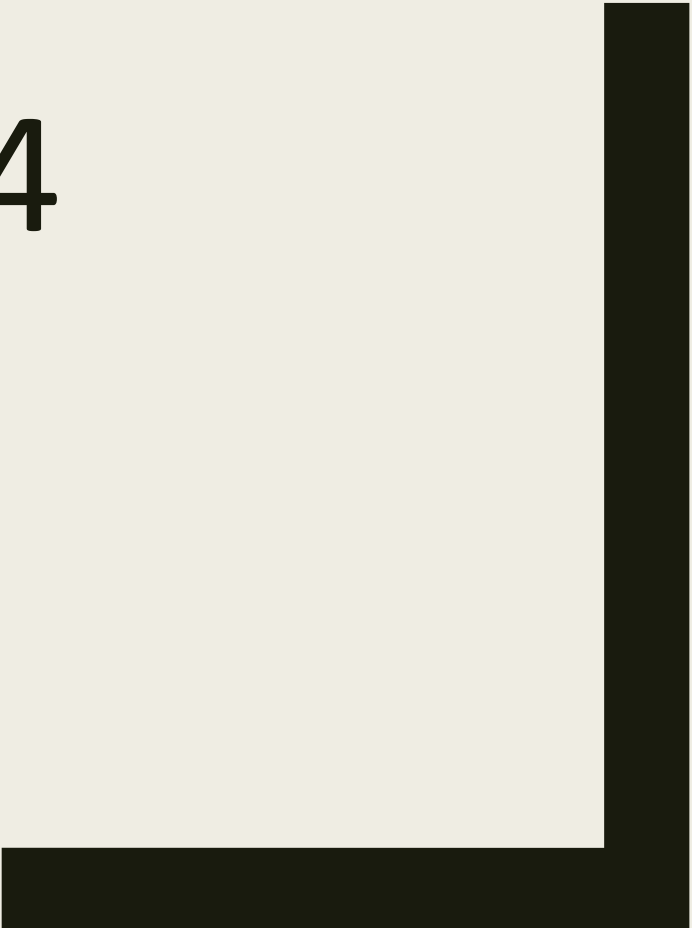




WORKSHOP 4

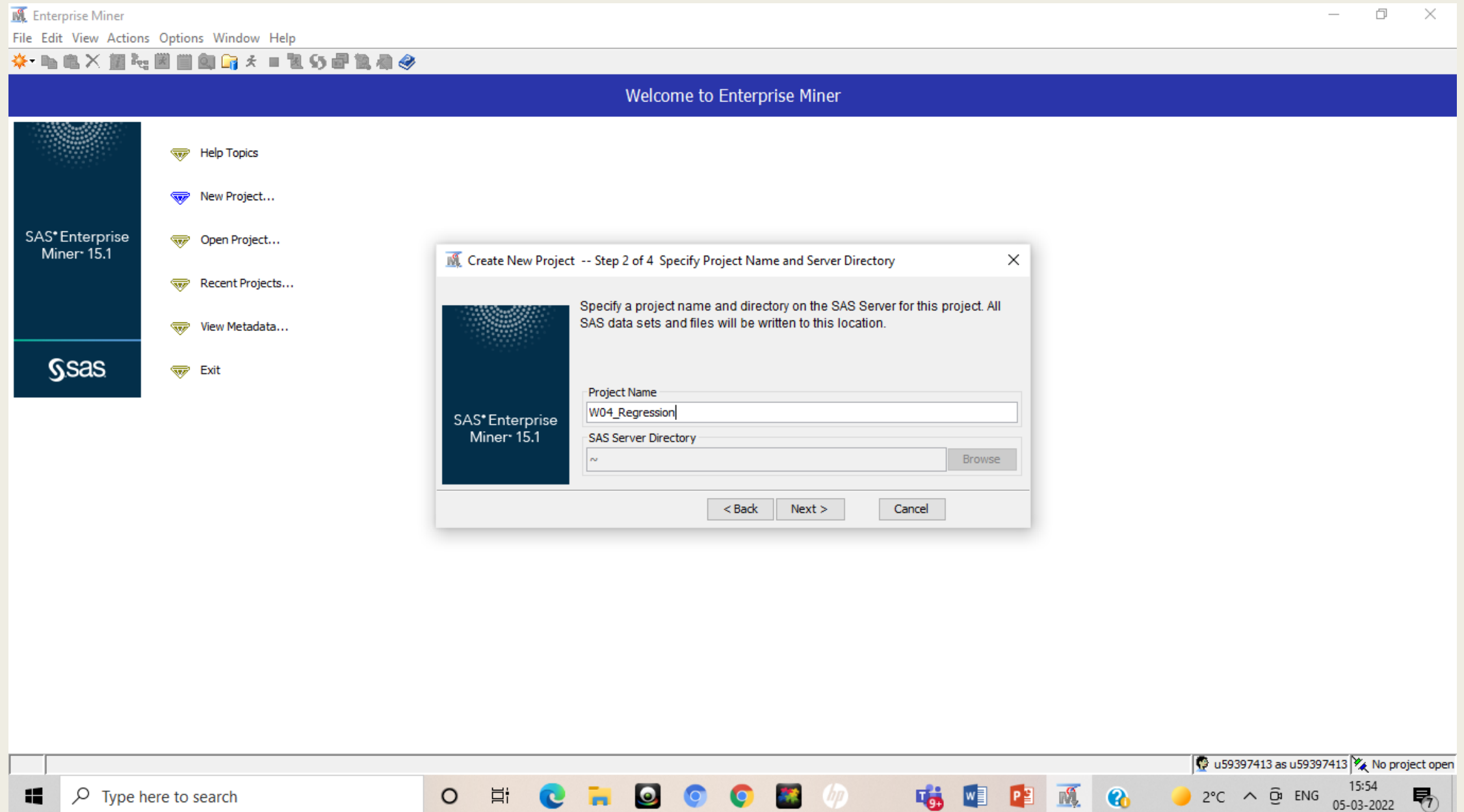
By Group 5-
Anand Mohan Thakur
Josh Shaji
Poonam Bhaliyan
Poornima Singh
Prateek Ramjanam Singh





PART I: LINEAR REGRESSION

Answer 1: Start a new project named *W04_Regression*



Answer 2: Create a diagram and name it as *Regression*

The screenshot displays the SAS Enterprise Miner interface. The main window is titled 'Enterprise Miner - W04_Regression'. The left sidebar shows a tree view with 'W04_Regression' expanded, containing 'Data Sources', 'Diagrams', and 'Model Packages'. The 'Diagrams' folder is selected, and a 'Regression' diagram is being created. The right pane shows the 'Regression' diagram, which is currently empty. The bottom status bar indicates 'Diagram Regression opened'.

Enterprise Miner - W04_Regression

File Edit View Actions Options Window Help

W04_Regression

- Data Sources
- Diagrams
- Regression
- Model Packages

Property	Value
ID	EMWS1
Name	Regression
Status	Open
Notes	
History	
Create Date	5/3/22 8:55 PM
Encoding	utf-8 Unicode (UTF-8)
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA
Native OS	Yes

Regression

Sample Explore Modify Model Assess Utility Credit Scoring HPDM Applications Text Mining Time Series

Diagram Regression

Log

Diagram Regression opened

u59397413 as u59397413 Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)

Type here to search

15:56 05-03-2022

Answer 3: Importing the file W04_2D.xlsx using File Import

The screenshot displays the Enterprise Miner - W04_Regression application window. The interface includes a menu bar (File, Edit, View, Actions, Options, Window, Help), a toolbar, and a left-hand tree view showing the project structure: W04_Regression, Data Sources, Diagrams, Regression (selected), and Model Packages. Below the tree view is a property table for the selected node.

Property	Value
General	
Node ID	FIMPORT
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Import File	C:\Users\Poornima Singh\Desktop\Se...
Maximum Rows to Import	1000000
Maximum Columns to Import	10000
Delimiter	,
Name Row	Yes
Number of Rows to Skip	0
Guessing Rows	500
File Location	Local
File Type	xlsx
Advanced Advisor	No
Rerun	No
Score	
Role	Train
Report	
Summarize	No
Status	
Create Time	5/3/22 8:58 PM
Run ID	
Last Error	
Last Status	
Last Run Time	

The main workspace shows a diagram titled "Regression" with a single "File Import" node. The bottom status bar indicates the user is connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com).

Answer 4: Click on Y. Choose Target as the Role for Y. Then click Explore

Enterprise Miner - W04_Regression

File Edit View Actions Options Window Help

W04_Regression

- Data Sources
- Diagrams
- Regression
- Model Packages

.. Property

- General**
- Node ID
- Imported Data
- Exported Data
- Notes
- Train**
- Variables
- Import File
- Maximum Rows to Import
- Maximum Columns to Import
- Delimiter
- Name Row
- Number of Rows to Skip
- Guessing Rows
- File Location
- File Type
- Advanced Advisor
- Rerun
- Score**
- Role
- Report**
- Summarize
- Status**
- Create Time
- Run ID
- Last Error
- Last Status
- Last Run Time

Variables - FIMPORT

(none) ☐ not Equal to

Apply Reset

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
X	Input	Interval	No		No	.	.
Y	Target	Interval	No		No	.	.

Explore... OK Cancel

Diagram Log

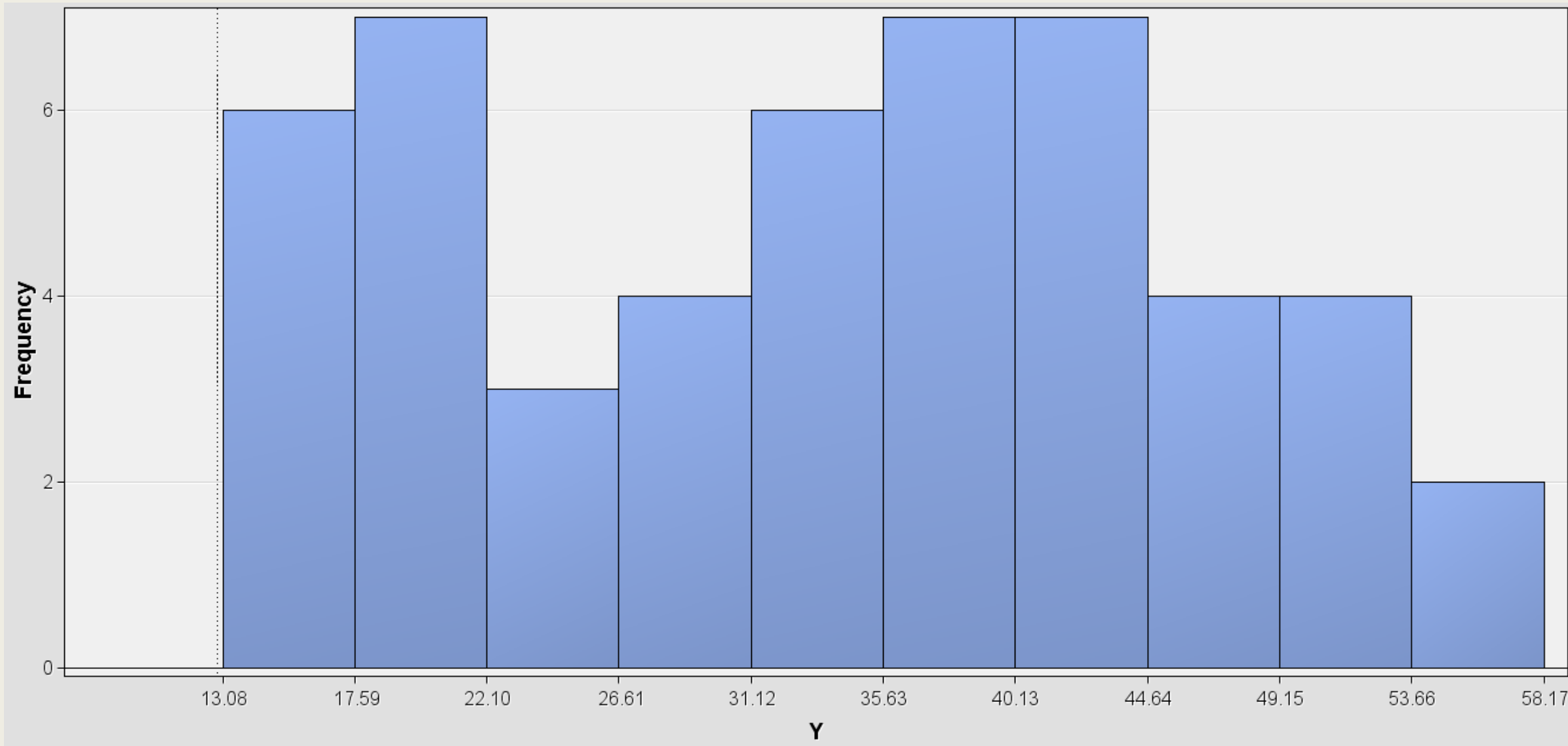
Diagram Regression opened

u59397413 as u59397413 Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)

Type here to search

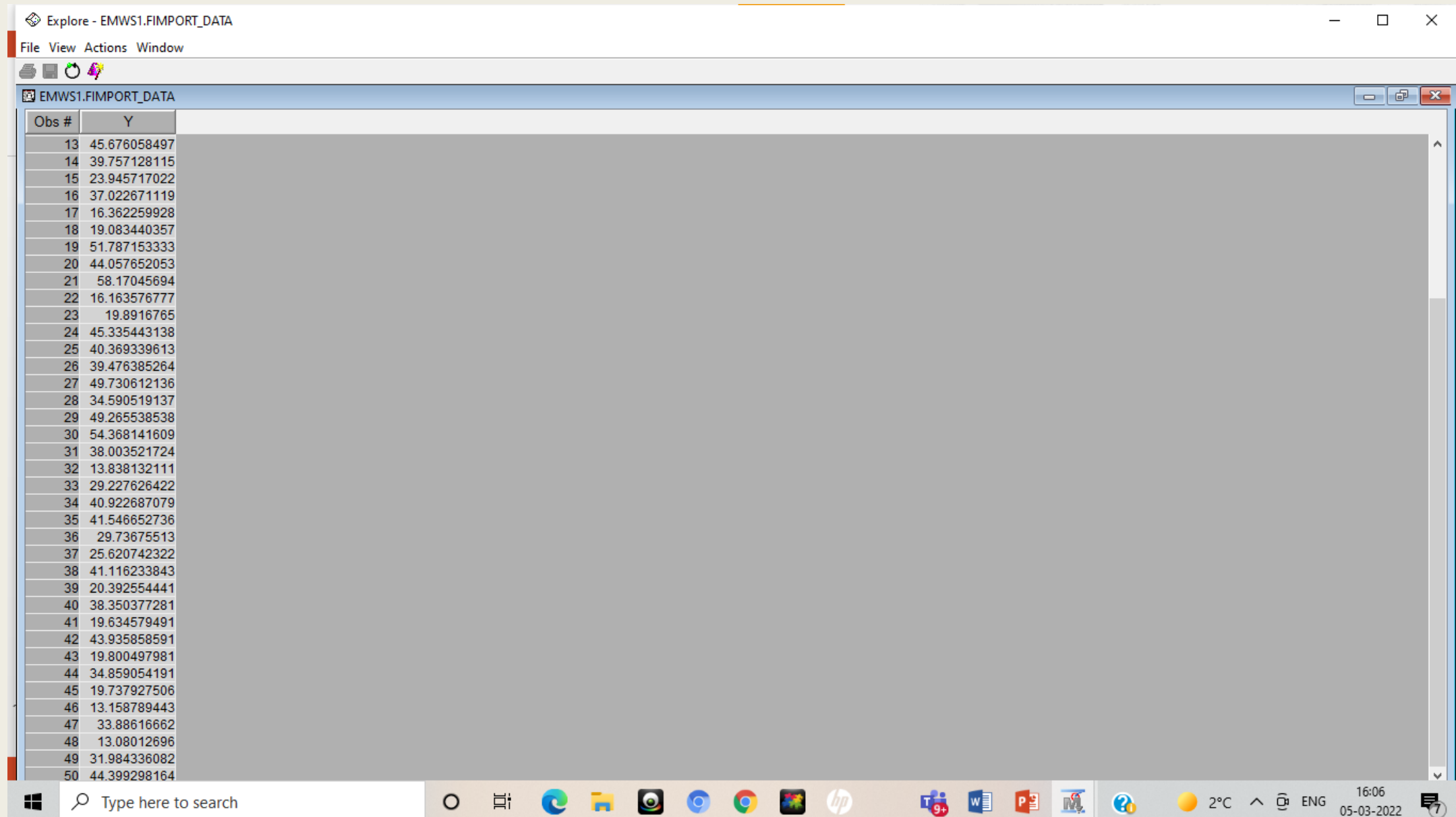
16:02 05-03-2022

Answer 4: Paste the frequency plot of Y here



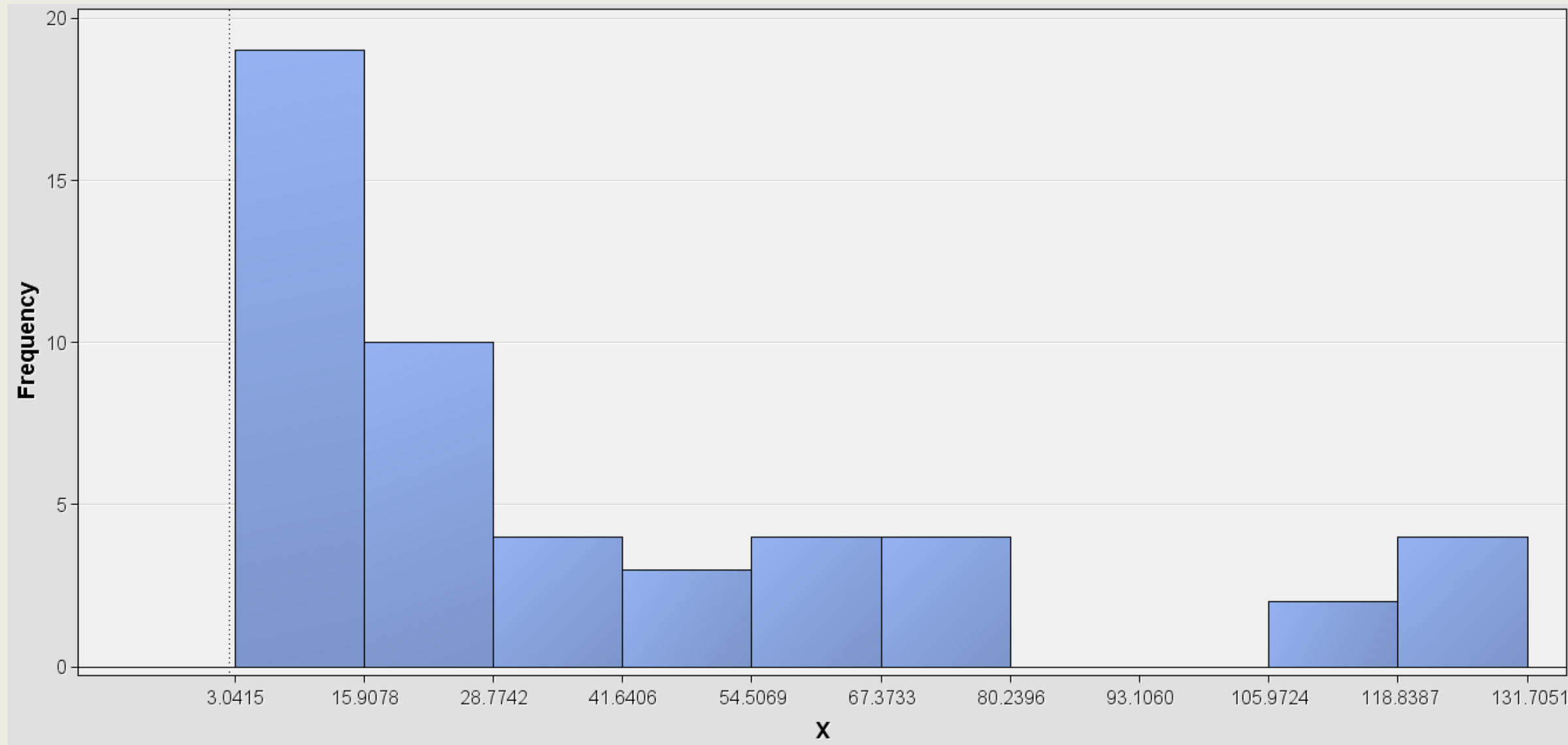
Answer 4: How many rows were imported?

50 rows were imported



Obs #	Y
13	45.676058497
14	39.757128115
15	23.945717022
16	37.022671119
17	16.362259928
18	19.083440357
19	51.787153333
20	44.057652053
21	58.17045694
22	16.163576777
23	19.8916765
24	45.335443138
25	40.369339613
26	39.476385264
27	49.730612136
28	34.590519137
29	49.265538538
30	54.368141609
31	38.003521724
32	13.838132111
33	29.227626422
34	40.922687079
35	41.546652736
36	29.73675513
37	25.620742322
38	41.116233843
39	20.392554441
40	38.350377281
41	19.634579491
42	43.935858591
43	19.800497981
44	34.859054191
45	19.737927506
46	13.158789443
47	33.88616662
48	13.08012696
49	31.984336082
50	44.399298164

Answer 4: Paste the frequency plot of X here



Answer 4: Is the distribution of X suitable for a linear regression model?

The frequency plot of variable X is right skewed which indicates that :

1. There can be a presence of certain amount of outliers that can affect the analysis and would further require complexity to remove the abnormality.
2. There can be an influence on the distribution of the residuals(the difference between estimated value of the regression and the observed dataset) which can lead to non normality.

Hence, it is recommend to transform the skewed variables(left or right) before performing regression analysis in order to eliminate the harmful effects.

Answer 5: Transformation of Variable X

Enterprise Miner - W04_Regression

File Edit View Actions Options Window Help

W04_Regression

- Data Sources
- Diagrams
- Regression
- Model Packages

Property	Value
General	
Node ID	Trans
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Formulas	...
Interactions	...
SAS Code	...
Default Methods	
Interval Inputs	None
Interval Targets	None
Class Inputs	None
Class Targets	None
Treat Missing as Level	No
Sample Properties	
Method	First N
Size	Default
Random Seed	12345
Optimal Binning	
Number of Bins	4
Missing Values	Use in Search
Grouping Method	
Cutoff Value	0.1
Group Missing	No
Number of Bins	Variables
Add Minimum Value to Offset Value	Yes

Regression

Sample Explore Modify Model Assess Utility Credit Scoring HPDM Applications Text Mining Time Series

Diagram

Log

File Import → Transform Variables

Diagram Regression opened

u59397413 as u59397413 Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)

Type here to search

16:25 05-03-2022

Answer 6: Transformation of Variable X and formula selected

The formula used is $\log(X+1)$

Results - Node: Transform Variables Diagram: Regression

File Edit View Window

Transformations Statistics

Source	Method	Variable Name	Formula	Number of Levels	Non Missing	Missing	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis	Label
Input	Original	X			50	0	3.041483	131.7051	37.79881	37.77642	1.219186	0.452479	X
Output	Computed	LOG_X	$\log(X + 1)$		50	0	1.396612	4.888129	3.14505	1.080949	-0.03632	-1.20813	Transformed: X

Output

```
19
20
21
22 Computed Transformations
23 (maximum 500 observations printed)
24
25 Input      Input
26 Name      Role   Level   Name   Level   Formula
27
28   X        INPUT  INTERVAL  LOG_X  INTERVAL  log(X + 1)
29
30
31 *-----*
32 * Score Output
33 *-----*
34
35
```

Windows taskbar: 16:33 05-03-2022

Answer 7: Explore of the Transformed of Variable X

Explore - EMWS1.Trans_TRAIN

File View Actions Window

Sample Properties

Property	Value
Rows	Unknown
Columns	3
Library	EMWS1
Member	TRANS_TRAIN
Type	VIEW
Sample Method	Top
Fetch Size	Default
Fetch Rows	50
Random Seed	12345

Sample Statistics

Obs #	Variable ...	Label	Type	Percent ...	Minimum	Maximum	Mean
1	LOG_X	Transforme...	VAR	0	1.396612	4.888129	3.14505
2	X	X	VAR	0	3.041483	131.7051	37.79881
3	Y	Y	VAR	0	13.08013	58.17046	33.70918

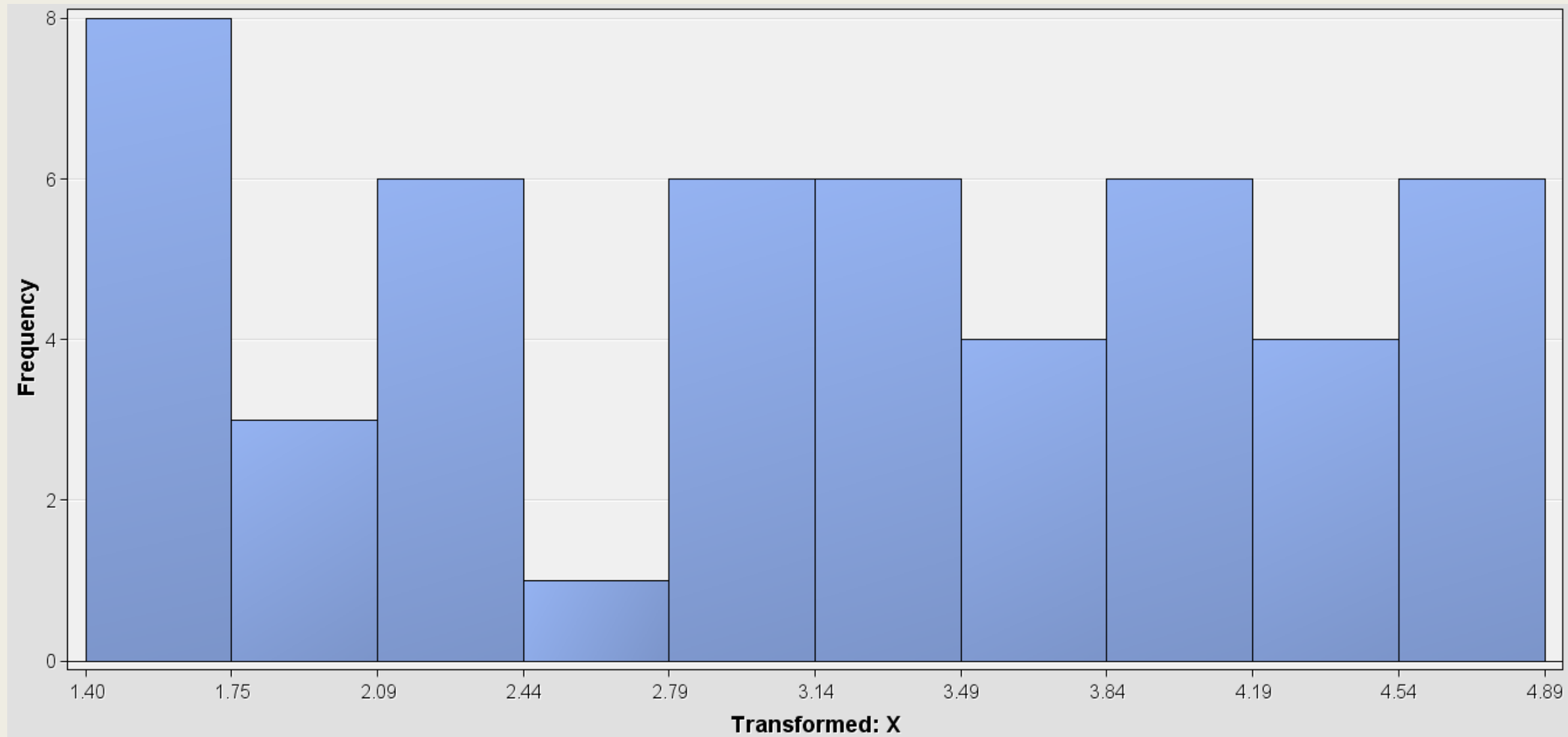
EMWS1.Trans_TRAIN

Obs #	X	Y	Transformed: X
1	60.071379228	47.151987124	4.112043
2	17.361285444	35.147580909	2.910244
3	5.1469832412	18.789461049	1.815961
4	108.7606069	49.596801284	4.698302
5	23.456424516	29.537318131	3.196893
6	12.067147686	27.621244784	2.570101
7	7.3072527012	24.441240638	2.117129
8	46.56192309	39.736594019	3.862033
9	55.831021054	37.358432516	4.040082
10	75.151396007	48.504850793	4.332723
11	4.3521845867	17.286373673	1.677505
12	22.751543265	32.001368327	3.167648
13	51.770642287	45.676058497	3.965955
14	29.696993276	39.757128115	3.424165
15	9.7735850562	23.945717022	2.377097
16	16.768338501	37.022671119	2.877418
17	4.6425191154	16.362259928	1.730331

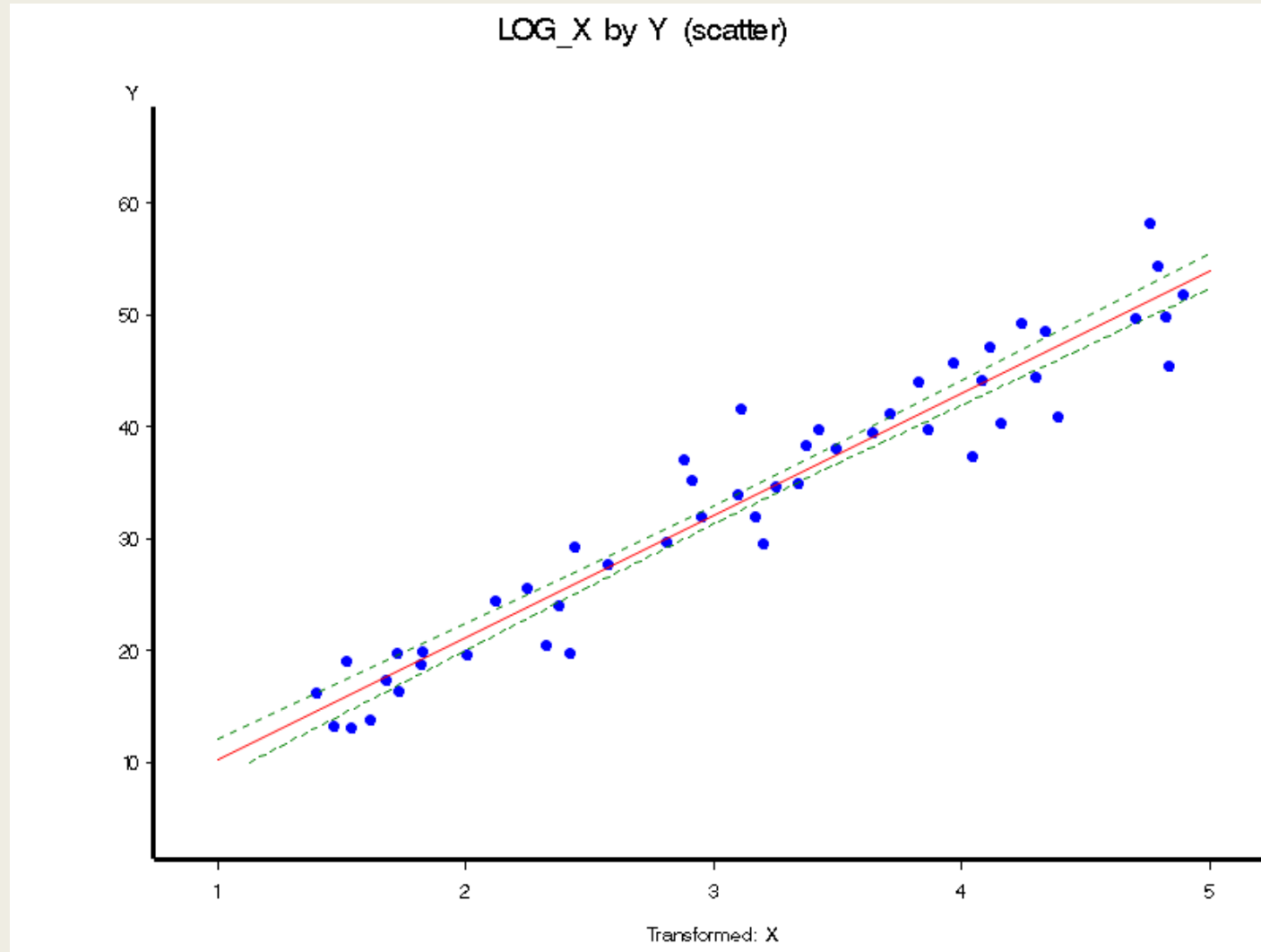
Type here to search

16:41 05-03-2022

Answer 8: Histogram Plot of Transformed X



Answer 10: Scatter Plot between log X and Y



Answer 12: Drag the Regression node from Model tab and connect to the Transform node

The screenshot displays the SAS Enterprise Miner interface. On the left, the 'W04_Regression' project tree shows the 'Regression' node selected. Below it, the 'Property' table lists various settings for the Regression node.

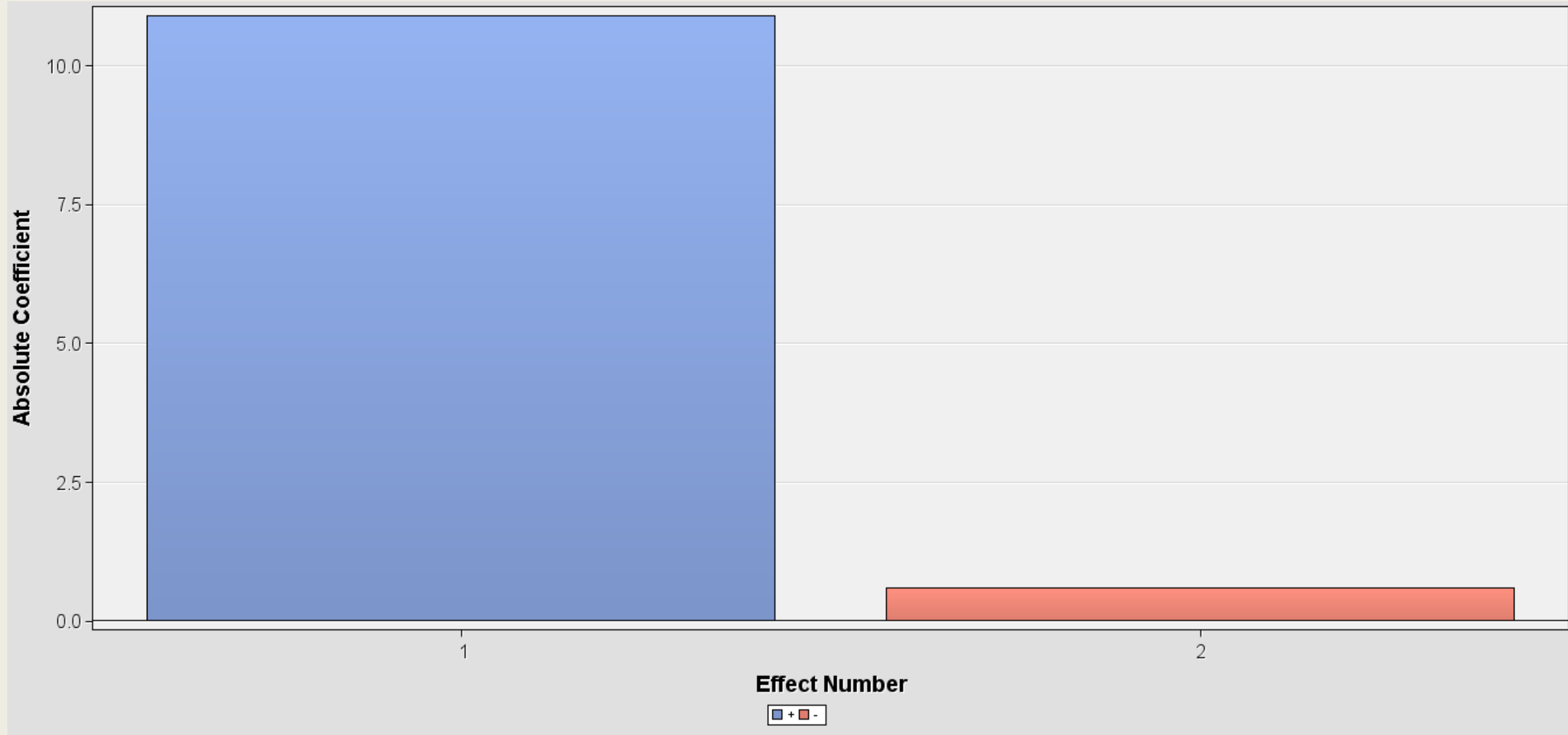
Property	Value
General	
Node ID	Reg
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Equation	
Main Effects	Yes
Two-Factor Interactions	No
Polynomial Terms	No
Polynomial Degree	2
User Terms	No
Term Editor	
Class Targets	
Regression Type	Linear Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	None
Selection Criterion	Default
Use Selection Defaults	Yes
Selection Options	
Optimization Options	
Technique	Default
Default Optimization	Yes

The main workspace shows a workflow diagram with the following nodes and connections:

- File Import** (green checkmark) connects to **Transform Variables** (green checkmark).
- Transform Variables** connects to **Regression** (blue box with a regression line icon).
- MultiPlot** (blue box with a chart icon) is connected to the **Transform Variables** node.
- File Import (2)** (blue box with a folder icon) is shown below the main flow.

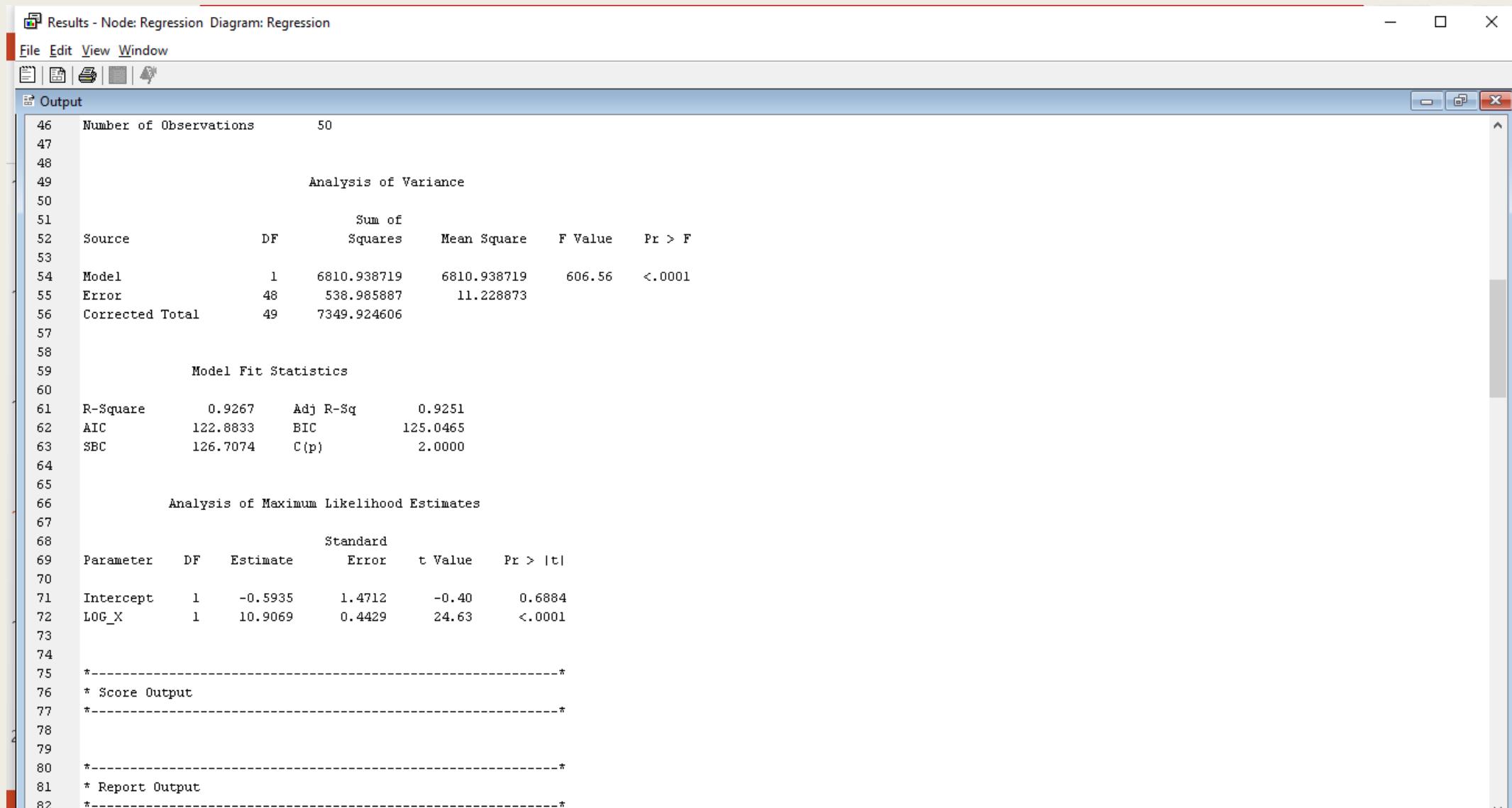
The bottom status bar indicates the user is 'u59397413 as u59397413' and is 'Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)'.

Answer 13: Linear Regression model – Effects Plot



Answer 13: Linear Regression model – Coefficient for log X

The coefficient for log X is 10.9069



The screenshot displays the SAS Results window for a regression analysis. The window title is "Results - Node: Regression Diagram: Regression". The menu bar includes "File", "Edit", "View", and "Window". The toolbar contains icons for file operations and a help icon. The "Output" pane shows the following results:

46 Number of Observations 50

47

48

49 Analysis of Variance

50

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6810.938719	6810.938719	606.56	<.0001
Error	48	538.985887	11.228873		
Corrected Total	49	7349.924606			

57

58

59 Model Fit Statistics

Statistic	Value
R-Square	0.9267
Adj R-Sq	0.9251
AIC	122.8833
BIC	125.0465
SBC	126.7074
C(p)	2.0000

63

64

65

66 Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	-0.5935	1.4712	-0.40	0.6884
LOG_X	1	10.9069	0.4429	24.63	<.0001

73

74

75 *-----*

76 * Score Output

77 *-----*

78

79

80 *-----*

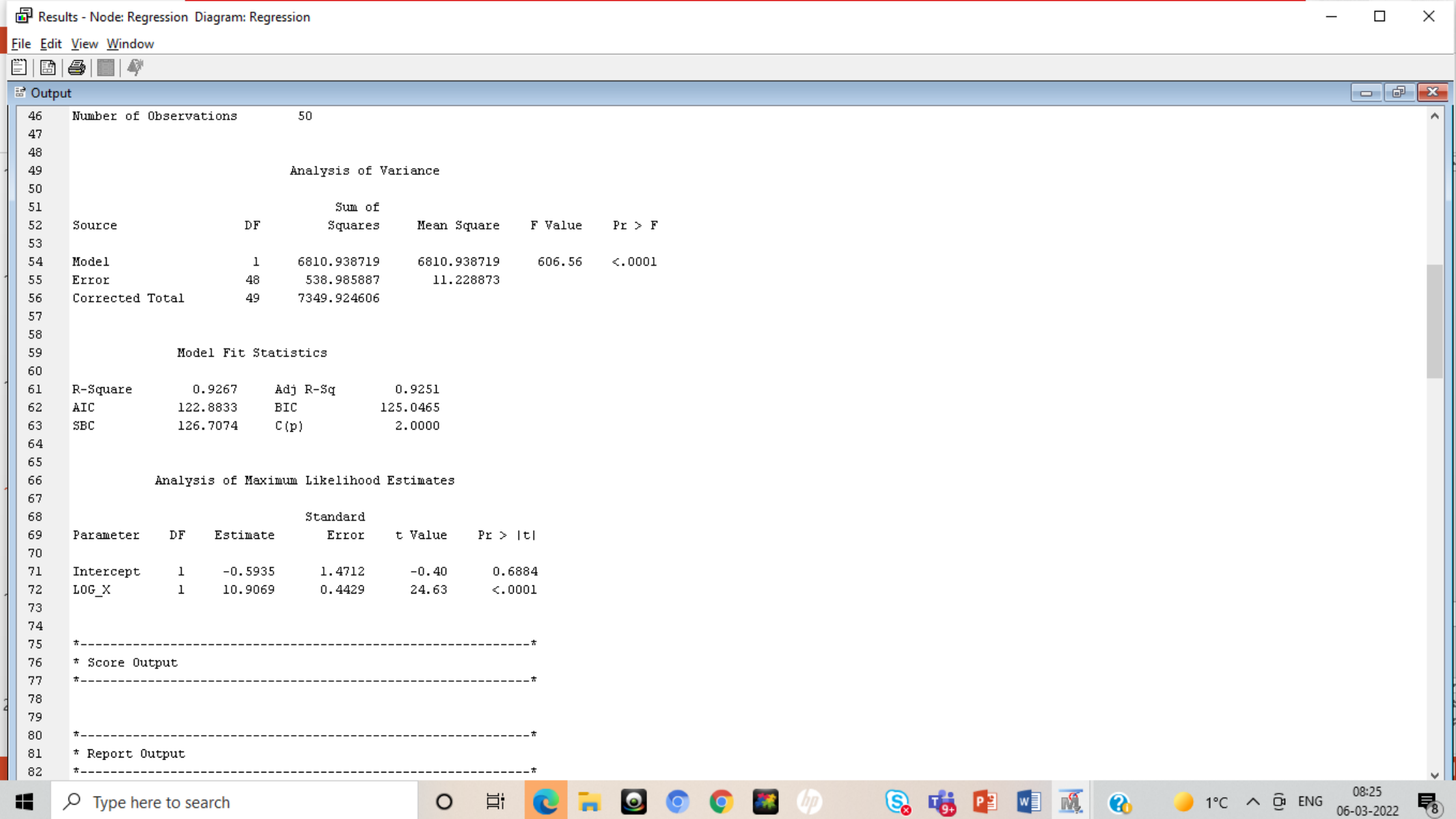
81 * Report Output

82 *-----*

The Windows taskbar at the bottom shows the search bar, system tray with weather (1°C), date (06-03-2022), and time (08:25). Various application icons are visible in the taskbar.

Answer 13: Linear Regression model – intercept

The intercept is -0.5935



The screenshot shows the SAS Results window for a regression model. The window title is "Results - Node: Regression Diagram: Regression". The menu bar includes File, Edit, View, and Window. The toolbar contains icons for saving, printing, and other functions. The main content area is titled "Output" and displays the following statistics:

46 Number of Observations 50

47

48

49 Analysis of Variance

50

51

52 Source DF Sum of Squares Mean Square F Value Pr > F

53

54 Model 1 6810.938719 6810.938719 606.56 <.0001

55 Error 48 538.985887 11.228873

56 Corrected Total 49 7349.924606

57

58

59 Model Fit Statistics

60

61 R-Square 0.9267 Adj R-Sq 0.9251

62 AIC 122.8833 BIC 125.0465

63 SBC 126.7074 C(p) 2.0000

64

65

66 Analysis of Maximum Likelihood Estimates

67

68

69 Parameter DF Estimate Standard Error t Value Pr > |t|

70

71 Intercept 1 -0.5935 1.4712 -0.40 0.6884

72 LOG_X 1 10.9069 0.4429 24.63 <.0001

73

74

75 *-----*

76 * Score Output

77 *-----*

78

79

80 *-----*

81 * Report Output

82 *-----*

The Windows taskbar at the bottom shows the Start button, a search bar, and several application icons. The system tray on the right displays the date and time as 08:25 on 06-03-2022, along with weather information (1°C) and language settings (ENG).

Answer 13: Linear Regression model – Predicted Y for X = 10

The linear equation obtained after running the regression model is:

$$Y = -0.5935 + (10.9069 * \log X)$$

Now for predicting value of Y at X = 10, we have :

$$Y = -0.5935 + (10.9069 * \log 10)$$

$$Y = -0.5935 + (10.9069 * 2.302585)$$

$$Y = 24.5205$$

The predicted value of Y is **24.5205**



PART II: MULTIVARIATE REGRESSION



Answer 15: Importing the file W04_3D.xlsx using File Import

The screenshot displays the Enterprise Miner interface. On the left, a tree view shows the project structure: W04_Regression, Data Sources, Diagrams, Regression, and Model Packages. Below this, a property table for the 'Regression' node is visible.

Property	Value
General	
Node ID	FIMPORT2
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Import File	
Maximum Rows to Import	1000000
Maximum Columns to Import	10000
Delimiter	,
Name Row	Yes
Number of Rows to Skip	0
Guessing Rows	500
File Location	Local
File Type	xlsx
Advanced Advisor	No
Rerun	No
Score	
Role	Train
Report	
Summarize	No
Status	
Create Time	6/3/22 1:37 PM
Run ID	
Last Error	
Last Status	
Last Run Time	

The main workspace shows a workflow diagram with three nodes: 'File Import', 'Transform Variables', and 'Regression'. A 'File Import (2)' node is also visible. A 'File Import' dialog box is open, prompting the user to select the source of the data file. The 'My Computer' option is selected. The file path is 'nima Singh\Desktop\Seneca-2022\Semester 2\BAN 210- Predictive Analytics\Week 8\W04_3D.xlsx'. The 'Browse...' button is highlighted.

Enterprise Miner - W04_Regression

File Edit View Actions Options Window Help

Sample Explore Modify Model Assess Utility Credit Scoring HPDM Applications Text Mining Time Series

Regression

File Import

Select My Computer if the data file you want to import is located on your local machine. Select SAS Server to import a data file located on your SAS workspace server.

☒ My Computer

☐ SAS Servers

nima Singh\Desktop\Seneca-2022\Semester 2\BAN 210- Predictive Analytics\Week 8\W04_3D.xlsx

Browse...

View File Import Types Preview OK Cancel

Diagram Log

Path updated

u59397413 as u59397413 Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)

Type here to search

1°C ENG 08:44 06-03-2022

Answer 16: Click on Y. Choose Target as the Role for Y

Enterprise Miner - W04_Regression

File Edit View Actions Options Window Help

W04_Regression

- Data Sources
- Diagrams
- Regression
- Model Packages

Property

General

- Node ID
- Imported Data
- Exported Data
- Notes

Train

- Variables
- Import File
- Maximum Rows to Import
- Maximum Columns to Import
- Delimiter
- Name Row
- Number of Rows to Skip
- Guessing Rows
- File Location
- File Type
- Advanced Advisor
- Rerun

Score

- Role

Report

- Summarize

Status

- Create Time
- Run ID
- Last Error
- Last Status
- Last Run Time

Variables - FIMPORT2

(none) ☐ not Equal to ☐ ☐ ☐

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
X1	Input	Interval	No		No	.	.
X2	Input	Interval	No		No	.	.
Y	Target	Interval	No		No	.	.

Explore... OK Cancel

Diagram Log

Path updated

u59397413 as u59397413 Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)

Type here to search

1°C 08:46 06-03-2022

Answer 17: Transformation of Variable

Enterprise Miner - W04_Regression

File Edit View Actions Options Window Help

W04_Regression

- Data Sources
- Diagrams
- Regression
- Model Packages

Property Value

General

Node ID Trans2

Imported Data

Exported Data

Notes

Train

Variables

Formulas

Interactions

SAS Code

Default Methods

Interval Inputs None

Interval Targets None

Class Inputs None

Class Targets None

Treat Missing as Level No

Sample Properties

Method First N

Size Default

Random Seed 12345

Optimal Binning

Number of Bins 4

Missing Values Use in S

Grouping Method

Cutoff Value 0.1

Group Missing No

Number of Bins Variable

Add Minimum Value to Offset Value Yes

Diagram Log

Path updated

u59397413 as u59397413 Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)

08:49 06-03-2022

Variables - Trans2

(none) ☐ not Equal to

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

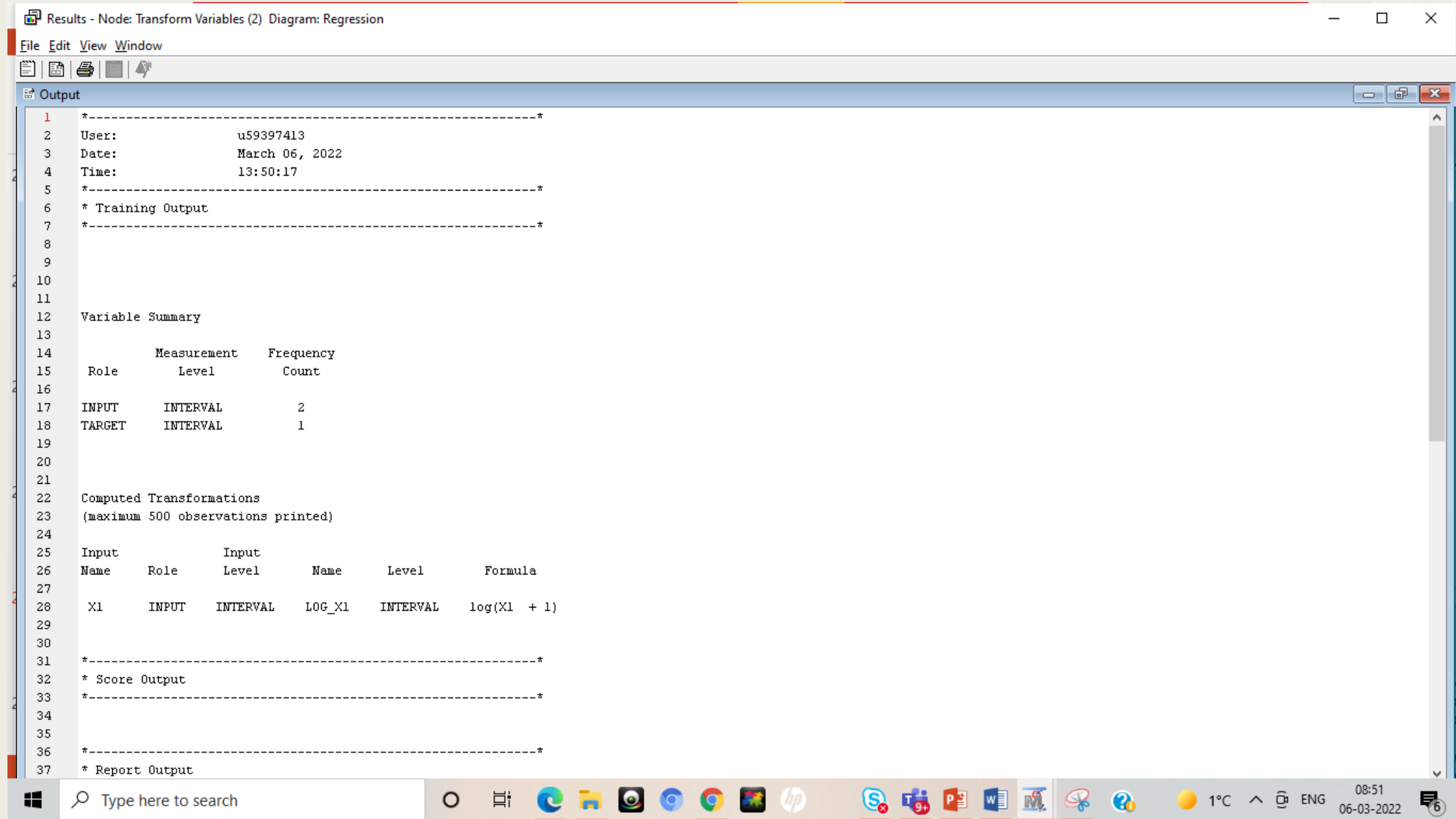
Name	Method	Number of Bins	Role	Level
X1	Log	4	Input	Interval
X2	Max. Normal	4	Input	Interval
Y	Default	4	Target	Interval

Apply Reset

Explore... Update Path OK Cancel

Answer 18: Transformation of Variable X1 and X2 and formula selected

The formula selected is $\log(X1 + 1)$



Results - Node: Transform Variables (2) Diagram: Regression

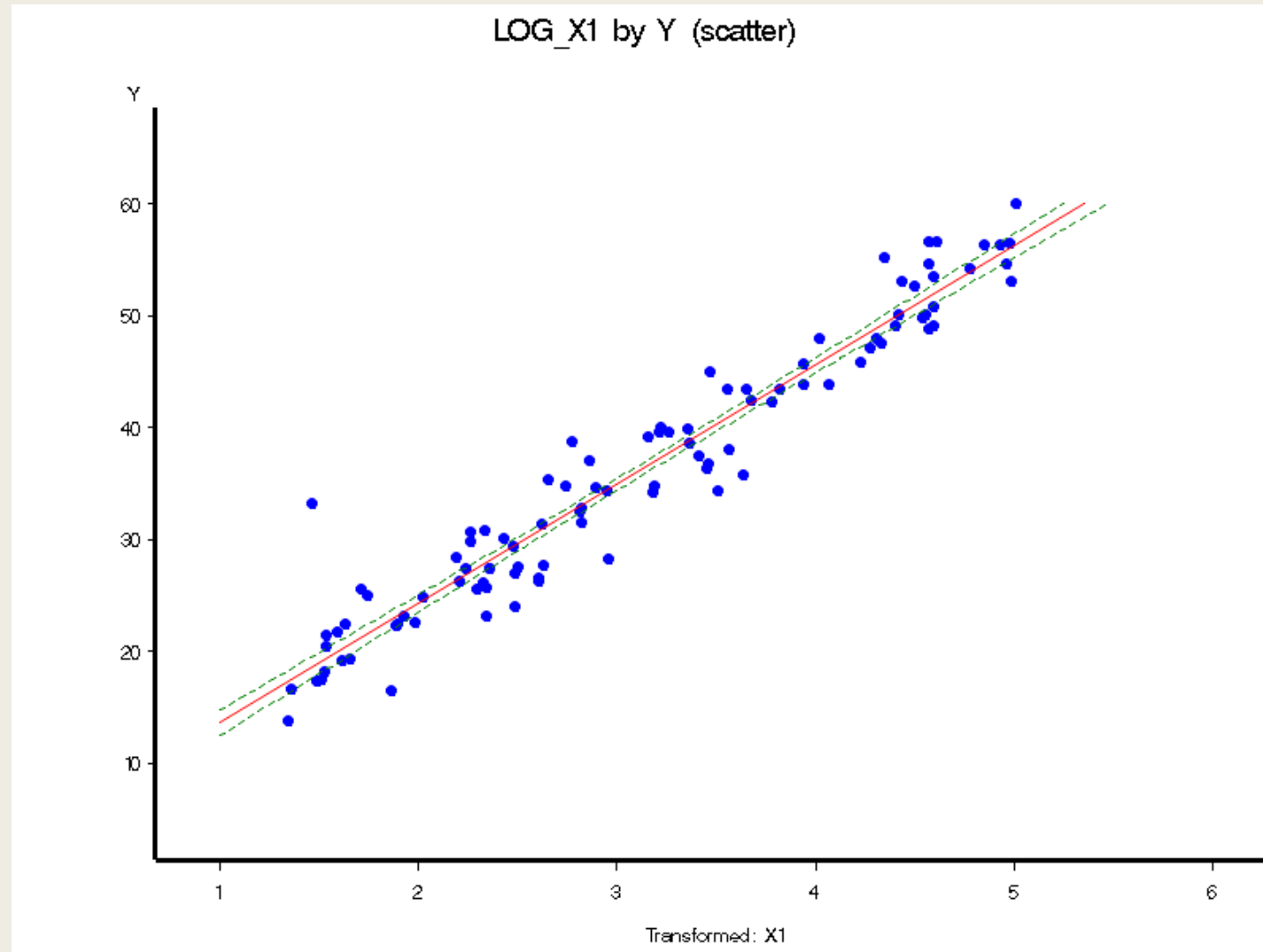
File Edit View Window

Output

```
1 *-----*
2 User:          u59397413
3 Date:          March 06, 2022
4 Time:          13:50:17
5 *-----*
6 * Training Output
7 *-----*
8
9
10
11
12 Variable Summary
13
14      Measurement      Frequency
15 Role      Level      Count
16
17 INPUT      INTERVAL      2
18 TARGET      INTERVAL      1
19
20
21
22 Computed Transformations
23 (maximum 500 observations printed)
24
25 Input      Input
26 Name      Role      Level      Name      Level      Formula
27
28 X1      INPUT      INTERVAL      LOG_X1      INTERVAL      log(X1 + 1)
29
30
31 *-----*
32 * Score Output
33 *-----*
34
35
36 *-----*
37 * Report Output
```

Windows taskbar: Type here to search, 08:51, 06-03-2022, 1°C, ENG

Answer 19: Scatter Plot between log X1 and X2



Answer 20: Drag the Regression node from Model tab and connect to the Transform node.

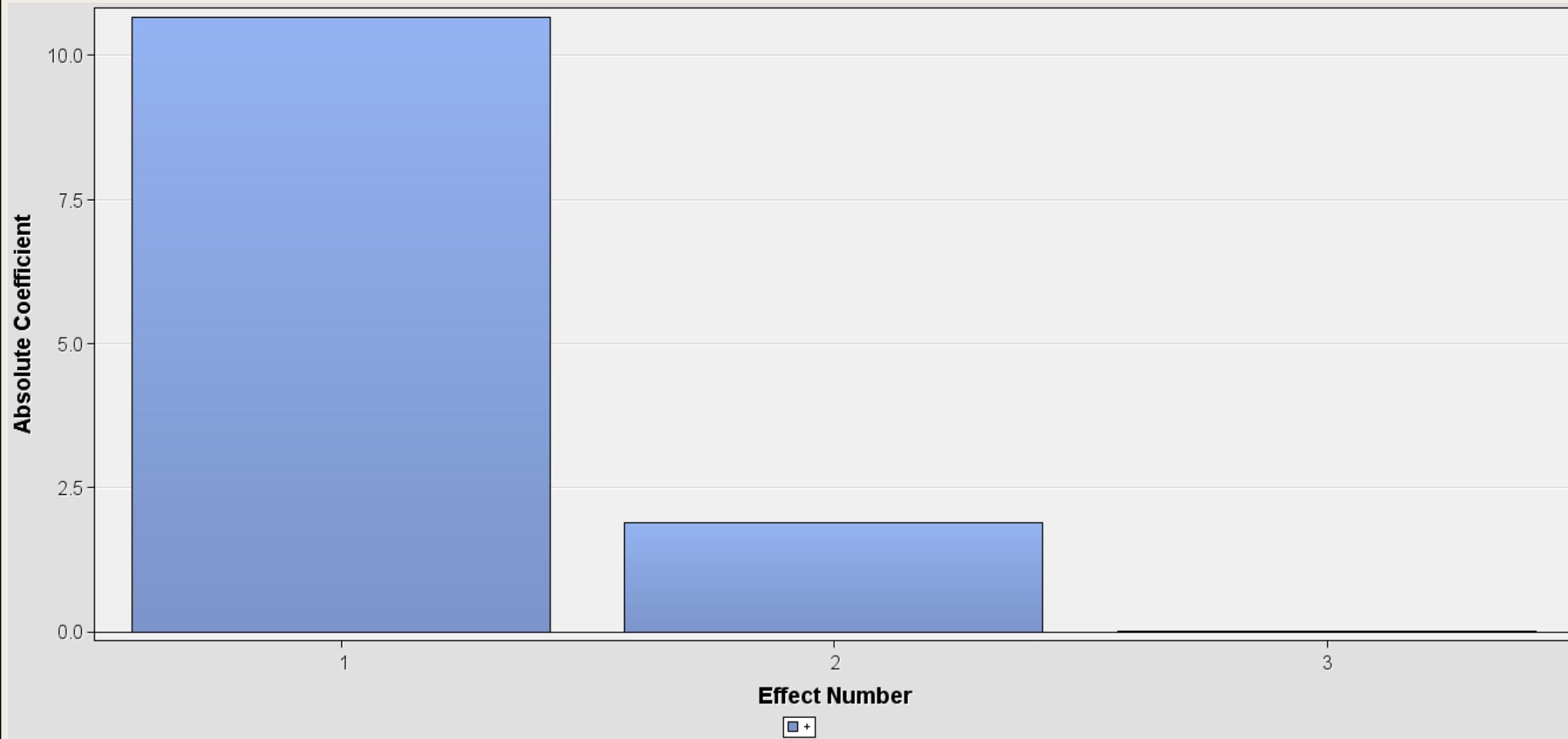
The screenshot displays the SAS Enterprise Miner interface. On the left, a tree view shows the project structure: W04_Regression, Data Sources, Diagrams, Regression, and Model Packages. The 'Regression' node is highlighted. Below the tree, a properties panel for the 'Regression' node is visible, showing various settings under 'General', 'Train', 'Class Targets', 'Model Options', 'Model Selection', and 'Optimization Options'.

The main workspace shows a workflow diagram titled 'Regression'. The diagram consists of two parallel paths. The top path starts with a 'File Import' node, followed by a 'Transform Variables' node, and then a 'Regression' node. The bottom path starts with a 'File Import (2)' node, followed by a 'Transform Variables (2)' node, and then a 'Regression (2)' node. Both 'Transform Variables' nodes are connected to 'MultiPlot' nodes. The 'Regression' node is highlighted with a green checkmark, indicating it is the selected node.

The bottom status bar shows the user is logged in as 'u59397413' and is connected to the 'SASApp - Logical Workspace Server'.

Property	Value
General	
Node ID	Reg2
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Equation	
Main Effects	Yes
Two-Factor Interactions	No
Polynomial Terms	No
Polynomial Degree	2
User Terms	No
Term Editor	
Class Targets	
Regression Type	Linear Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	None
Selection Criterion	Default
Use Selection Defaults	Yes
Selection Options	
Optimization Options	
Technique	Default
Default Optimization	Yes

Answer 21: Linear Regression model – Effects Plot



Answer 21: Linear Regression model – the formula for the model

The formula is $Y = 1.8937 + (10.6744 * \log X1) + (0.000186 * X2)$

Results - Node: Regression (2) Diagram: Regression

File Edit View Window

Output

```
44 Link Function Identity
45 Number of Model Parameters 3
46 Number of Observations 100
47
48
49 Analysis of Variance
50
51
52 Source DF Sum of Squares Mean Square F Value Pr > F
53
54 Model 2 13335 6667.301132 699.75 <.0001
55 Error 97 924.234046 9.528186
56 Corrected Total 99 14259
57
58
59 Model Fit Statistics
60
61 R-Square 0.9352 Adj R-Sq 0.9338
62 AIC 228.3795 BIC 230.5632
63 SBC 236.1950 C(p) 3.0000
64
65
66 Analysis of Maximum Likelihood Estimates
67
68
69 Parameter DF Estimate Standard Error t Value Pr > |t|
70
71 Intercept 1 1.8937 1.1350 1.67 0.0984
72 LOG_X1 1 10.6744 0.2853 37.41 <.0001
73 X2 1 0.000186 0.000114 1.63 0.1053
74
75
76 *-----*
77 * Score Output
78 *-----*
79
80
```

Windows taskbar: Type here to search, 09:01, 06-03-2022

Answer 21: Linear Regression model – Predicted Y for X1 = 10 and X2 =5000

$$Y = 1.8937 + (10.6744 * \log X1) + (0.000186 * X2)$$

$$Y = 1.8937 + (10.6744 * \log 10) + (0.000186 * 5000)$$

$$Y = 1.8937 + (10.6744 * 2.302585) + (0.000186 * 5000)$$

$$Y = 27.4024133$$

The predicted value of Y is **27.4024133**

Answer 22: Paste the diagram here

Enterprise Miner - W04_Regression

File Edit View Actions Options Window Help

W04_Regression

- Data Sources
- Diagrams
- Regression
- Model Packages

Property	Value
General	
Node ID	Reg2
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Equation	
Main Effects	Yes
Two-Factor Interactions	No
Polynomial Terms	No
Polynomial Degree	2
User Terms	No
Term Editor	
Class Targets	
Regression Type	Linear Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	None
Selection Criterion	Default
Use Selection Defaults	Yes
Selection Options	
Optimization Options	
Technique	Default
Default Optimization	Yes

Regression

```
graph LR; FI[File Import] --> TV[Transform Variables]; TV --> R[Regression]; MP[MultiPlot] --> TV; FI2[File Import (2)] --> TV2[Transform Variables (2)]; TV2 --> R2[Regression (2)]; MP2[MultiPlot (2)] --> TV2
```

Diagram Log

Running...

u59397413 as u59397413 Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)

Type here to search

09:22 06-03-2022



PART III: REGRESSION ASSESSMENT



Answer 23: Importing the file W04_3D.xlsx using File Import

First Regression Model

Model Fit Statistics			
R-Square	0.9267	Adj R-Sq	0.9251
AIC	122.8833	BIC	125.0465
SBC	126.7074	C(p)	2.0000

Fit		
Statistics	Statistics Label	Train
AIC	Akaike's Information Criterion	122.883
ASE	Average Squared Error	10.780

Second Regression Model

Model Fit Statistics			
R-Square	0.9352	Adj R-Sq	0.9338
AIC	228.3795	BIC	230.5632
SBC	236.1950	C(p)	3.0000

Fit		
Statistics	Statistics Label	Train
AIC	Akaike's Information Criterion	228.380
ASE	Average Squared Error	9.242

	R squared	Average squared error
First regression model	0.9267	10.780
Second regression model	0.9352	9.242

- R-squared value determine how close the data are to the fitted regression line and average squared error is the mean square error(of the predicted and the observed value). The more the value of the R-Square , the better the model will behave.
- It is indicated that the addition of more independent variables can lead to a better analysis and hence good R-square value. However, comparing the univariate and the multivariate regression using different dataset would not result in best result as these are tow different statistical approach.
- The better fit on training data does not normally comply that a resulted model would be performing best on test or validation dataset as the model would be performing on different dataset and hence could result in different statistics.

GROUP WORK DECLARATION

We, **Group 5 (Anand Mohan Thankur, Josh Shaji, Poonam Bhaliyan, Prateek Ramjanam Singh, and Poornima Singh)** declare that the attached assignment is our own work in accordance with the Seneca Academic Policy. We have not copied any part of this assignment, manually or electronically, from any other source including web sites, unless specified as references. We have not distributed our work to other students.

	Name	Task(s)
1	Anand Mohan Thakur (149200206)	Consolidated the Workshop together on MS Teams
2	Josh Shaji (133557215)	Consolidated the Workshop together on MS Teams
3	Poonam Bhaliyan (121114219)	Consolidated the Workshop together on MS Teams
4	Prateek Ramjanam Singh (124483215)	Consolidated the Workshop together on MS Teams
5	Poornima Singh (125638213)	Consolidated the Workshop together on MS Teams



THANK YOU

