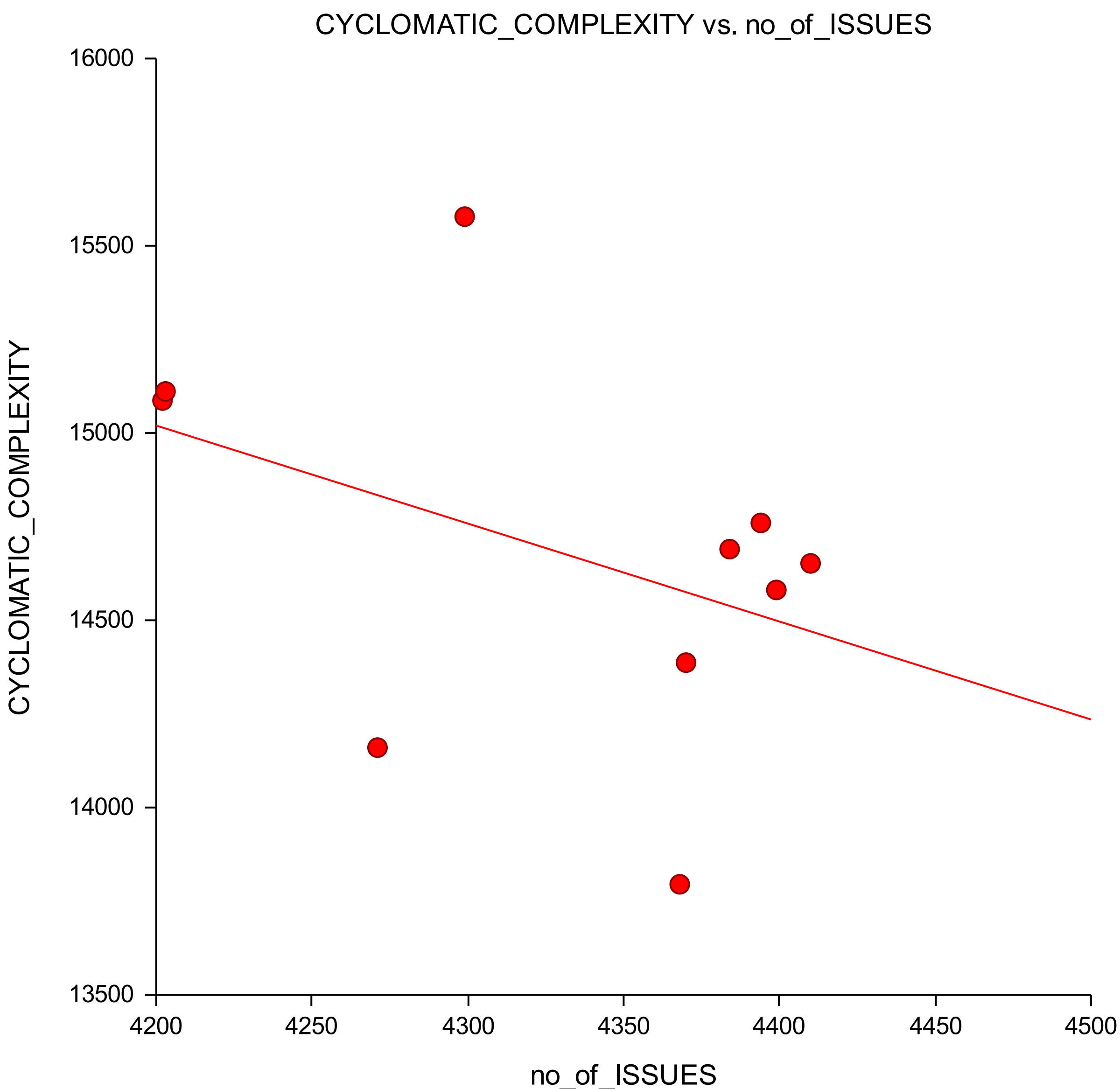
Linear Regression Report

Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES

Linear Regression Plot Section



Linear Regression Report

Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES

Run Summary Section

Parameter	Value	Parameter	Value
Dependent Variable	CYCLOMATIC_COMPLEXITY	Rows Processed	10
Independent Variable	no_of_ISSUES	Rows Used in Estimation	10
Frequency Variable	None	Rows with X Missing	0
Weight Variable	None	Rows with Freq Missing	0
Intercept	26050.1089	Rows Prediction Only	0
Slope	-2.6259	Sum of Frequencies	10
R-Squared	0.1733	Sum of Weights	10.0000
Correlation	-0.4163	Coefficient of Variation	0.0333
Mean Square Error	239158.3	Square Root of MSE	489.0381

Summary Statement

The equation of the straight line relating CYCLOMATIC_COMPLEXITY and no_of_ISSUES is estimated as: CYCLOMATIC_COMPLEXITY = (26050.1089) + (-2.6259) no_of_ISSUES using the 10 observations in this dataset. The y-intercept, the estimated value of CYCLOMATIC_COMPLEXITY when no_of_ISSUES is zero, is 26050.1089 with a standard error of 8780.9394. The slope, the estimated change in CYCLOMATIC_COMPLEXITY per unit change in no_of_ISSUES, is -2.6259 with a standard error of 2.0276. The value of R-Squared, the proportion of the variation in CYCLOMATIC_COMPLEXITY that can be accounted for by variation in no_of_ISSUES, is 0.1733. The correlation between CYCLOMATIC_COMPLEXITY and no_of_ISSUES is -0.4163.

A significance test that the slope is zero resulted in a t-value of -1.2951. The significance level of this t-test is 0.2314. Since 0.2314 > 0.0500, the hypothesis that the slope is zero is not rejected.

The estimated slope is -2.6259. The lower limit of the 95% confidence interval for the slope is -7.3016 and the upper limit is 2.0498. The estimated intercept is 26050.1089. The lower limit of the 95% confidence interval for the intercept is 5801.2264 and the upper limit is 46298.9915.

Linear Regression Report

Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES

Regression Estimation Section

Parameter	Intercept B(0)	Slope B(1)
Regression Coefficients	26050.1089	-2.6259
Lower 95% Confidence Limit	5801.2264	-7.3016
Upper 95% Confidence Limit	46298.9915	2.0498
Standard Error	8780.9394	2.0276
Standardized Coefficient	0.0000	-0.4163
T Value	2.9667	-1.2951
Prob Level (T Test)	0.0180	0.2314
Reject H0 (Alpha = 0.0500)	Yes	No
Power (Alpha = 0.0500)	0.7387	0.2081
Regression of Y on X	26050.1089	-2.6259
Inverse Regression from X on Y	80284.7054	-15.1512
Orthogonal Regression of Y and X	78927.5500	-14.8377

Notes:
The above report shows the least-squares estimates of the intercept and slope followed by the corresponding standard errors, confidence intervals, and hypothesis tests. Note that these results are based on several assumptions that should be validated before they are used.

Estimated Model
(26050.1089390069) + (-2.6258681152467) * (no_of_ISSUES)

Linear Regression Report

Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES

Correlation and R-Squared Section

Parameter	Pearson Correlation Coefficient	R-Squared	Spearman Rank Correlation Coefficient
Estimated Value	-0.4163	0.1733	-0.2848
Lower 95% Conf. Limit (r dist'n)	0.2748		
Upper 95% Conf. Limit (r dist'n)	-0.8065		
Lower 95% Conf. Limit (Fisher's z)	-0.8287		-0.7754
Upper 95% Conf. Limit (Fisher's z)	0.2891		0.4201
Adjusted (Rbar)		0.0700	
T-Value for H0: Rho = 0	1.2951	1.2951	0.8405
Prob Level for H0: Rho = 0	0.2314	0.2314	0.4250

Notes:
The confidence interval for the Pearson correlation assumes that X and Y follow the bivariate normal distribution. This is a different assumption from linear regression which assumes that X is fixed and Y is normally distributed.

Two confidence intervals are given. The first is based on the exact distribution of Pearson's correlation. The second is based on Fisher's z transformation which approximates the exact distribution using the normal distribution. Why are both provided? Because most books only mention Fisher's approximate method, it will often be needed to do homework. However, the exact methods should be used whenever possible.

The confidence limits can be used to test hypotheses about the correlation. To test the hypothesis that rho is a specific value, say r0, check to see if r0 is between the confidence limits. If it is, the null hypothesis that rho = r0 is not rejected. If r0 is outside the limits, the null hypothesis is rejected.

Spearman's Rank correlation is calculated by replacing the orginal data with their ranks. This correlation is used when some of the assumptions may be invalid.

Summary Matrices

Index	X'X 0	X'X 1	X'Y 2	X'X Inverse 0	X'X Inverse 1
0	10	43300	146801	322.4011	-0.07443444
1	43300	1.875472E+08	6.354956E+08	-0.07443444	1.71904E-05
2 (Y'Y)			2.157368E+09		
Determinant		581720			1.71904E-06

Variance - Covariance Matrix of Regression Coefficients

Index	VC(b) 0	VC(b) 1
0	7.71049E+07	-17801.61
1	-17801.61	4.111227

Linear Regression Report

Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES

Tests of Assumptions Section

Assumption/Test	Test Value	Prob Level	Is the Assumption Reasonable at the 0.2000 Level of Significance?
Residuals follow Normal Distribution?			
Shapiro Wilk	0.9004	0.221058	Yes
Anderson Darling	0.6543	0.087717	No
D'Agostino Skewness	-0.4855	0.627350	Yes
D'Agostino Kurtosis	0.6901	0.490132	Yes
D'Agostino Omnibus	0.7119	0.700505	Yes
Constant Residual Variance?			
Modified Levene Test	3.4156	0.101769	No
Relationship is a Straight Line?			
Lack of Linear Fit F(0, 0) Test	0.0000	0.000000	No

No Serial Correlation?
Evaluate the Serial-Correlation report and the Durbin-Watson test if you have equal-spaced, time series data.

Notes:
A 'Yes' means there is not enough evidence to make this assumption seem unreasonable. This lack of evidence may be because the sample size is too small, the assumptions of the test itself are not met, or the assumption is valid.
A 'No' means the that the assumption is not reasonable. However, since these tests are related to sample size, you should assess the role of sample size in the tests by also evaluating the appropriate plots and graphs. A large dataset (say N > 500) will often fail at least one of the normality tests because it is hard to find a large dataset that is perfectly normal.

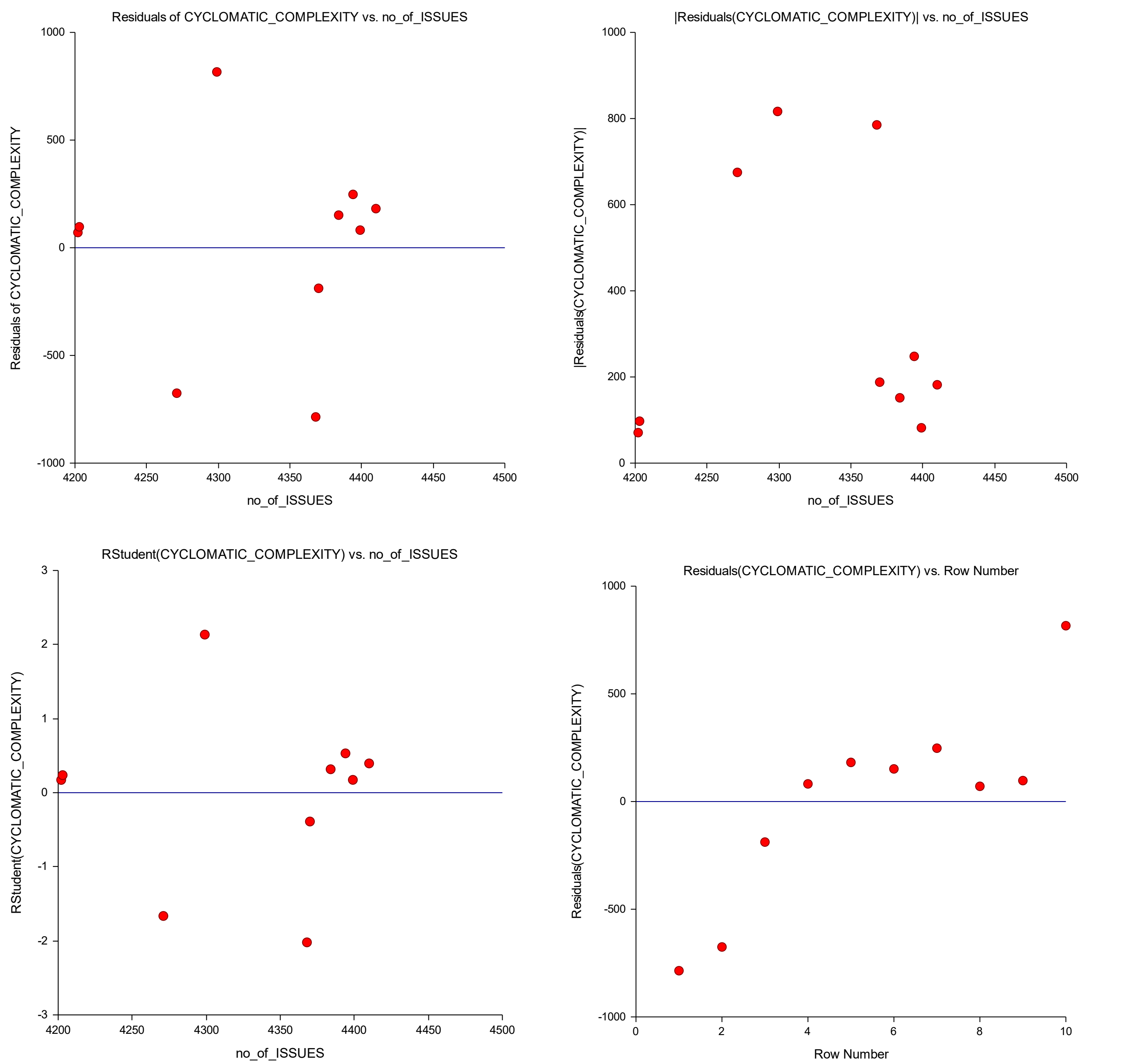
Normality and Constant Residual Variance:
Possible remedies for the failure of these assumptions include using a transformation of Y such as the log or square root, correcting data-recording errors found by looking into outliers, adding additional independent variables, using robust regression, or using bootstrap methods.

Straight-Line:
Possible remedies for the failure of this assumption include using nonlinear regression or polynomial regression.

Linear Regression Report

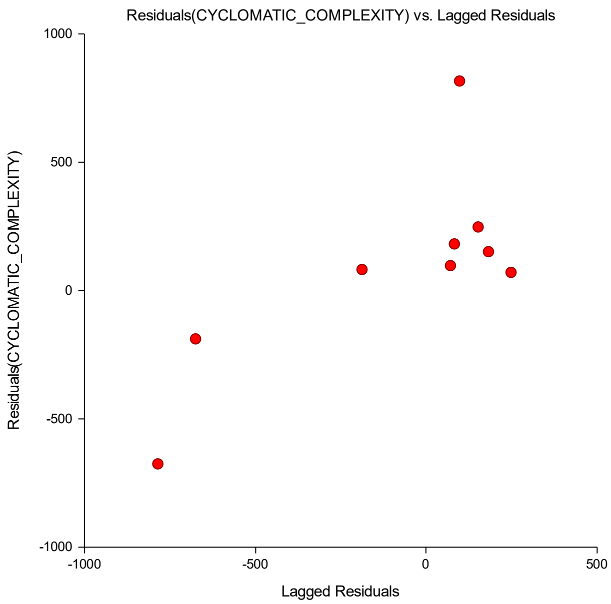
Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES

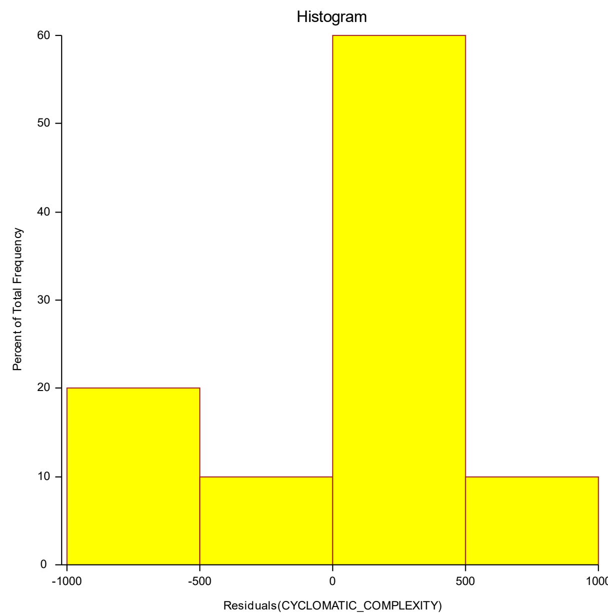
Residual Plots Section

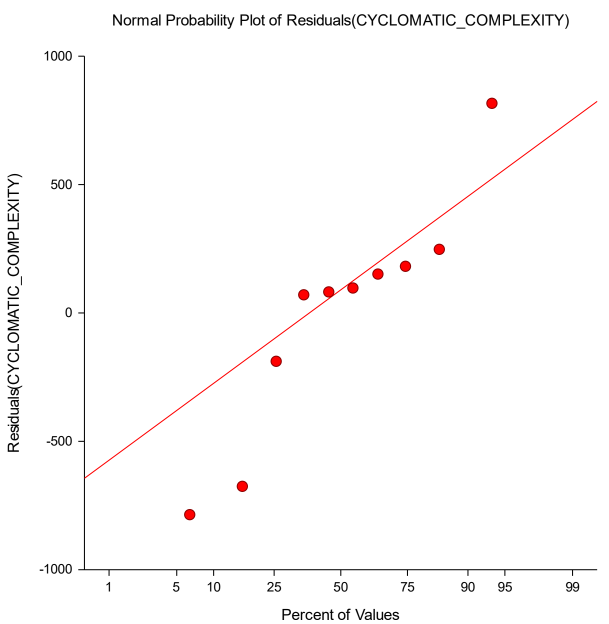


Linear Regression Report

Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES







Linear Regression Report

Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES

Procedure Input Settings

Autosaved Template File
C:\Users\KASATLA\Documents\NCSS 12\Procedure Templates\Autosave\Linear Regression and Correlation - Autosaved 2018_3_30-0_43_26.t153

Variables Tab

-- Variables	
Y: Dependent Variable(s):	CYCLOMATIC_COMPLEXITY
X: Independent Variable:	no_of_ISSUES
Frequency Variable:	<Empty>
Weight Variable:	<Empty>
-- Model Specification	
Remove Intercept	Unchecked
-- Resampling (Increases computation time)	
Calculate Bootstrap C.I.'s	Unchecked
Run Randomization Tests	Unchecked
-- Alpha Levels	
Alpha for C.I.'s and Tests:	0.050
Alpha for Assumptions:	0.20

Reports Tab

-- Select Report / Plot Group	
Select a Group of Reports and Plots:	Display only those items that are CHECKED BELOW
Show Notes	Checked
Show All Rows	Checked
-- Select Reports	
·· Summaries	
Run Summary	Checked
Summary Statement	Checked
Descriptive Statistics	Unchecked
Correlation and R-Squared	Checked
Summary Matrices	Checked
·· Estimation	
Regression Estimation	Checked
·· ANOVA	
ANOVA	Unchecked
·· Assumptions	
Assumptions	Checked
Levene Groups:	2
Durbin-Watson	Unchecked
PRESS	Unchecked

Linear Regression Report

Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES

Procedure Input Settings (Continued)

Reports Tab (Continued)

·· Prediction ··································

Predict Y at these X values:	<Empty>
Predicted Y - C.L.	Unchecked
Predicted Y - P.L.	Unchecked

·· Row-by-Row Lists ··································

Original Data	Unchecked
Predicted Y Means	Unchecked
Predicted Y Individuals	Unchecked
Simultaneous Bands	Unchecked
Predicted X Means	Unchecked
Predicted X Individuals	Unchecked

·· Regression Diagnostics ··································

Residuals	Unchecked
Residual Diagnostics	Unchecked
Leave One Row Out	Unchecked
Outlier Detection Chart	Unchecked
Influence Detection Chart	Unchecked
Outlier-Influence Chart	Unchecked

Report Options Tab

-- Report Options ----------------------------------

Precision:	Single
Variable Names:	Names

·· Decimal Places ··································

Probability:	4
Beta (Coefficients):	4
SE:	4
T:	4
R2:	4
X:	4
Y:	4
Residuals:	4
Std Residuals:	4
Sum Squares:	All
Matrix:	All

Linear Regression Report

Dataset Untitled
Y = CYCLOMATIC_COMPLEXITY X = no_of_ISSUES

Procedure Input Settings (Continued)

Plots Tab

-- Select Plots --

Y vs X	Checked
RStudent vs X	Checked
Histogram	Checked
Residuals vs X	Checked
Residuals vs Row	Checked
Probability Plot	Checked
|Residuals| vs X	Checked
Serial Correlation	Checked

-- Plot Options --

Y vs X Plot Size:	Medium
All Other Plot Sizes:	Small

Resampling Tab

-- Bootstrap Calculation Options --

·· Sampling ································

Samples (N):	3000
Sampling Method:	Observations
Retries:	50

·· Estimation ································

Percentile Type:	Ave X(p[n+1])
C.I. Method:	Reflection
Bootstrap Confidence Coefficients:	0.90 0.95 0.99

-- Randomization Test Options --

Monte Carlo Samples:	1000

Storage Tab

-- Data Storage Options --

Storage Option:	Do not store data