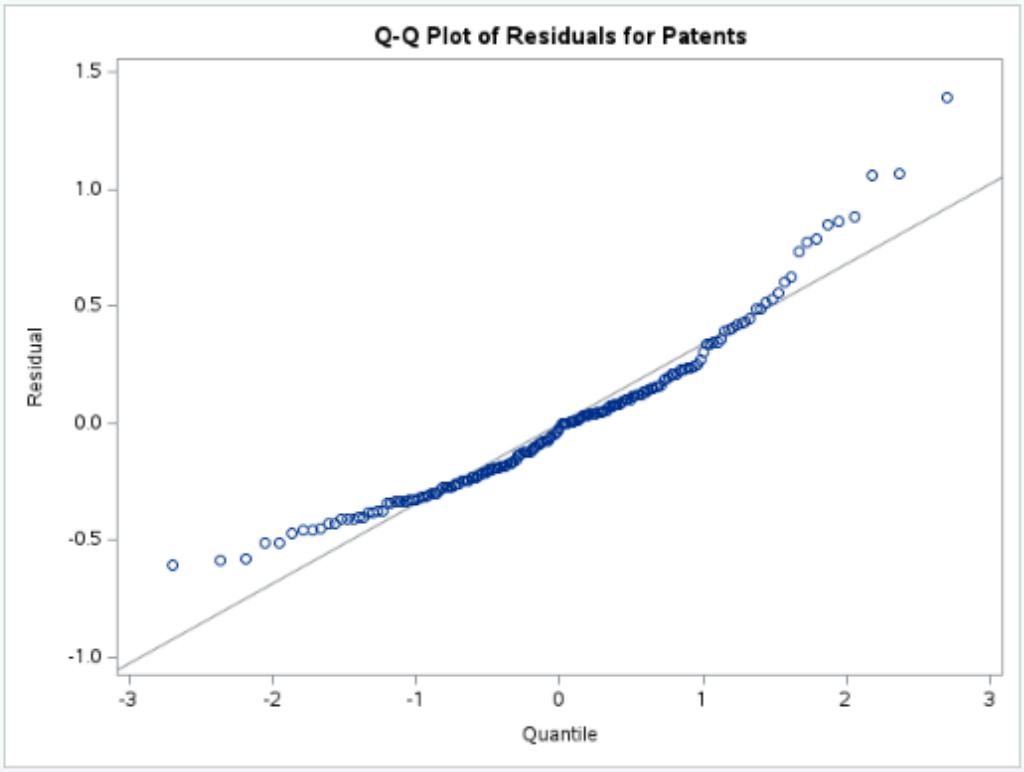

Residual by Predicted for Patents
Residual
1.5
1.0
0.5
0.0
-0.5
2.2
2.4
2.6
2.8
Predicted Value

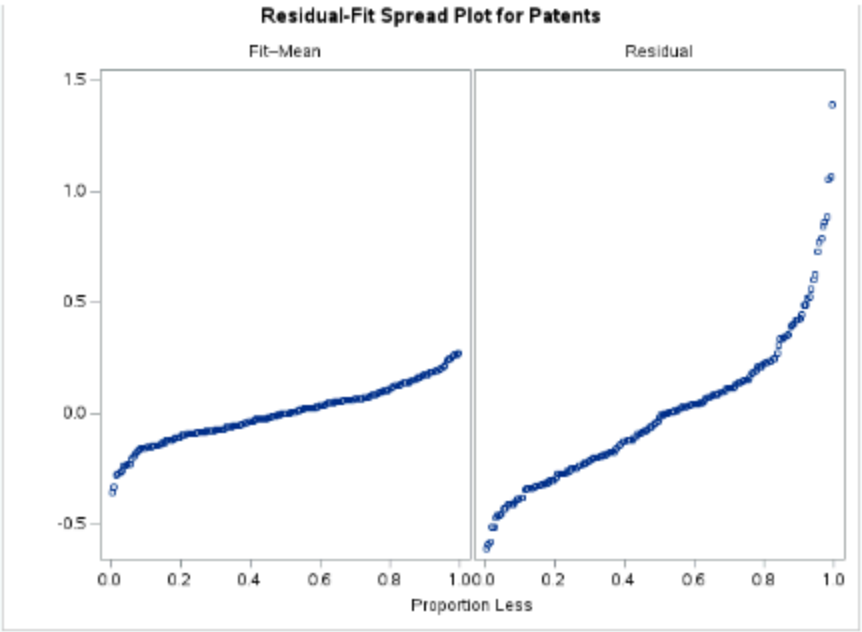

RStudent by Predicted for Patents
RStudent
4
2
0
-2
2.2
2.4
2.6
2.8
Predicted Value

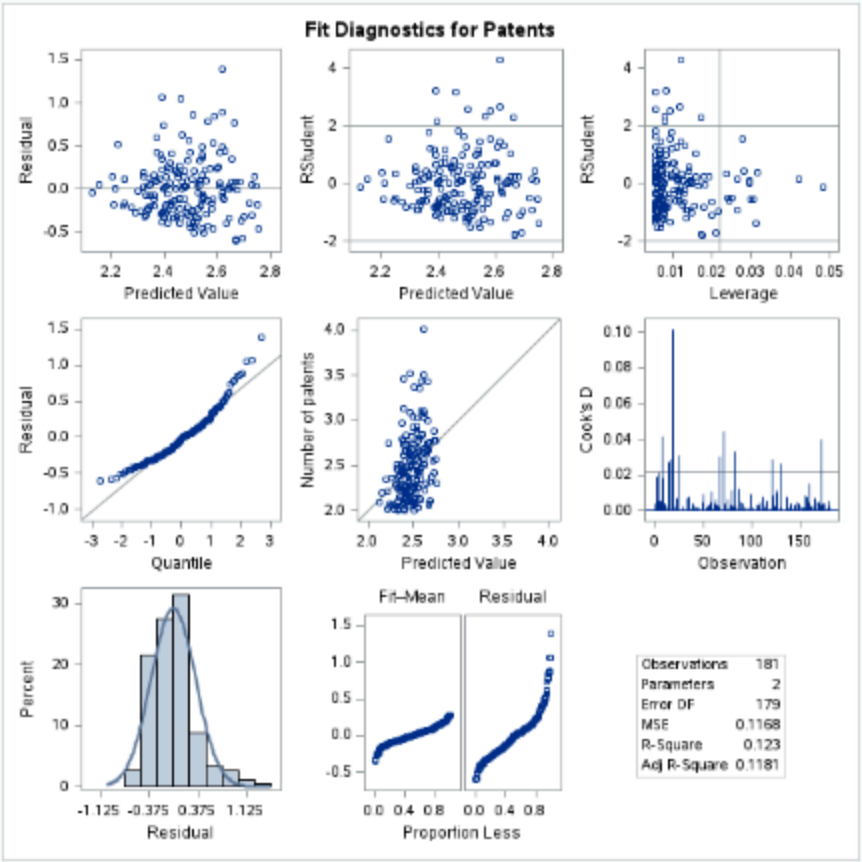

Observed by Predicted for Patents
Number of patents
Predicted Value
4.0
3.5
3.0
2.5
2.0
2.0
2.5
3.0
3.5
4.0

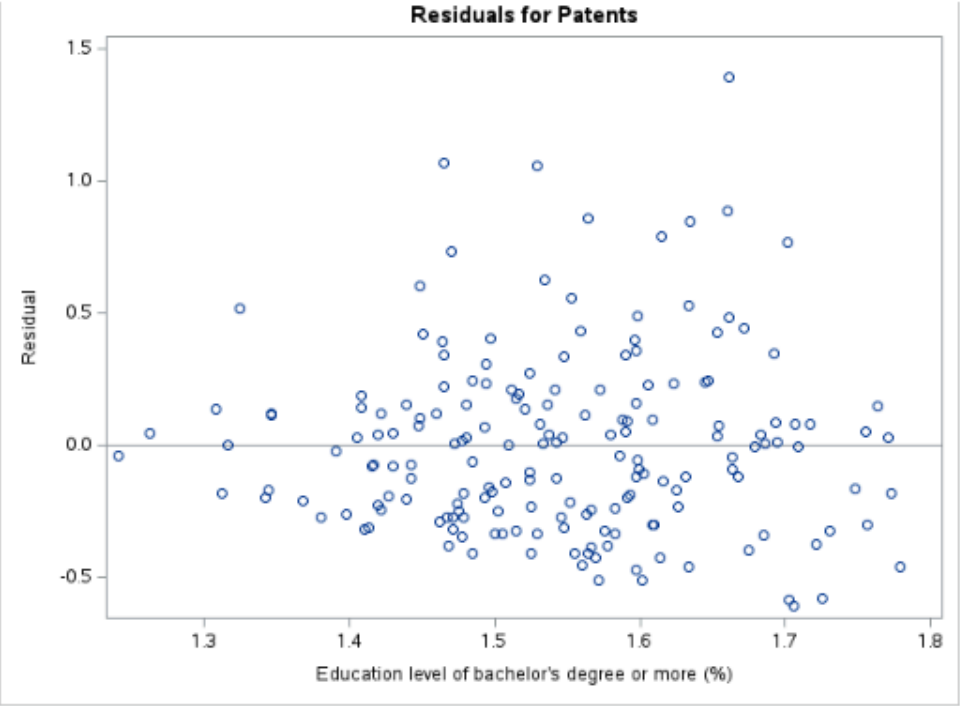

Cook's D for Patents
Cook's D
Observation
0.10
0.08
0.06
0.04
0.02
0.00
0
50
100
150

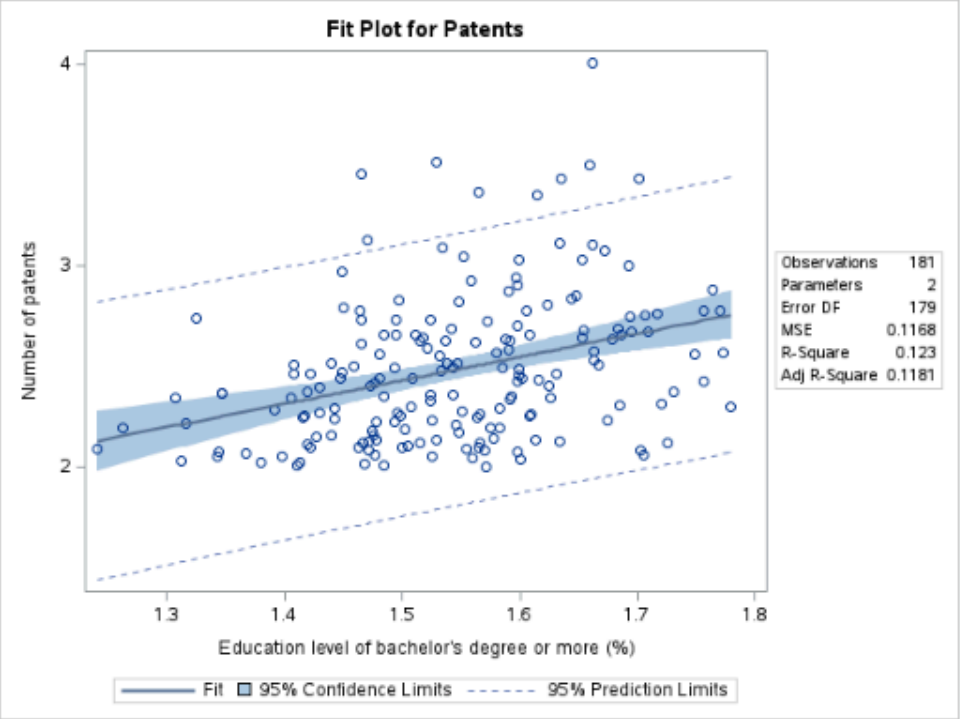

Outlier and Leverage Diagnostics for Patents
RStudent
4
2
0
-2
0.01
0.02
0.03
0.04
0.05
Leverage
Outlier
Leverage
Outlier and Leverage


Q-Q Plot of Residuals for Patents
Residual
1.5
1.0
0.5
0.0
-0.5
-1.0
-3
-2
-1
0
1
2
3
Quantile


Residual-Fit Spread Plot for Patents
Fit–Mean
Residual
Proportion Less


Fit Diagnostics for Patents
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Number of patents
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 181
Parameters 2
Error DF 179
MSE 0.1168
R-Square 0.123
Adj R-Square 0.1181


Residuals for Patents
Residual
1.5
1.0
0.5
0.0
-0.5
1.3
1.4
1.5
1.6
1.7
1.8
Education level of bachelor's degree or more (%)


Fit Plot for Patents
Number of patents
4
3
2
1.3
1.4
1.5
1.6
1.7
1.8
Education level of bachelor's degree or more (%)
Observations 181
Parameters 2
Error DF 179
MSE 0.1168
R-Square 0.123
Adj R-Square 0.1181
Fit
95% Confidence Limits
95% Prediction Limits

The UNIVARIATE Procedure
Variable: patent_resid (Residual)

Moments			
N	181	Sum Weights	181
Mean	0	Sum Observations	0
Std Deviation	0.34077617	Variance	0.1161284
Skewness	1.06273982	Kurtosis	1.71087356
Uncorrected SS	20.9031111	Corrected SS	20.9031111
Coeff Variation	.	Std Error Mean	0.02532969

Basic Statistical Measures			
Location		Variability	
Mean	0.00000	Std Deviation	0.34078
Median	-0.02063	Variance	0.11613
Mode	.	Range	2.00018
		Interquartile Range	0.39776

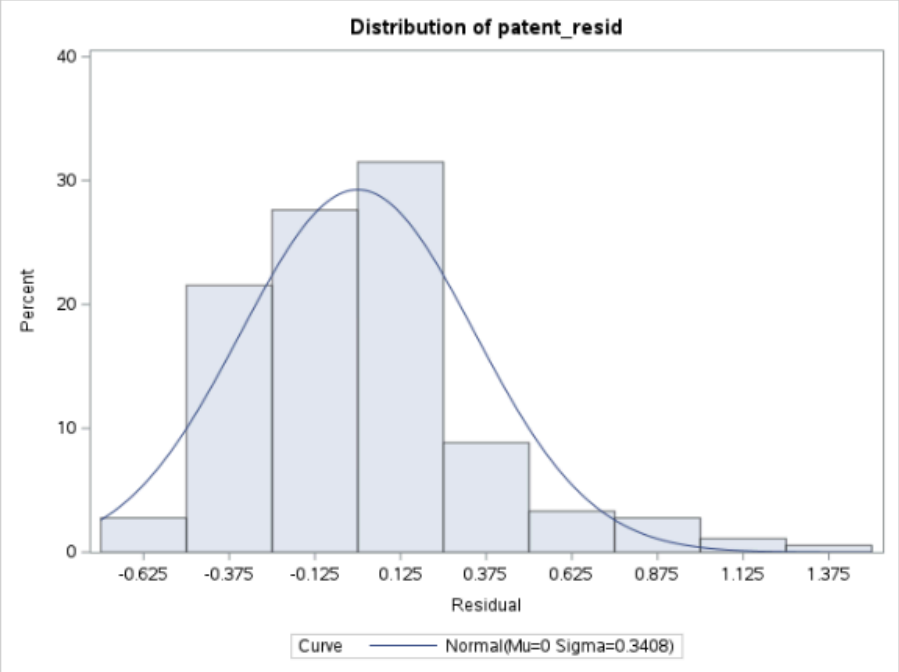
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	-3.5	Pr >= |M|	0.6557
Signed Rank	S	-831.5	Pr >= |S|	0.2399

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.940261	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.084237	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.338668	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.327616	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	1.3918465
99%	1.0654237
95%	0.6244911
90%	0.4241977
75% Q3	0.1507372
50% Median	-0.0206271
25% Q1	-0.2470277
10%	-0.3791268
5%	-0.4281486
1%	-0.5868518
0% Min	-0.6083359

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-0.608336	121	0.859289	9
-0.586852	131	0.883840	173
-0.578477	25	1.054972	14
-0.513252	35	1.065424	7
-0.512538	22	1.391846	18

The UNIVARIATE Procedure

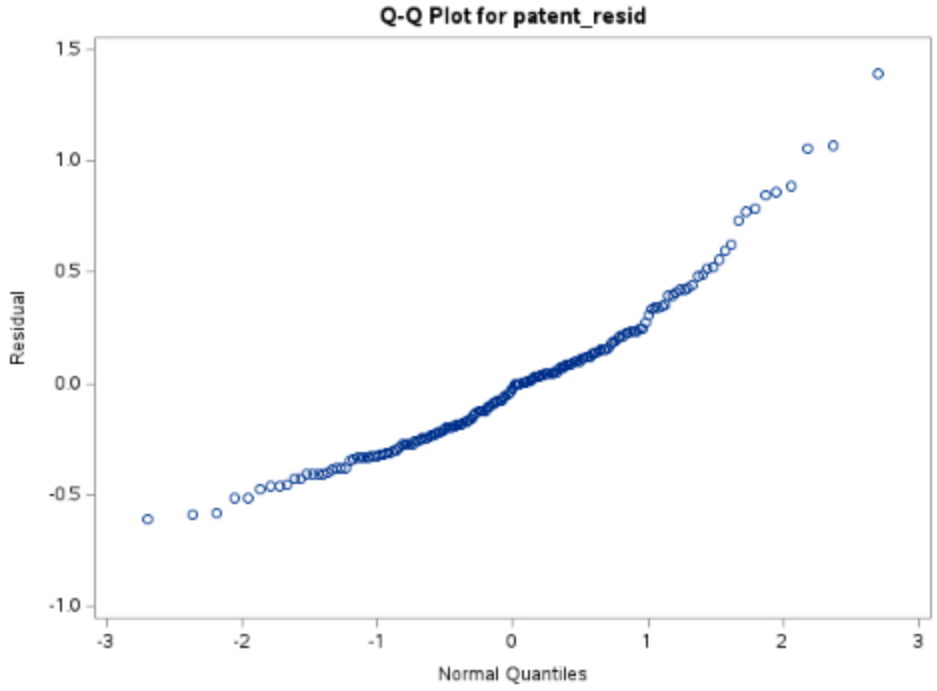


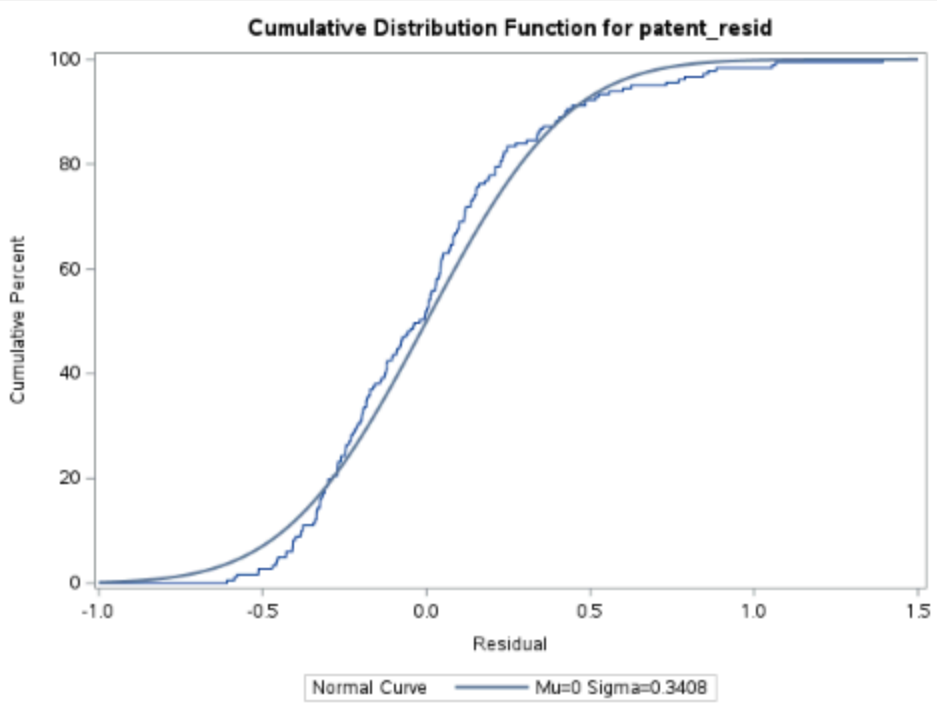
The UNIVARIATE Procedure
Fitted Normal Distribution for patent_resid (Residual)

Parameters for Normal Distribution		
Parameter	**Symbol**	**Estimate**
Mean	Mu	0
Std Dev	Sigma	0.340776

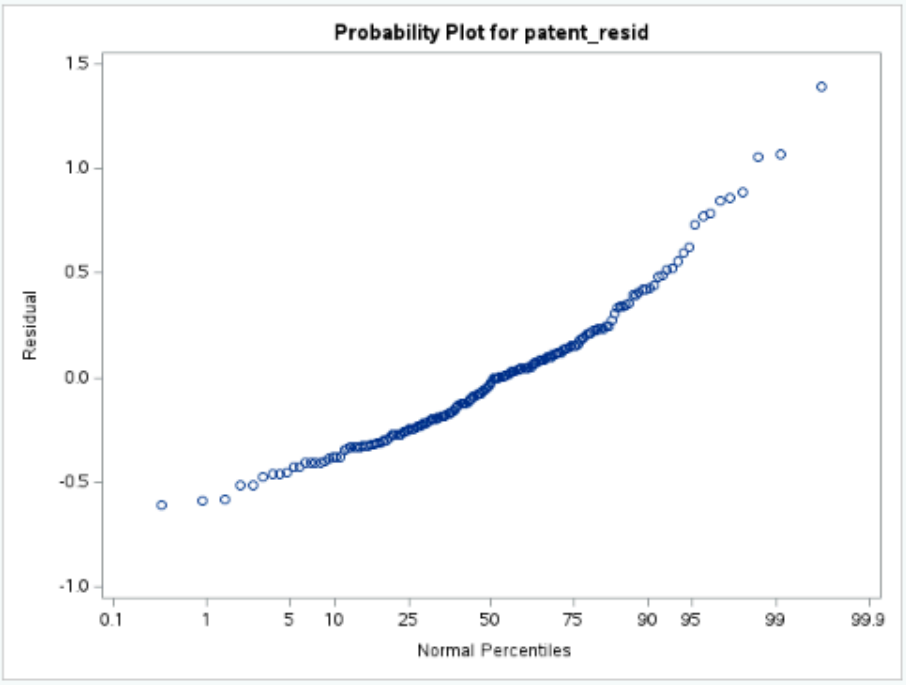
Goodness-of-Fit Tests for Normal Distribution				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.08423721	**Pr > D**	<0.010
Cramer-von Mises	**W-Sq**	0.33866817	**Pr > W-Sq**	<0.005
Anderson-Darling	**A-Sq**	2.32761554	**Pr > A-Sq**	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	**Observed**	**Estimated**
1.0	-0.58685	-0.79276
5.0	-0.42815	-0.56053
10.0	-0.37913	-0.43672
25.0	-0.24703	-0.22985
50.0	-0.02063	0.00000
75.0	0.15074	0.22985
90.0	0.42420	0.43672
95.0	0.62449	0.56053
99.0	1.06542	0.79276

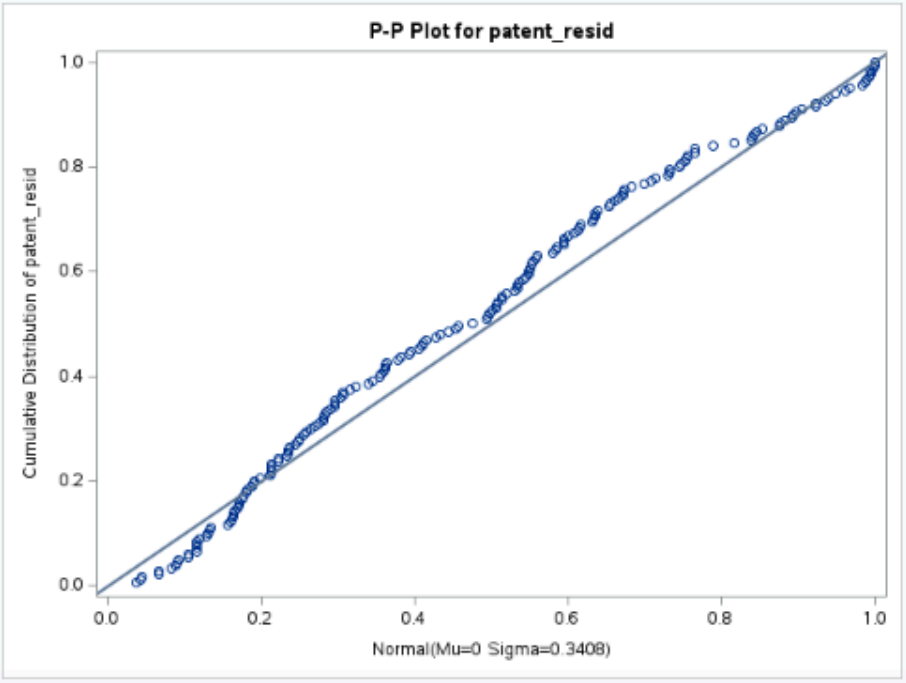

Q-Q Plot for patent_resid
Residual
Normal Quantiles
1.5
1.0
0.5
0.0
-0.5
-1.0
-3
-2
-1
0
1
2
3


Cumulative Distribution Function for patent_resid
Cumulative Percent
Residual
100
80
60
40
20
0
-1.0
-0.5
0.0
0.5
1.0
1.5
Normal Curve
Mu=0 Sigma=0.3408

The UNIVARIATE Procedure



The UNIVARIATE Procedure



28F

/* 28F */

/*

Based on the results, all normality tests (Anderson-Darling, Cramer-von Mises, Kolmogorov-Smirnov)

yielded p-values <0.05, which means that we reject the null hypothesis that the residuals are normally distributed.

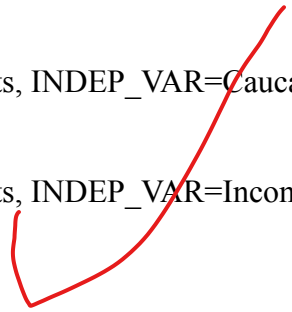
*/

28G

CODE

```
/* 28G */
PROC OPTIONS OPTION=MACRO;
RUN;
PROC OPTIONS OPTION = SYMBOLGEN;
RUN;


%MACRO PATENT_REG (DEP_VAR=, INDEP_VAR=, DATASET_NAME=);

	%LET RESID_DATASET = &DEP_VAR._resid;

	PROC REG DATA = &DATASET_NAME ALPHA = 0.05
		CORR EDF ALL SIMPLE RSQUARE
```

```
        PLOTS(ONLY)=(FITPLOT RESIDUALS DIAGNOSTICS COOKSD
OBSERVEDBYPREDICTED QQPLOT RESIDUALBYPREDICTED
RESIDUALHISTOGRAM RFPLOT RSTUDENTBYLEVERAGE
RSTUDENTBYPREDICTED);
        MODEL &DEP_VAR = &INDEP_VAR;
        OUTPUT OUT= &RESID_DATASET
            RESIDUAL = resid;
        TITLE "REGRESSION ANALYSES FOR &DEP_VAR AND &INDEP_VAR AS
REGRESSOR (&DATASET_NAME.) " ;
    RUN;

    PROC UNIVARIATE NORMAL NORMALTEST DATA = &RESID_DATASET
ALPHA= 0.05;
        VAR resid;
        CDFPLOT resid / NORMAL;
        HISTOGRAM resid / NORMAL;
        PPPLOT resid / NORMAL;
        PROBPLOT resid / NORMAL;
        QQPLOT resid / NORMAL;
        TITLE "RESIDUAL NORMALITY ANALYSES FOR &DEP_VAR
(&DATASET_NAME.) " ;
    RUN;

%MEND PATENT_REG;

%PATENT_REG(DEP_VAR=Patents, INDEP_VAR=Caucasian,
DATASET_NAME=PATENTS)
%PATENT_REG(DEP_VAR=Patents, INDEP_VAR=Income, DATASET_NAME=PATENTS)
```



LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* 28G */
70         PROC OPTIONS OPTION=MACRO;
71         RUN;
    SAS (r) Proprietary Software Release 9.4  TS1M7
 MACRO             Enables the macro facility.
NOTE: PROCEDURE OPTIONS used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              404.28k
      OS Memory           28068.00k
      Timestamp           05/02/2024 04:46:58 AM
      Step Count                        280  Switch Count  0
      Page Faults                       0
      Page Reclaims                     14
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0

72         PROC OPTIONS OPTION = SYMBOLGEN;
73         RUN;
    SAS (r) Proprietary Software Release 9.4  TS1M7
 NOSYMBOLGEN       Does not display the results of resolving
macro variable references in the SAS log.
NOTE: PROCEDURE OPTIONS used (Total process time):
      real time           0.00 seconds
```

```
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              405.25k
      OS Memory           28068.00k
      Timestamp           05/02/2024 04:46:58 AM
      Step Count                        281  Switch Count  0
      Page Faults                       0
      Page Reclaims                     14
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0
```

```
74
75
76         %MACRO PATENT_REG (DEP_VAR=, INDEP_VAR=,
DATASET_NAME=);
77
78         %LET RESID_DATASET = &DEP_VAR._resid;
79
80         PROC REG DATA = &DATASET_NAME ALPHA = 0.05
81         CORR EDF ALL SIMPLE RSQUARE
82         PLOTS(ONLY)=(FITPLOT RESIDUALS DIAGNOSTICS COOKSD
OBSERVEDBYPREDICTED QQPLOT RESIDUALBYPREDICTED RESIDUALHISTOGRAM
82       ! RFPLOT RSTUDENTBYLEVERAGE RSTUDENTBYPREDICTED);
83         MODEL &DEP_VAR = &INDEP_VAR;
84         OUTPUT OUT= &RESID_DATASET
85         RESIDUAL = resid;
86         TITLE "REGRESSION ANALYSES FOR &DEP_VAR AND
&INDEP_VAR AS REGRESSOR (&DATASET_NAME.) " ;
87         RUN;
```

```
88
89          PROC UNIVARIATE NORMAL NORMALTEST DATA =
&RESID_DATASET ALPHA= 0.05;
90          VAR resid;
91          CDFPLOT resid / NORMAL;
92          HISTOGRAM resid / NORMAL;
93          PPPLOT resid / NORMAL;
94          PROBPLOT resid / NORMAL;
95          QQPLOT resid / NORMAL;
96          TITLE "RESIDUAL NORMALITY ANALYSES FOR &DEP_VAR
(&DATASET_NAME.) " ;
97          RUN;
98
99          %MEND PATENT_REG;
100
101         %PATENT_REG(DEP_VAR=Patents, INDEP_VAR=Caucasian,
DATASET_NAME=PATENTS)
NOTE: The data set WORK.PATENTS_RESID has 808 observations and
15 variables.
NOTE: PROCEDURE REG used (Total process time):
      real time           1.82 seconds
      user cpu time       1.44 seconds
      system cpu time     0.08 seconds
      memory              20675.25k
      OS Memory           43496.00k
      Timestamp           05/02/2024 04:47:00 AM
      Step Count                        282  Switch Count  52
      Page Faults                       0
      Page Reclaims                     30270
      Page Swaps                        0
      Voluntary Context Switches        3394
      Involuntary Context Switches      3
```

```
      Block Input Operations            0
      Block Output Operations           7112

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.42 seconds
      user cpu time       0.29 seconds
      system cpu time     0.02 seconds
      memory              25299.84k
      OS Memory           59652.00k
      Timestamp           05/02/2024 04:47:01 AM
      Step Count                        283  Switch Count  1
      Page Faults                       0
      Page Reclaims                     8310
      Page Swaps                        0
      Voluntary Context Switches        1204
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           2112

102         %PATENT_REG(DEP_VAR=Patents, INDEP_VAR=Income,
DATASET_NAME=PATENTS)
NOTE: The data set WORK.PATENTS_RESID has 808 observations and
15 variables.
NOTE: PROCEDURE REG used (Total process time):
      real time           1.79 seconds
      user cpu time       1.42 seconds
      system cpu time     0.09 seconds
      memory              13058.31k
      OS Memory           44264.00k
      Timestamp           05/02/2024 04:47:02 AM
      Step Count                        284  Switch Count  52
      Page Faults                       0
```

```
      Page Reclaims                     28424
      Page Swaps                        0
      Voluntary Context Switches        3240
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           7320

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.42 seconds
      user cpu time       0.29 seconds
      system cpu time     0.03 seconds
      memory              25123.56k
      OS Memory           59652.00k
      Timestamp           05/02/2024 04:47:03 AM
      Step Count                        285  Switch Count  1
      Page Faults                       0
      Page Reclaims                     8263
      Page Swaps                        0
      Voluntary Context Switches        1205
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           2360

103
104        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
114
```

User: u62223361

Messages: 33

RESULTS

REGRESSION ANALYSES FOR Patents AND Caucasian AS REGRESSOR (PATENTS)

The REG Procedure

Number of Observations Read	808
Number of Observations Used	808

Descriptive Statistics

Variable	Sum	Mean	Uncorrected SS	Variance	Standard Deviation	Label
Intercept	808.00000	1.00000	808.00000	0	0	Intercept
Caucasian	66129	81.84295	5578447	206.00622	14.35292	Caucasian (%)
Patents	103978	128.68564	191949046	221275	470.39830	Number of patents

Uncorrected Sums of Squares and Crossproducts

Variable	Label	Intercept	Caucasian	Patents
Intercept	Intercept	808	66129.1	103978
Caucasian	Caucasian (%)	66129.1	5578447.35	7703618.8
Patents	Number of patents	103978	7703618.8	191949046

Correlation

Variable	Label	Caucasian	Patents
Caucasian	Caucasian (%)	1.0000	-0.1480
Patents	Number of patents	-0.1480	1.0000

REGRESSION ANALYSES FOR Patents AND Caucasian AS REGRESSOR (PATENTS)

The REG Procedure
Model: MODEL1

Model Crossproducts X'X X'Y Y'Y

Variable	Label	Intercept	Caucasian	Patents
Intercept	Intercept	808	66129.1	103978
Caucasian	Caucasian (%)	66129.1	5578447.35	7703618.8
Patents	Number of patents	103978	7703618.8	191949046

REGRESSION ANALYSES FOR Patents AND Caucasian AS REGRESSOR (PATENTS)

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Number of Observations Read	808
Number of Observations Used	808

X'X Inverse, Parameter Estimates, and SSE

Variable	Label	Intercept	Caucasian	Patents
Intercept	Intercept	0.0415286783	-0.000492297	525.59879564
Caucasian	Caucasian (%)	-0.000492297	6.0151454E-6	-4.849692902
Patents	Number of patents	525.59879564	-4.849692902	174658519.84

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3910050	3910050	18.04	<.0001
Error	806	174658520	216698		
Corrected Total	807	178568570			

Root MSE	465.50823	R-Square	0.0219
Dependent Mean	128.68564	Adj R-Sq	0.0207
Coeff Var	361.74061		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent			Type I SS	Type II SS	Standardized Estimate	Squared Semi-partial Corr Type I	Squared Partial Corr Type I	Squared Semi-partial Corr Type II	Squared Partial Corr Type II	Tolerance	Variance Inflation	95% Confidence Limits		Heteroscedasticity Consistent 95% Confidence Limits	
							Standard Error	t Value	Pr > |t|													
Intercept	Intercept	1	525.59880	94.86400	5.54	<.0001	165.86075	3.17	0.0016	13380476	6652128	0	.	.	.	.	.	0	339.38915	711.80844	200.02880	851.16879
Caucasian	Caucasian (%)	1	-4.84969	1.14170	-4.25	<.0001	1.84937	-2.62	0.0089	3910050	3910050	-0.14798	0.02190	0.02190	0.02190	0.02190	1.00000	1.00000	-7.09074	-2.60864	-8.47984	-1.21954

Covariance of Estimates

Variable	Label	Intercept	Caucasian
Intercept	Intercept	8999.178011	-106.6797812
Caucasian	Caucasian (%)	-106.6797812	1.303469474

Correlation of Estimates

Variable	Label	Intercept	Caucasian
Intercept	Intercept	1.0000	-0.9850
Caucasian	Caucasian (%)	-0.9850	1.0000

Sequential Parameter Estimates

Intercept	Caucasian
128.685644	0
525.598796	-4.849693

REGRESSION ANALYSES FOR Patents AND Caucasian AS REGRESSOR (PATENTS)

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Heteroscedasticity Consistent Covariance of Estimates			
Variable	Label	Intercept	Caucasian
Intercept	Intercept	27509.78965	-306.3885783
Caucasian	Caucasian (%)	-306.3885783	3.4201715076

Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
2	6.30	0.0429

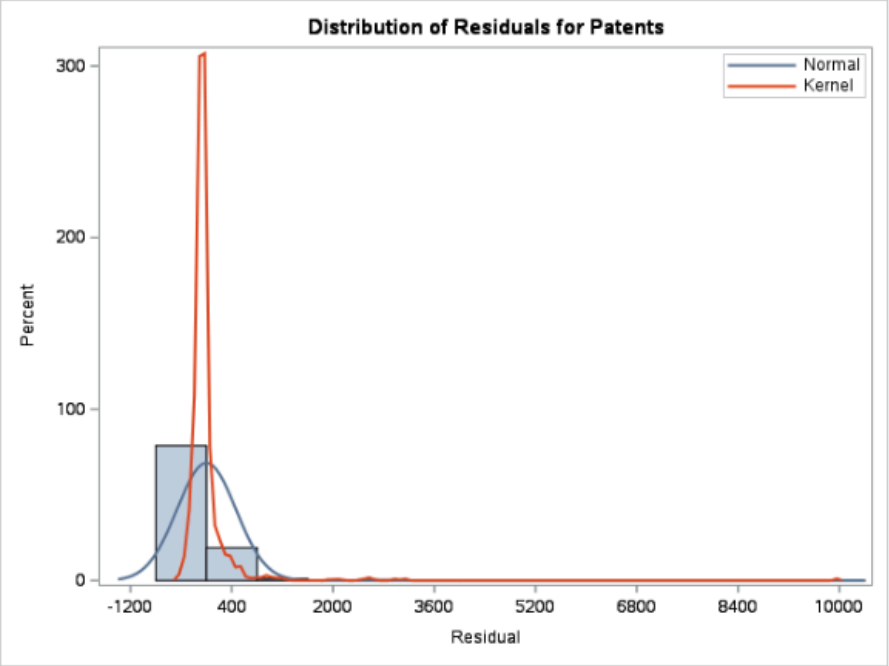
REGRESSION ANALYSES FOR Patents AND Caucasian AS REGRESSOR (PATENTS)

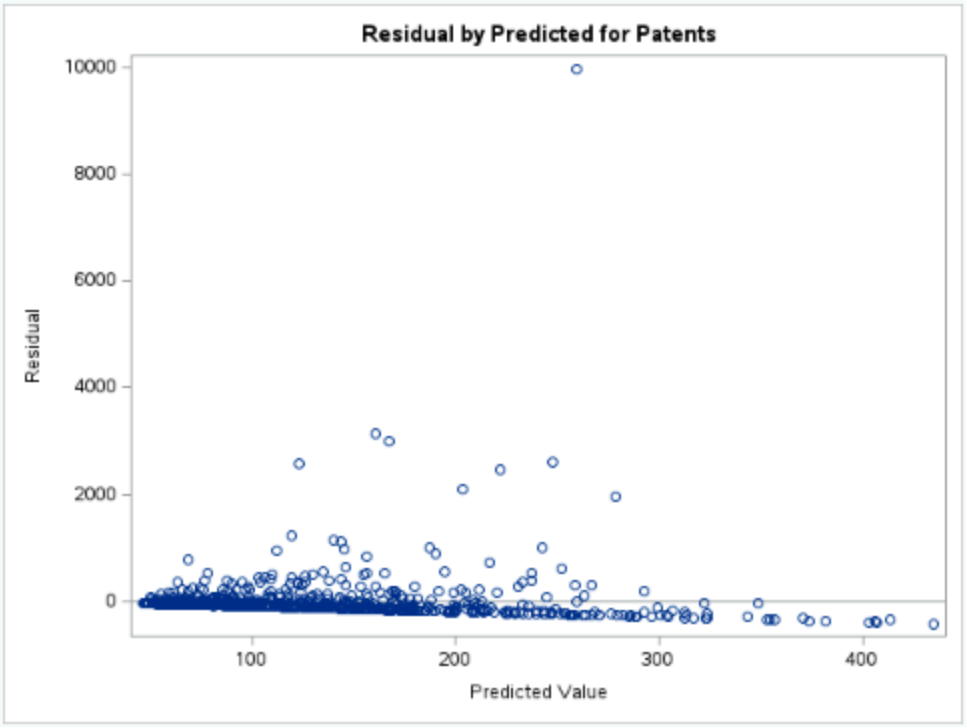
The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

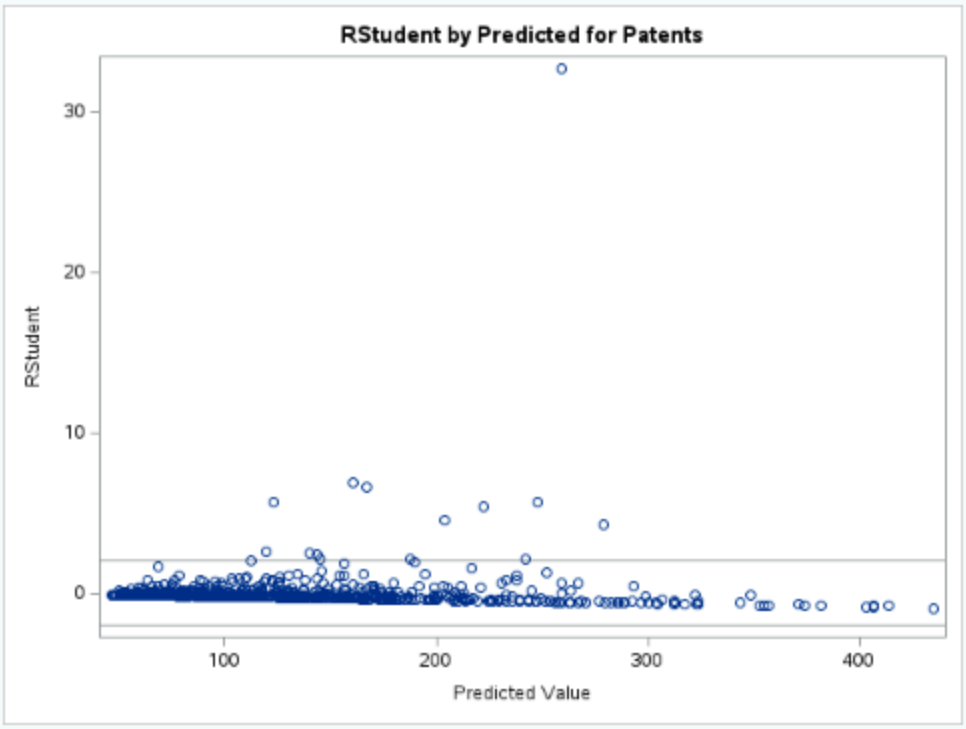
Output Statistics												
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D
1	8	98.8258	17.8215	63.8439	133.8078	-815.5951	1013	-90.8258	465.2	-0.195	| | |	0.000
2	1	153.1424	17.3591	119.0679	187.2168	-761.2442	1068	-152.1424	465.2	-0.327	| | |	0.000
3	4	58.0884	23.3325	12.2888	103.8880	-856.8102	972.9870	-54.0884	464.9	-0.116	| | |	0.000
4	2	77.0022	20.4017	36.9555	117.0489	-837.6264	991.6309	-75.0022	465.1	-0.161	| | |	0.000
5	2	154.5973	17.4757	120.2940	188.9006	-759.7979	1069	-152.5973	465.2	-0.328	| | |	0.000
6	2	127.9240	16.3775	95.7764	160.0716	-786.3929	1042	-125.9240	465.2	-0.271	| | |	0.000
7	3	180.7856	20.4603	140.6238	220.9474	-733.8481	1095	-177.7856	465.1	-0.382	| | |	0.000
8	51	268.0801	36.6751	196.0902	340.0700	-648.5029	1185	-217.0801	464.1	-0.468	| | |	0.001
9	5	99.7958	17.7326	64.9882	134.6033	-814.6185	1014	-94.7958	465.2	-0.204	| | |	0.000
10	24	178.3608	20.1233	138.8605	217.8611	-736.2441	1093	-154.3608	465.1	-0.332	| | |	0.000
11	27	126.4691	16.3848	94.3071	158.6311	-787.8483	1041	-99.4691	465.2	-0.214	| | |	0.000
12	122	179.3307	20.2569	139.5683	219.0932	-735.2855	1094	-57.3307	465.1	-0.123	| | |	0.000
13	6	72.1525	21.1025	30.7302	113.5748	-842.5374	986.8424	-66.1525	465.0	-0.142	| | |	0.000
14	13	226.8577	28.3253	171.2577	282.4578	-688.5838	1142	-213.8577	464.6	-0.460	| | |	0.000
15	1	323.3666	48.6690	227.8336	418.8996	-595.3653	1242	-322.3666	463.0	-0.696	| *| |	0.003
16	7	115.3148	16.6763	82.5807	148.0489	-799.0229	1030	-108.3148	465.2	-0.233	| | |	0.000
17	0	87.1866	19.0692	49.7554	124.6177	-827.3313	1002	-87.1866	465.1	-0.187	| | |	0.000
18	29	117.7396	16.5780	85.1984	150.2808	-796.5911	1032	-88.7396	465.2	-0.191	| | |	0.000
19	0	208.4289	24.9120	159.5287	257.3290	-706.6302	1123	-208.4289	464.8	-0.448	| | |	0.000
20	11	196.7896	22.9181	151.8034	241.7759	-718.0686	1112	-185.7896	464.9	-0.400	| | |	0.000

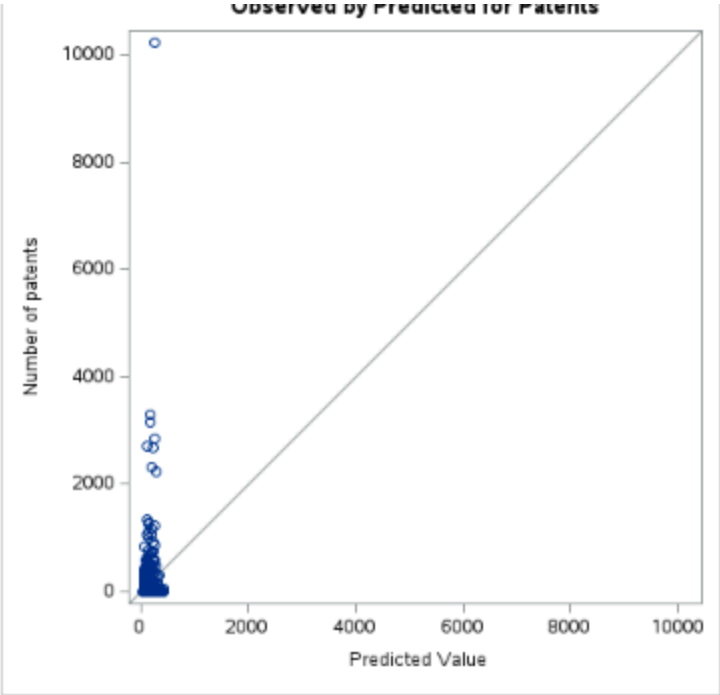
804	277	65.8479	22.0686	22.5292	109.1667	-848.9298	980.6257	211.1521	465.0	0.454				0.000
805	129	66.8178	21.9162	23.7983	109.8374	-847.9458	981.5815	62.1822	465.0	0.134				0.000
806	2	58.5734	23.2513	12.9331	104.2137	-856.3173	973.4640	-56.5734	464.9	-0.122				0.000
807	0	88.1565	18.9532	50.9530	125.3600	-826.3521	1003	-88.1565	465.1	-0.190				0.000
808	3	62.4531	22.6121	18.0676	106.8387	-852.3758	977.2820	-59.4531	465.0	-0.128				0.000

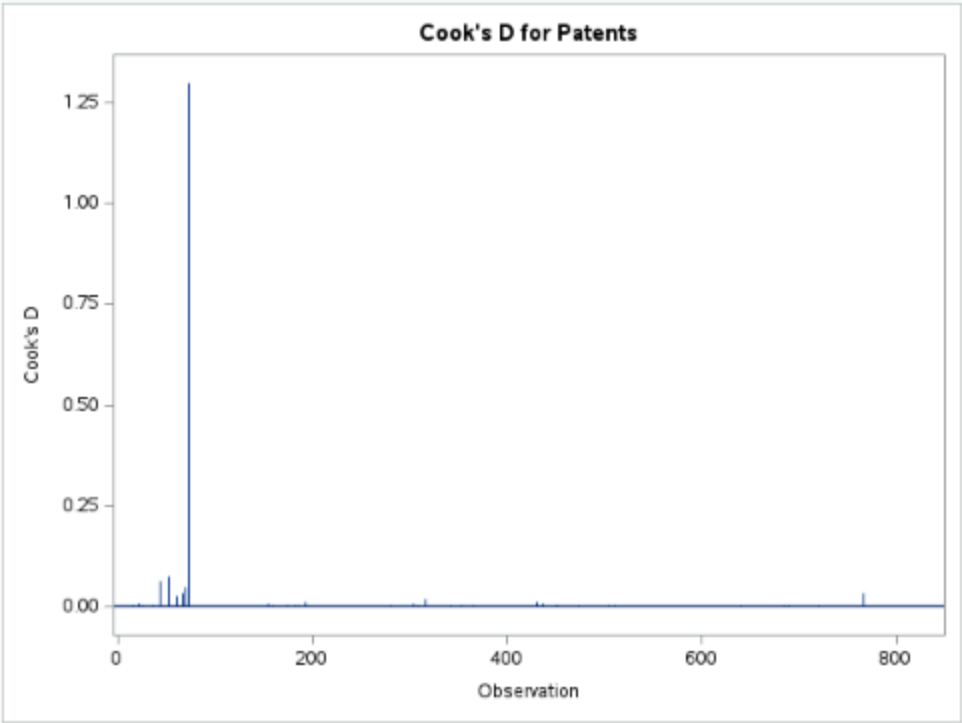
Sum of Residuals	0
Sum of Squared Residuals	174658520
Predicted Residual SS (PRESS)	176241613

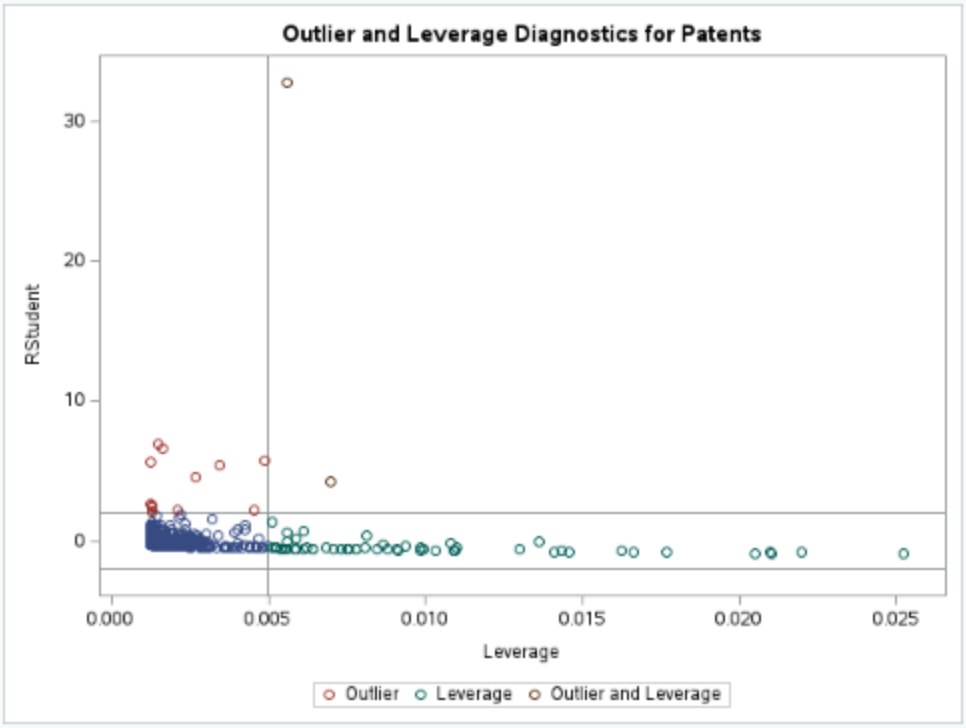


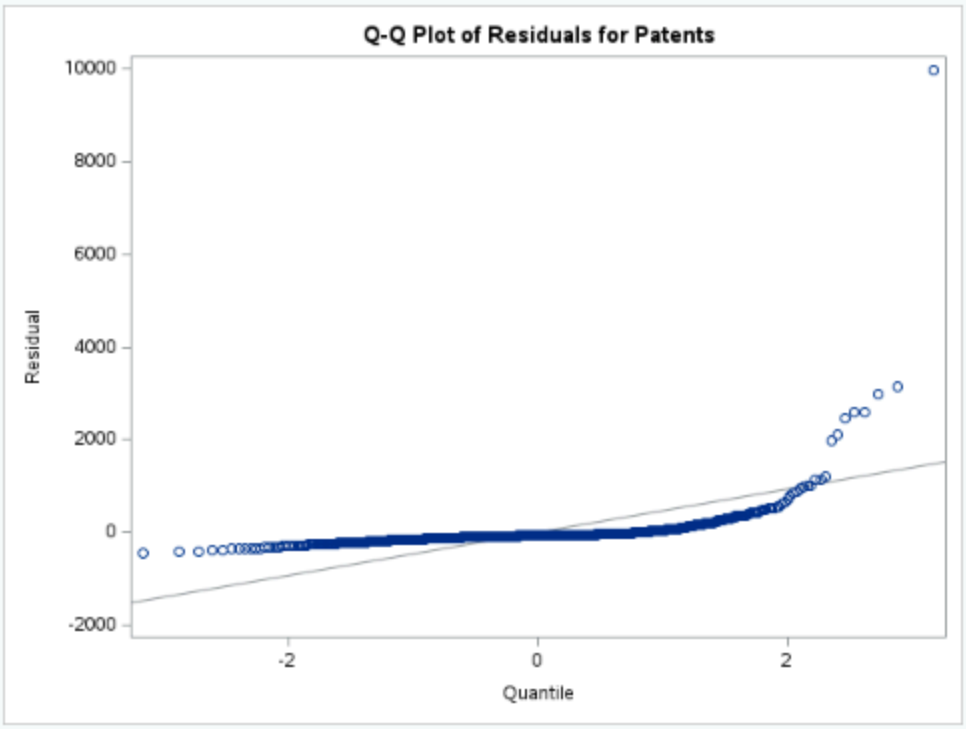

Residual by Predicted for Patents
Residual
10000
8000
6000
4000
2000
0
100
200
300
400
Predicted Value

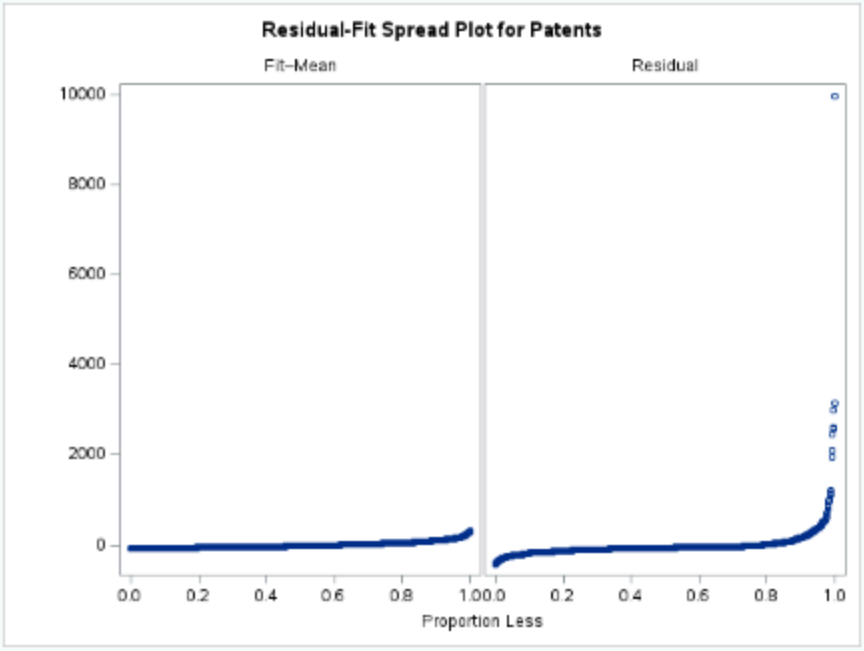

RStudent by Predicted for Patents
RStudent
30
20
10
0
100
200
300
400
Predicted Value

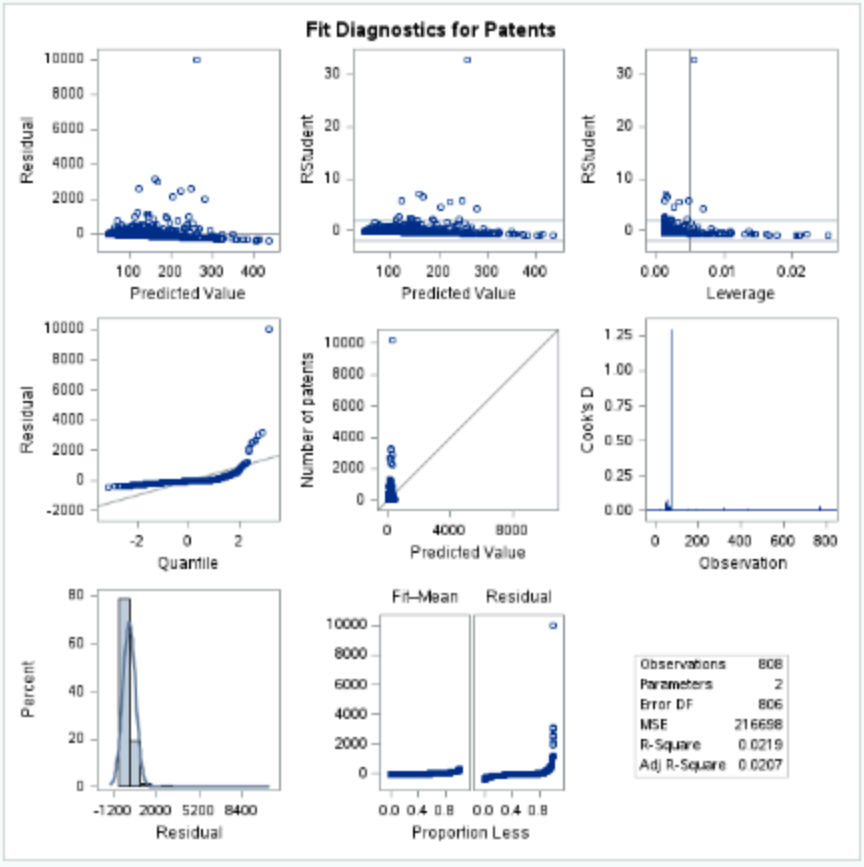

Observed by Predicted for Patents
Number of patents
Predicted Value
0
2000
4000
6000
8000
10000

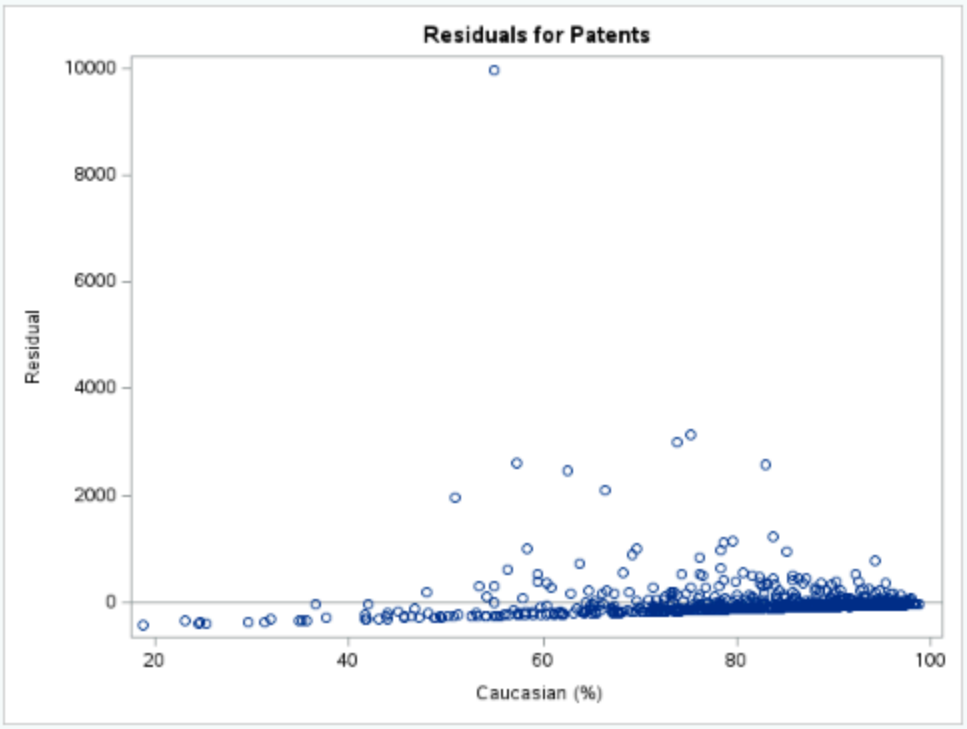

Cook's D for Patents
Cook's D
Observation
0.00
0.25
0.50
0.75
1.00
1.25
0
200
400
600
800

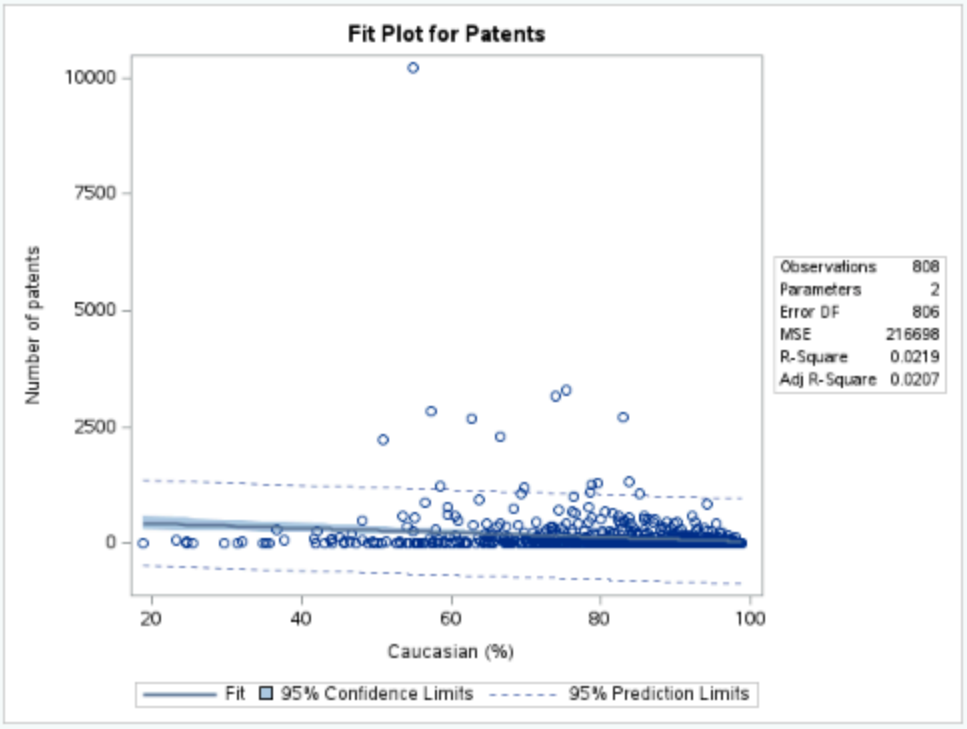

Outlier and Leverage Diagnostics for Patents
RStudent
30
20
10
0
0.000
0.005
0.010
0.015
0.020
0.025
Leverage
Outlier
Leverage
Outlier and Leverage


Q-Q Plot of Residuals for Patents
Residual
10000
8000
6000
4000
2000
0
-2000
-2
0
2
Quantile


Residual-Fit Spread Plot for Patents
Fit–Mean
Residual
10000
8000
6000
4000
2000
0
0.0
0.2
0.4
0.6
0.8
1.0
Proportion Less


Fit Diagnostics for Patents
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Number of patents
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 808
Parameters 2
Error DF 806
MSE 216698
R-Square 0.0219
Adj R-Square 0.0207


Residuals for Patents
Residual
10000
8000
6000
4000
2000
0
20
40
60
80
100
Caucasian (%)


Fit Plot for Patents
Number of patents
10000
7500
5000
2500
0
20
40
60
80
100
Caucasian (%)
Observations 808
Parameters 2
Error DF 806
MSE 216698
R-Square 0.0219
Adj R-Square 0.0207
Fit
95% Confidence Limits
95% Prediction Limits

RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

The UNIVARIATE Procedure
Variable: resid (Residual)

Moments			
N	808	Sum Weights	808
Mean	0	Sum Observations	0
Std Deviation	465.219725	Variance	216429.393
Skewness	13.8780636	Kurtosis	268.810623
Uncorrected SS	174658520	Corrected SS	174658520
Coeff Variation	.	Std Error Mean	16.3663728

Basic Statistical Measures			
Location		Variability	
Mean	0.000	Std Deviation	465.21973
Median	-62.878	Variance	216429
Mode	-221.918	Range	10396
		Interquartile Range	90.66625

Note: The mode displayed is the smallest of 10 modes with a count of 2.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	-232	Pr >= |M|	<.0001
Signed Rank	S	-80213	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.285971	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.303376	Pr > D	<0.0100
Cramer-von Mises	W-Sq	29.99115	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	150.3362	Pr > A-Sq	<0.0050

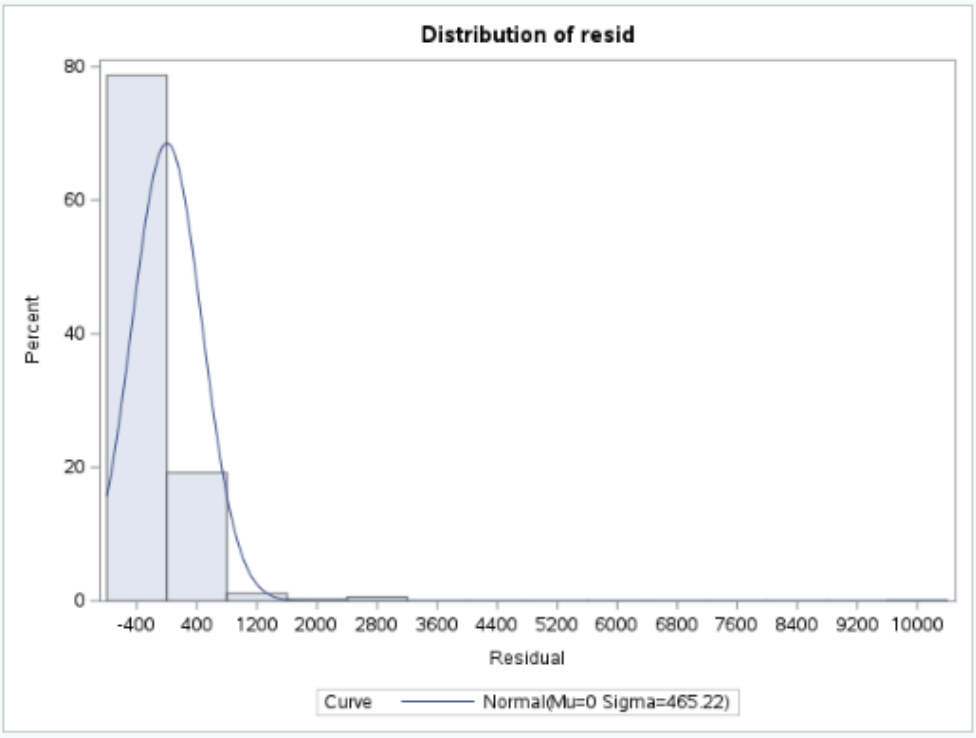
Quantiles (Definition 5)	
Level	Quantile
100% Max	9961.649
99%	1214.805
95%	368.323
90%	162.447
75% Q3	-21.378
50% Median	-62.878
25% Q1	-112.044
10%	-178.699
5%	-236.741
1%	-343.571
0% Min	-433.910

0% Min	-433.910

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-433.910	430	2588.93	316
-402.902	22	2596.29	53
-402.781	154	2989.79	767
-373.803	159	3132.58	66
-373.048	365	9961.65	72

RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

The UNIVARIATE Procedure



RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

The UNIVARIATE Procedure
Fitted Normal Distribution for resid (Residual)

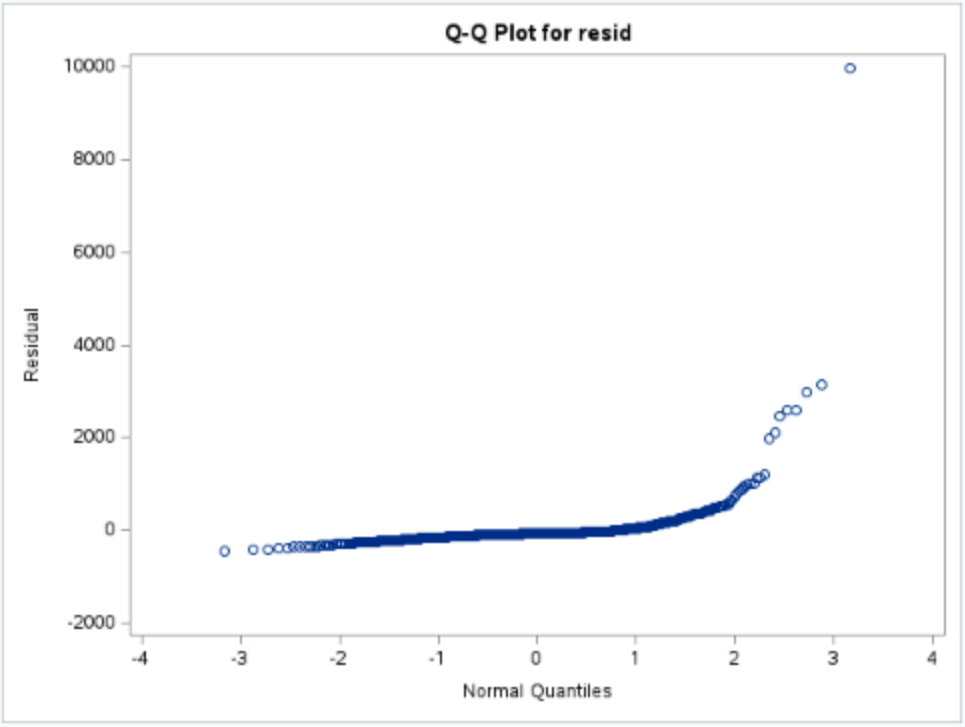
Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	0
Std Dev	Sigma	465.2197

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.303376	Pr > D	<0.010
Cramer-von Mises	W-Sq	29.991148	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	150.336217	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	-343.5709	-1082.26
5.0	-236.7415	-765.22
10.0	-178.6994	-596.20
25.0	-112.0442	-313.79
50.0	-62.8780	0.00
75.0	-21.3780	313.79
90.0	162.4469	596.20
95.0	368.3227	765.22
99.0	1214.8055	1082.26

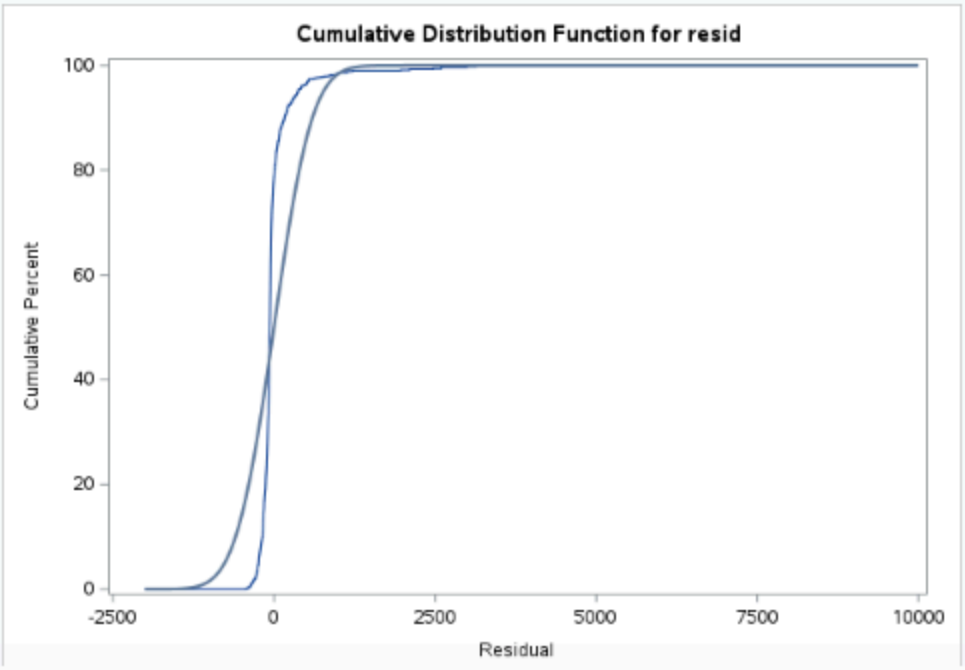
RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

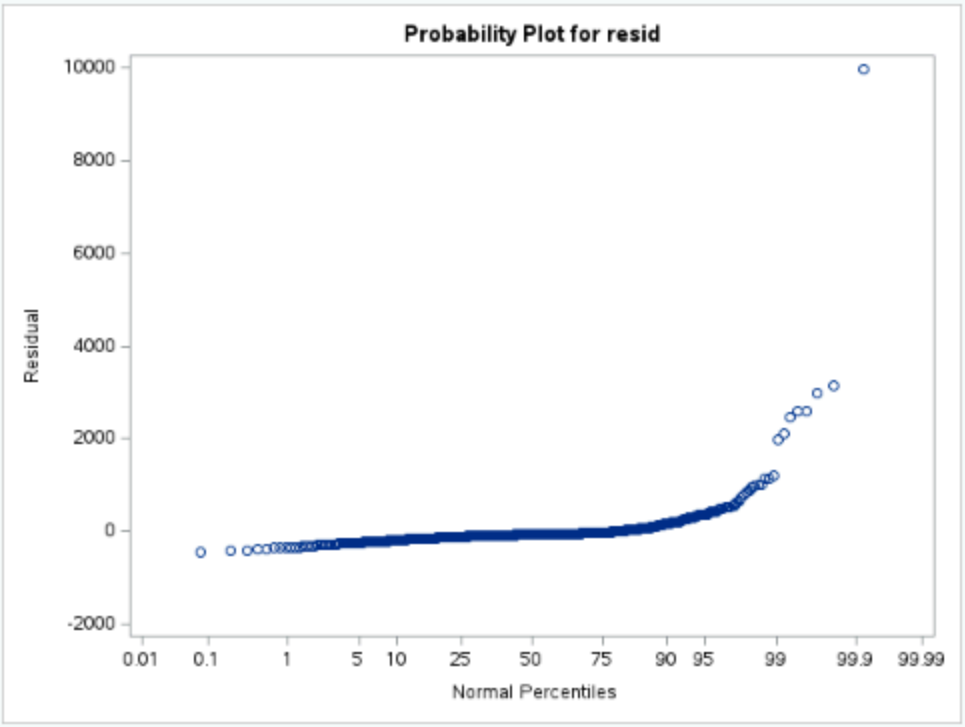
The UNIVARIATE Procedure



RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

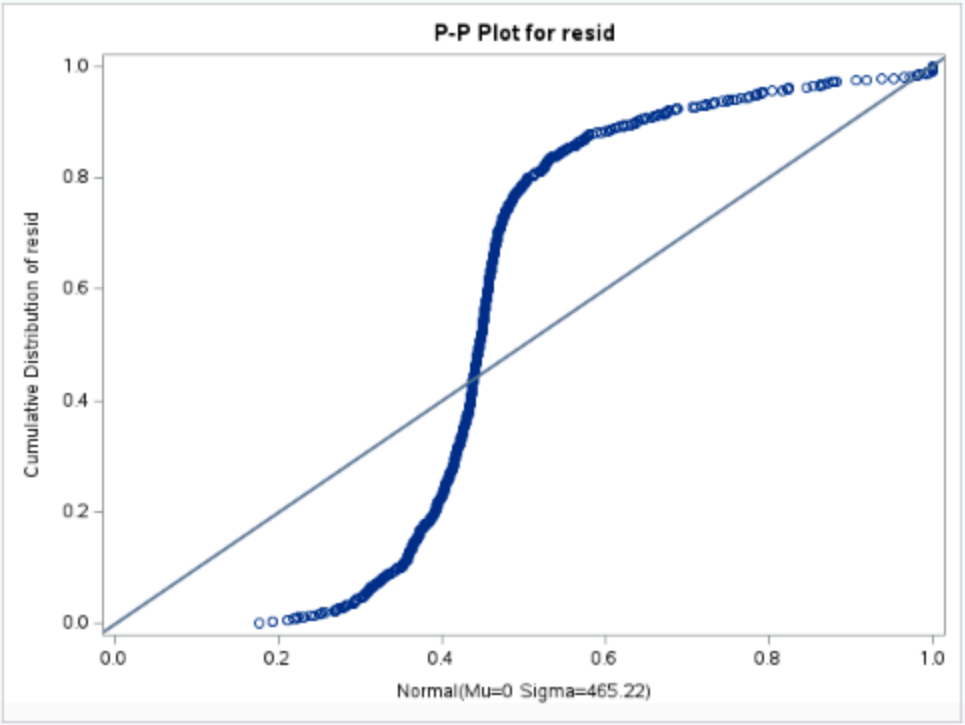
The UNIVARIATE Procedure





RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

The UNIVARIATE Procedure



REGRESSION ANALYSES FOR Patents AND Income AS REGRESSOR (PATENTS)

The REG Procedure

Number of Observations Read	808
Number of Observations Used	808

Descriptive Statistics

Variable	Sum	Mean	Uncorrected SS	Variance	Standard Deviation	Label
Intercept	808.00000	1.00000	808.00000	0	0	Intercept
Income	41160227	50941	2.237129E12	173966926	13190	Median household income
Patents	103978	128.68564	191949046	221275	470.39830	Number of patents

Uncorrected Sums of Squares and Crossproducts

Variable	Label	Intercept	Income	Patents
Intercept	Intercept	808	41160227	103978
Income	Median household income	41160227	2.2371293E12	6669637441
Patents	Number of patents	103978	6669637441	191949046

Correlation

Variable	Label	Income	Patents
Income	Median household income	1.0000	0.2742
Patents	Number of patents	0.2742	1.0000

REGRESSION ANALYSES FOR Patents AND Income AS REGRESSOR (PATENTS)

The REG Procedure
Model: MODEL1

Model Crossproducts X'X X'Y Y'Y

Variable	Label	Intercept	Income	Patents
Intercept	Intercept	808	41160227	103978
Income	Median household income	41160227	2.2371293E12	6669637441
Patents	Number of patents	103978	6669637441	191949046

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Number of Observations Read	808
Number of Observations Used	808

X'X Inverse, Parameter Estimates, and SSE

Variable	Label	Intercept	Income	Patents
Intercept	Intercept	0.0197214798	-3.628492E-7	-369.472622
Income	Median household income	-3.628492E-7	7.122948E-12	0.0097791462
Patents	Number of patents	-369.472622	0.0097791462	165142710.47

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	13425860	13425860	65.53	<.0001
Error	806	165142710	204892		
Corrected Total	807	178568570			

Root MSE	452.64964	R-Square	0.0752
Dependent Mean	128.68564	Adj R-Sq	0.0740
Coeff Var	351.74836		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent Standard Error	Heteroscedasticity Consistent t Value	Heteroscedasticity Consistent Pr > |t|	Type I SS	Type II SS	Standardized Estimate	Squared Semi-partial Corr Type I	Squared Partial Corr Type I	Squared Semi-partial Corr Type II	Squared Partial Corr Type II	Tolerance	Variance Inflation	95% Confidence Limits	Heteroscedasticity Consistent 95% Confidence Limits
Intercept	Intercept	1	-369.47262	63.56703	-5.81	<.0001	119.58467	-3.09	0.0021	13380476	6921895	0	.	.	.	.	.	0	-494.24909	-244.69616	-604.20676	-134.73848
Income	Median household income	1	0.00978	0.00121	8.09	<.0001	0.00262	3.73	0.0002	13425860	13425860	0.27420	0.07519	0.07519	0.07519	0.07519	1.00000	1.00000	0.00741	0.01215	0.00463	0.01493

Covariance of Estimates

Variable	Label	Intercept	Income
Intercept	Intercept	4040.7675299	-0.074344791
Income	Median household income	-0.074344791	1.4594329E-6

Correlation of Estimates

Variable	Label	Intercept	Income
Intercept	Intercept	1.0000	-0.9681
Income	Median household income	-0.9681	1.0000

Sequential Parameter Estimates

Intercept	Income
128.685644	0
-369.472622	0.009779

REGRESSION ANALYSES FOR Patents AND Income AS REGRESSOR (PATENTS)

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

Heteroscedasticity Consistent Covariance of Estimates			
Variable	Label	Intercept	Income
Intercept	Intercept	14300.493844	-0.313358078
Income	Median household income	-0.313358078	6.8894468E-6

Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
2	9.84	0.0073

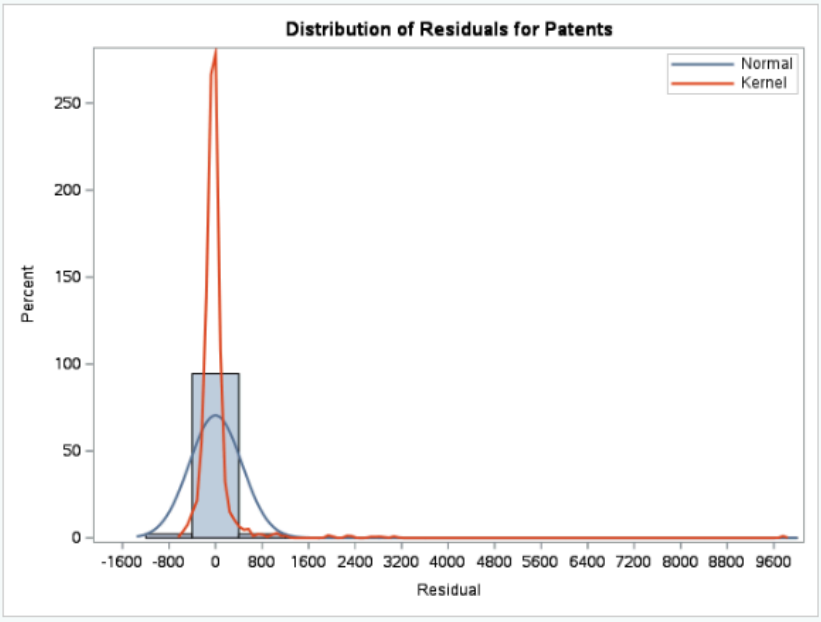
REGRESSION ANALYSES FOR Patents AND Income AS REGRESSOR (PATENTS)

The REG Procedure
Model: MODEL1
Dependent Variable: Patents Number of patents

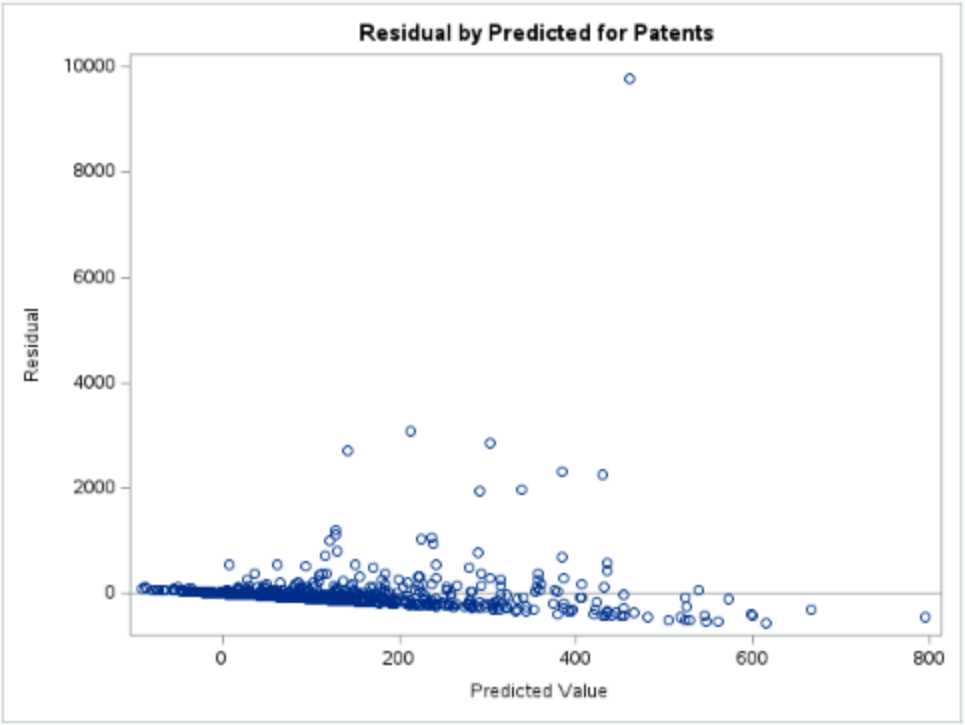
Output Statistics											
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D
1	8	128.2859	15.9242	97.0281	159.5438	-760.7750	1017	-120.2859	452.4	-0.266		0.000
2	1	12.2759	21.4566	-29.8414	54.3932	-877.2330	901.7848	-11.2759	452.1	-0.025		0.000
3	4	22.2213	20.6532	-18.3192	62.7618	-867.2143	911.6569	-18.2213	452.2	-0.040		0.000
4	2	-12.1328	23.5839	-58.4260	34.1603	-901.8492	877.5836	14.1328	452.0	0.031		0.000
5	2	191.8993	17.7359	157.0853	226.7132	-697.2938	1081	-189.8993	452.3	-0.420		0.000
6	2	-43.6999	26.5911	-95.8959	8.4960	-933.7430	846.3431	45.6999	451.9	0.101		0.000
7	3	24.9790	20.4380	-15.1390	65.0970	-864.4375	914.3955	-21.9790	452.2	-0.049		0.000
8	51	41.0168	19.2580	3.2150	78.8186	-848.2982	930.3318	9.9832	452.2	0.022		0.000
9	5	22.8765	20.6018	-17.5630	63.3160	-866.5545	912.3075	-17.8765	452.2	-0.040		0.000
10	24	50.6884	18.6124	14.1540	87.2228	-838.5736	939.9504	-26.6884	452.3	-0.059		0.000
11	27	83.6832	16.8667	50.5754	116.7911	-805.4446	972.8111	-56.6832	452.3	-0.125		0.000
12	122	162.9432	16.4769	130.6005	195.2859	-726.1565	1052	-40.9432	452.3	-0.091		0.000
13	6	14.3100	21.2890	-27.4784	56.0983	-875.1834	903.8034	-8.3100	452.1	-0.018		0.000
14	13	44.8894	18.9931	7.6075	82.1712	-844.4037	934.1824	-31.8894	452.3	-0.071		0.000
15	1	62.6190	17.8939	27.4949	97.7430	-826.5863	951.8242	-61.6190	452.3	-0.136		0.000
16	7	56.8004	18.2329	21.0107	92.5900	-832.4314	946.0321	-49.8004	452.3	-0.110		0.000
17	0	95.0368	16.4578	62.7317	127.3420	-794.0615	984.1352	-95.0368	452.4	-0.210		0.000
18	29	274.1419	24.0096	227.0131	321.2707	-615.6184	1164	-245.1419	452.0	-0.542	*	0.000
19	0	-72.5093	29.5184	-130.4512	-14.5673	-962.9078	817.8892	72.5093	451.7	0.161		0.000
20	11	41.7992	19.2039	4.1037	79.4946	-847.5113	931.1097	-30.7992	452.2	-0.068		0.000

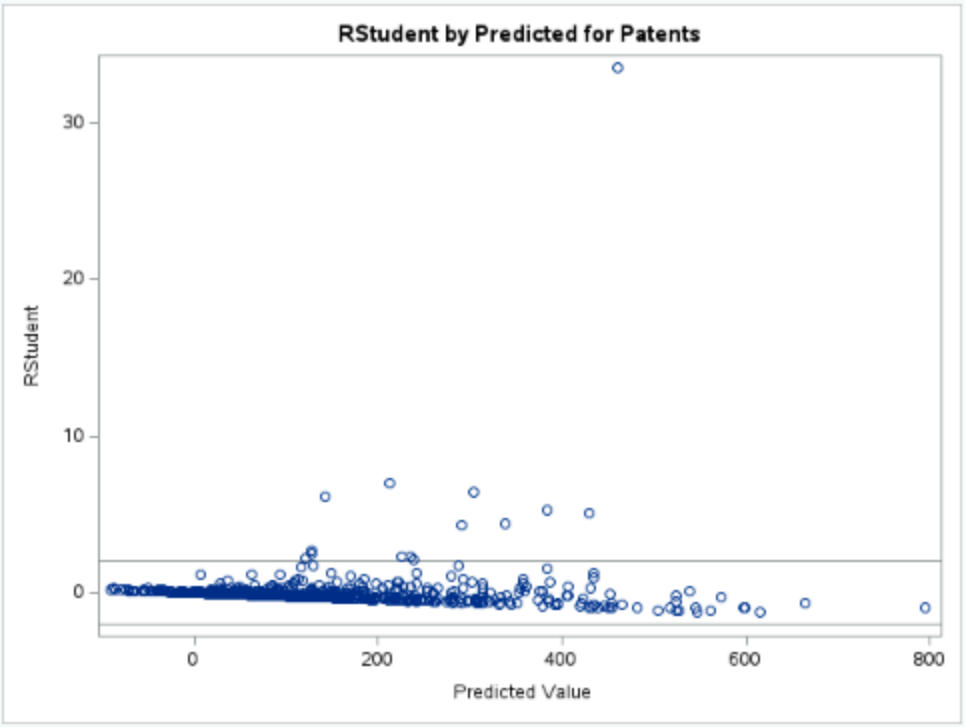
800	57	234.9862	20.6404	194.4710	275.5014	-654.4483	1124	-177.9862	452.2	-0.394	|	|	0.000
801	35	128.6282	15.9242	97.3705	159.8859	-760.4327	1018	-93.6282	452.4	-0.207	|	|	0.000
802	24	141.4878	16.0025	110.0763	172.8993	-747.5785	1031	-117.4878	452.4	-0.260	|	|	0.000
803	50	287.9989	25.3163	238.3053	337.6925	-601.9009	1178	-237.9989	451.9	-0.527	|	*|	0.000
804	277	340.1218	30.5912	280.0739	400.1697	-550.4162	1231	-63.1218	451.6	-0.140	|	|	0.000
805	129	97.6087	16.3804	65.4555	129.7620	-791.4841	986.7016	31.3913	452.4	0.069	|	|	0.000
806	2	79.3022	17.0527	45.8291	112.7752	-809.8394	968.4437	-77.3022	452.3	-0.171	|	|	0.000
807	0	159.1196	16.3620	127.0025	191.2367	-729.9719	1048	-159.1196	452.4	-0.352	|	|	0.000
808	3	144.9398	16.0503	113.4346	176.4450	-744.1298	1034	-141.9398	452.4	-0.314	|	|	0.000

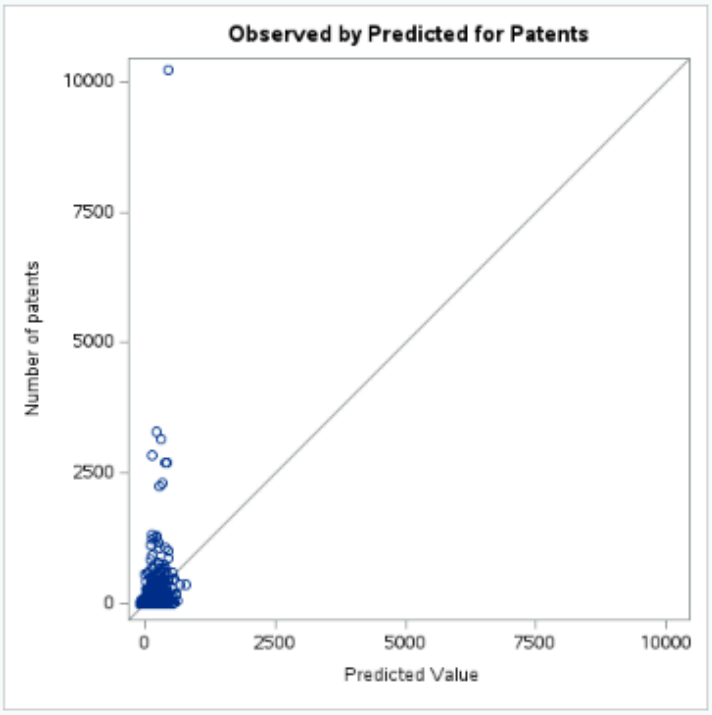
Sum of Residuals	0
Sum of Squared Residuals	165142710
Predicted Residual SS (PRESS)	167517651

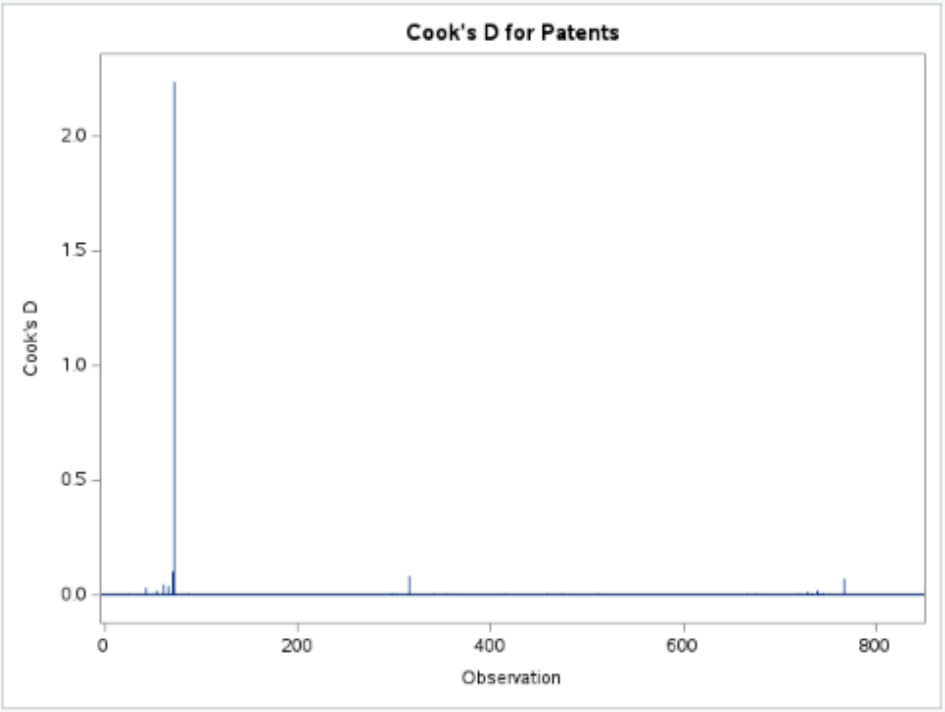


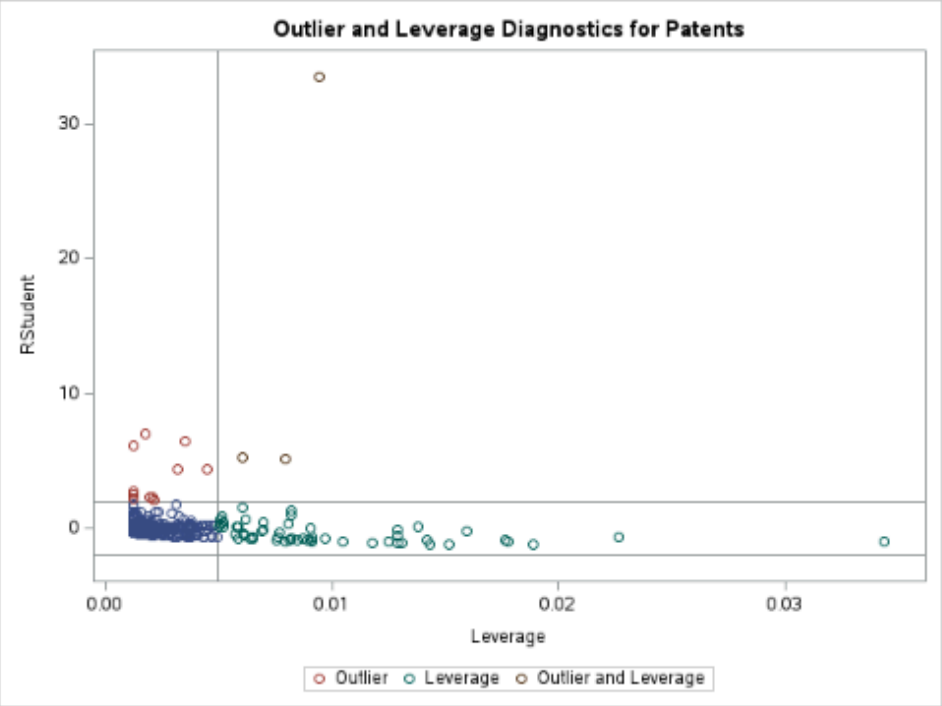
Residual

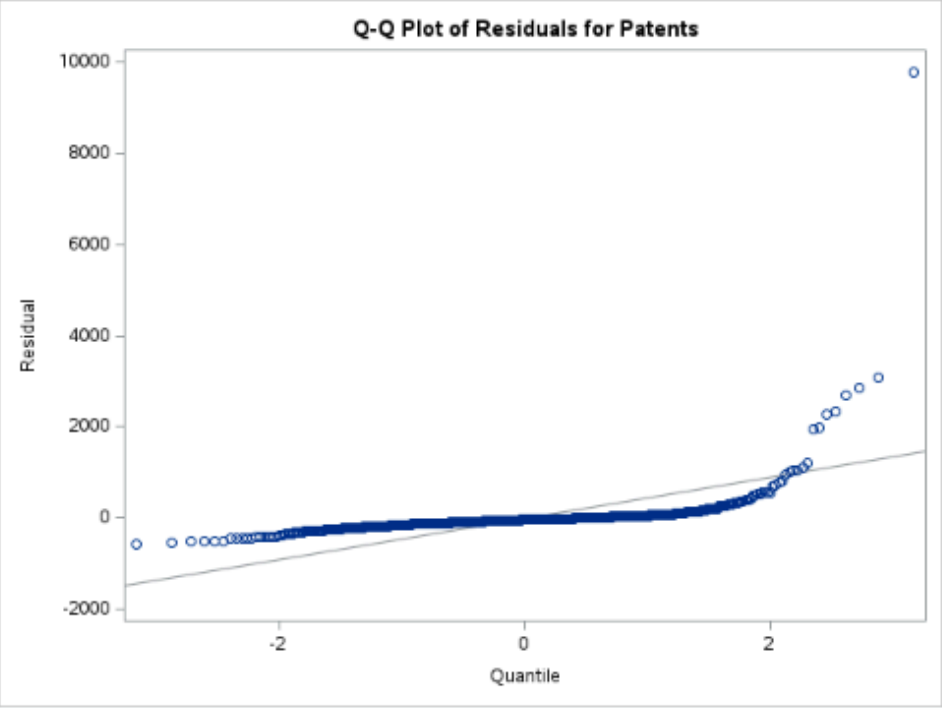

Residual by Predicted for Patents
Residual
10000
8000
6000
4000
2000
0
0
200
400
600
800
Predicted Value

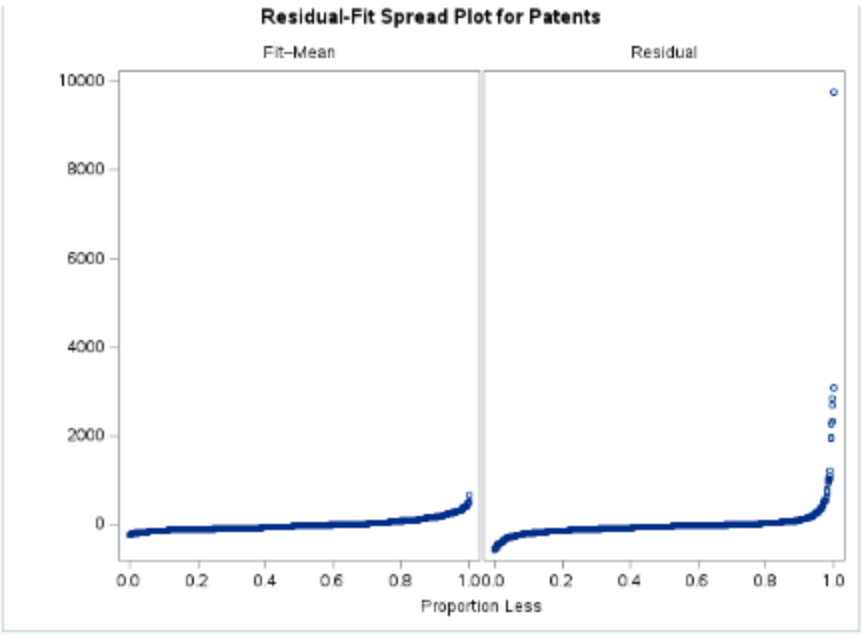

RStudent by Predicted for Patents
RStudent
30
20
10
0
0
200
400
600
800
Predicted Value

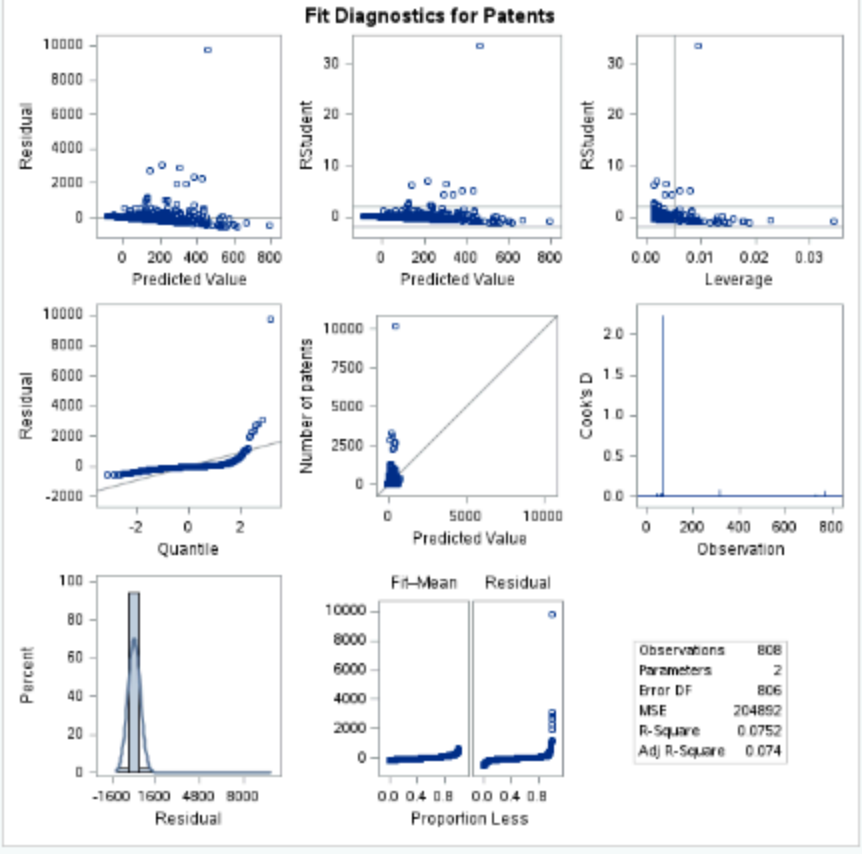

Observed by Predicted for Patents
Number of patents
Predicted Value
0
2500
5000
7500
10000

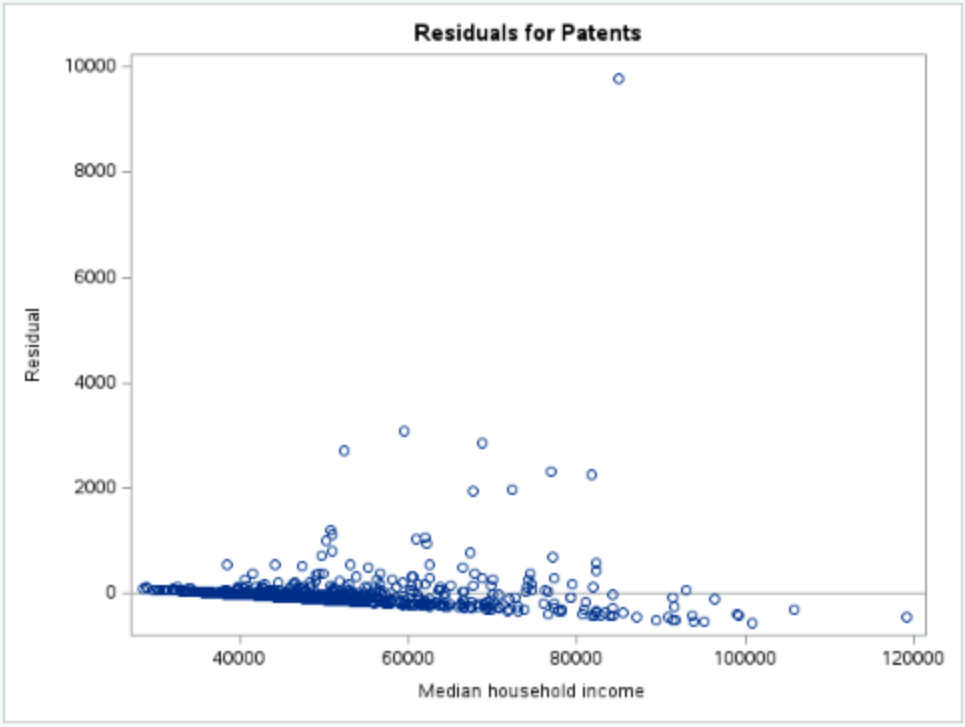

Cook's D for Patents
Cook's D
Observation
0.0
0.5
1.0
1.5
2.0
0
200
400
600
800

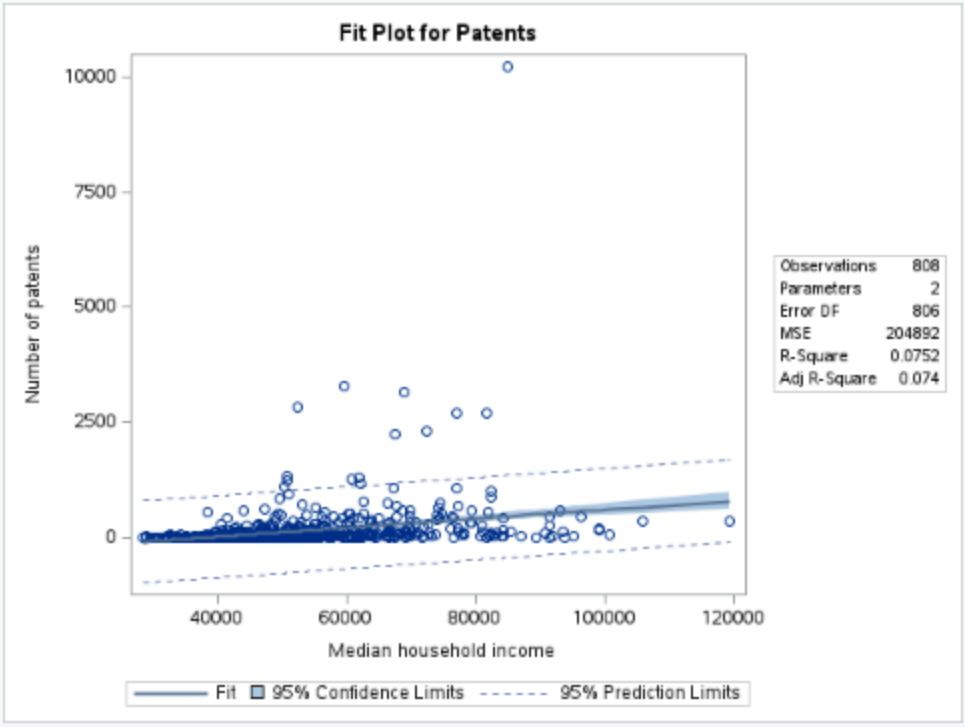

Outlier and Leverage Diagnostics for Patents
RStudent
30
20
10
0
0.00
0.01
0.02
0.03
Leverage
Outlier
Leverage
Outlier and Leverage


Q-Q Plot of Residuals for Patents
Residual
10000
8000
6000
4000
2000
0
-2000
-2
0
2
Quantile


Residual-Fit Spread Plot for Patents
Fit–Mean
Residual
10000
8000
6000
4000
2000
0
0.0
0.2
0.4
0.6
0.8
1.0
Proportion Less


Fit Diagnostics for Patents
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Number of patents
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 808
Parameters 2
Error DF 806
MSE 204892
R-Square 0.0752
Adj R-Square 0.074


Residuals for Patents
Residual
10000
8000
6000
4000
2000
0
40000
60000
80000
100000
120000
Median household income


Fit Plot for Patents
Number of patents
10000
7500
5000
2500
0
40000
60000
80000
100000
120000
Median household income
Observations 808
Parameters 2
Error DF 806
MSE 204892
R-Square 0.0752
Adj R-Square 0.074
Fit
95% Confidence Limits
95% Prediction Limits

RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

The UNIVARIATE Procedure
Variable: resid (Residual)

Moments			
N	808	Sum Weights	808
Mean	0	Sum Observations	0
Std Deviation	452.369105	Variance	204637.807
Skewness	14.077598	Kurtosis	276.666235
Uncorrected SS	165142710	Corrected SS	165142710
Coeff Variation	.	Std Error Mean	15.9142896

Basic Statistical Measures			
Location		Variability	
Mean	0.0000	Std Deviation	452.36911
Median	-37.5628	Variance	204638
Mode	-4.0441	Range	10322
		Interquartile Range	133.68471

Note: The mode displayed is the smallest of 2 modes with a count of 2.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	-155	Pr >= |M|	<.0001
Signed Rank	S	-66419	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.299679	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.314827	Pr > D	<0.0100
Cramer-von Mises	W-Sq	27.47168	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	139.7077	Pr > A-Sq	<0.0050

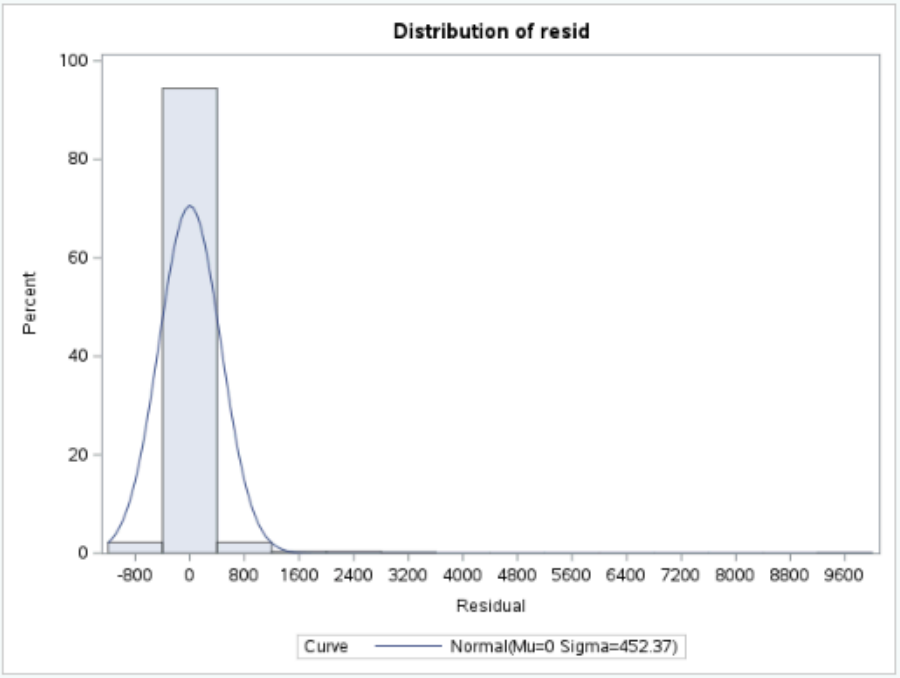
Quantiles (Definition 5)	
Level	Quantile
100% Max	9760.2720
99%	1206.9854
95%	280.2145
90%	118.4127
75% Q3	21.0301
50% Median	-37.5628
25% Q1	-112.6546
10%	-193.7638
5%	-267.4090
1%	-433.8382
0% Min	-561.6297

5%	-267.4090
1%	-433.8382
0% Min	-561.6297

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-561.630	729	2328.69	316
-535.440	734	2702.22	53
-522.974	741	2853.91	767
-514.833	745	3080.84	66
-507.598	299	9760.27	72

RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

The UNIVARIATE Procedure



RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

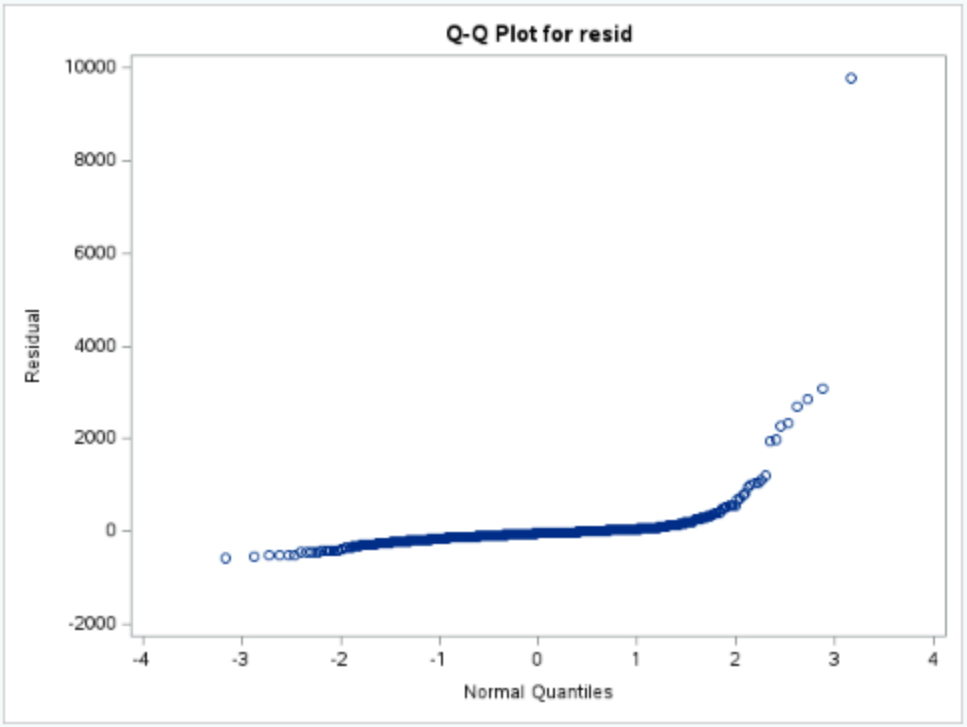
The UNIVARIATE Procedure
Fitted Normal Distribution for resid (Residual)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	0
Std Dev	Sigma	452.3691

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.314827	Pr > D	<0.010
Cramer-von Mises	W-Sq	27.471677	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	139.707682	Pr > A-Sq	<0.005

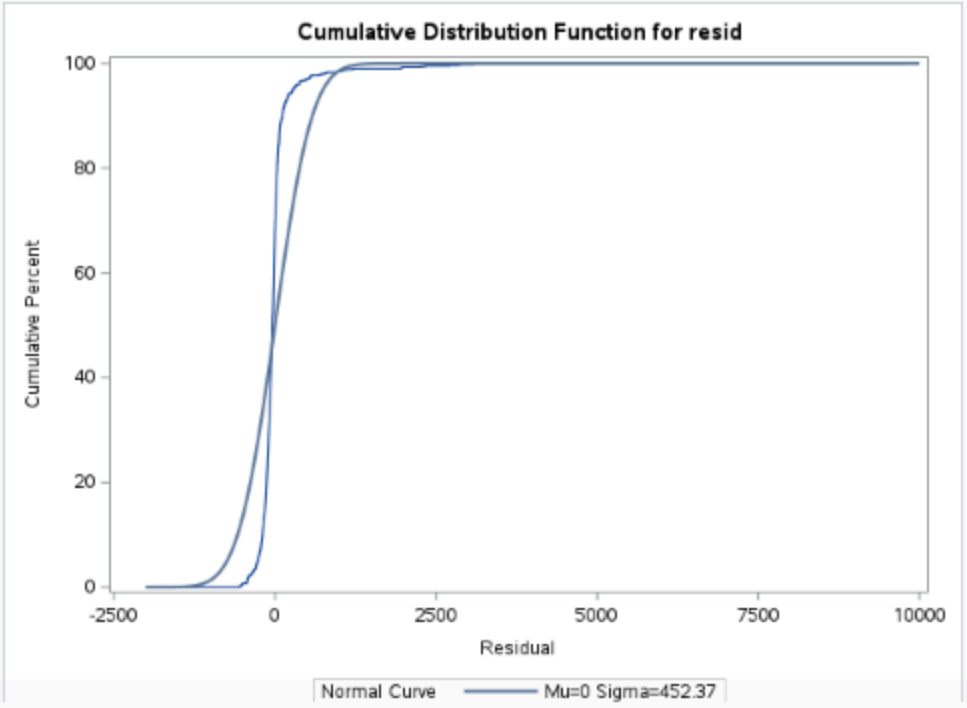
Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	-433.8382	-1052.37
5.0	-267.4090	-744.08
10.0	-193.7638	-579.73
25.0	-112.6546	-305.12
50.0	-37.5628	0.00
75.0	21.0301	305.12
90.0	118.4127	579.73
95.0	280.2145	744.08
99.0	1206.9854	1052.37

The UNIVARIATE Procedure



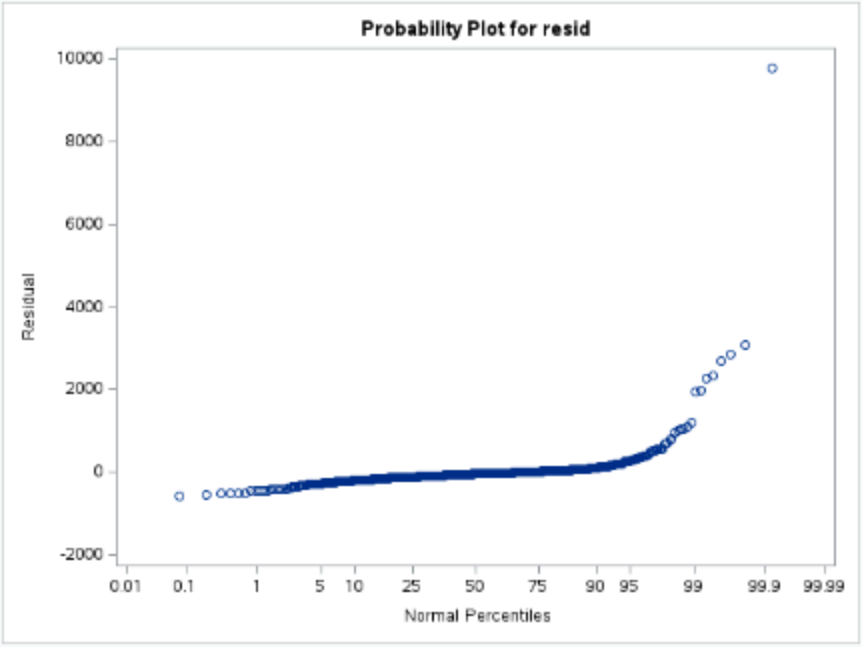
RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

The UNIVARIATE Procedure



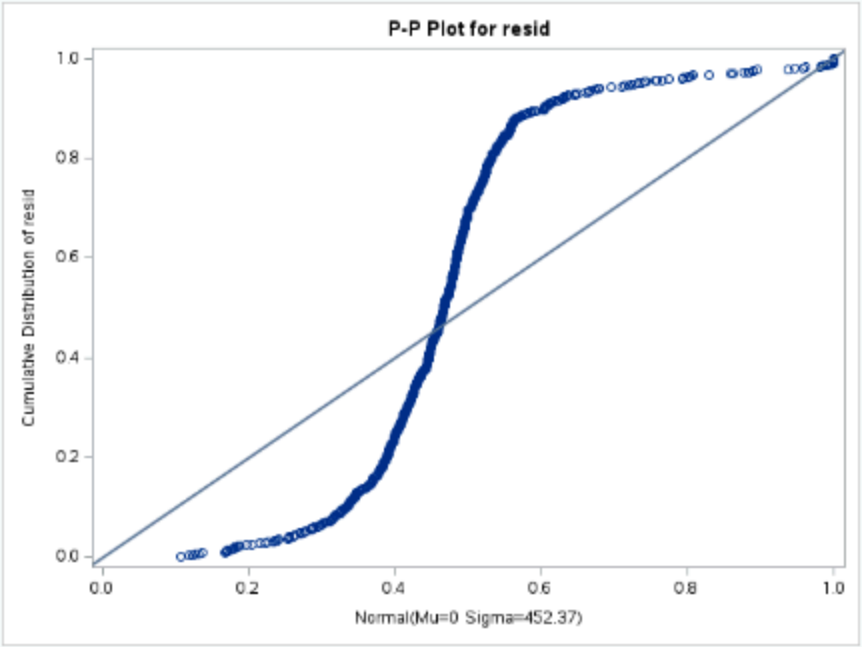
RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

The UNIVARIATE Procedure



RESIDUAL NORMALITY ANALYSES FOR Patents (PATENTS)

The UNIVARIATE Procedure



28H

CODE

```
/* 28H */
ODS PDF FILE = "/home/u62223361/Intro to SAS/HW9/28_ANALYSIS.PDF";
%PATENT_REG(DEP_VAR=Patents, INDEP_VAR=Education,
DATASET_NAME=PATENTS_100)
%PATENT_REG(DEP_VAR=Patents, INDEP_VAR=Education,
DATASET_NAME=PATENTS_LOG10)

ODS PDF CLOSE;
```

LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable
some output features.
69
70         /* 28H */
71         ODS PDF FILE = "/home/u62223361/Intro to
SAS/HW9/28_ANALYSIS.PDF";
NOTE: Writing ODS PDF output to DISK destination
"/home/u62223361/Intro to SAS/HW9/28_ANALYSIS.PDF", printer
"PDF".
72         %PATENT_REG(DEP_VAR=Patents, INDEP_VAR=Education,
DATASET_NAME=PATENTS_100)
NOTE: The data set WORK.PATENTS_RESID has 181 observations and
15 variables.
NOTE: PROCEDURE REG used (Total process time):
      real time           1.54 seconds
      user cpu time       1.03 seconds
      system cpu time     0.09 seconds
```

```
      memory                   30027.78k
      OS Memory                56296.00k
      Timestamp                05/02/2024 04:52:07 AM
      Step Count                        291  Switch Count  52
      Page Faults                       0
      Page Reclaims                     33445
      Page Swaps                        0
      Voluntary Context Switches        2864
      Involuntary Context Switches      4
      Block Input Operations            0
      Block Output Operations           4784

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time                0.47 seconds
      user cpu time            0.31 seconds
      system cpu time          0.02 seconds
      memory                   25481.09k
      OS Memory                73988.00k
      Timestamp                05/02/2024 04:52:07 AM
      Step Count                        292  Switch Count  1
      Page Faults                       0
      Page Reclaims                     8475
      Page Swaps                        0
      Voluntary Context Switches        1438
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1872

73         %PATENT_REG(DEP_VAR=Patents, INDEP_VAR=Education,
DATASET_NAME=PATENTS_LOG10)
NOTE: The data set WORK.PATENTS_RESID has 181 observations and
15 variables.
```

```
NOTE: PROCEDURE REG used (Total process time):
      real time           1.39 seconds
      user cpu time       0.90 seconds
      system cpu time     0.08 seconds
      memory              13934.96k
      OS Memory           58856.00k
      Timestamp           05/02/2024 04:52:09 AM
      Step Count                        293  Switch Count  52
      Page Faults                       0
      Page Reclaims                     28936
      Page Swaps                        0
      Voluntary Context Switches        2852
      Involuntary Context Switches      22
      Block Input Operations            0
      Block Output Operations           5072

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.50 seconds
      user cpu time       0.31 seconds
      system cpu time     0.03 seconds
      memory              24811.53k
      OS Memory           75524.00k
      Timestamp           05/02/2024 04:52:09 AM
      Step Count                        294  Switch Count  1
      Page Faults                       0
      Page Reclaims                     8353
      Page Swaps                        0
      Voluntary Context Switches        1443
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1864
```

```
74
75          ODS PDF CLOSE;
NOTE: ODS PDF printed 68 pages to /home/u62223361/Intro to
SAS/HW9/28_ANALYSIS.PDF.
76
77
78          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
88
```

RESULTS

INCLUDED IN SEPARATE PDF;