

Hannah Corrigan fall project PDF Analysis

Fitted Line Plot With Outliers

Worksheet 1

Regression Analysis: GPA versus TIME(MINUTES)

Regression Equation

$$\text{GPA} = 3.580 - 0.00321 \text{ TIME(MINUTES)}$$

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.580	0.310	11.54	0.000	
TIME(MINUTES)	-0.00321	0.00766	-0.42	0.678	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.723005	0.63%	0.00%	0.00%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.0921	0.09212	0.18	0.678
TIME(MINUTES)	1	0.0921	0.09212	0.18	0.678
Error	28	14.6366	0.52274		
Lack-of-Fit	8	4.6331	0.57914	1.16	0.370

Pure Error	20	10.003	0.50018
		5	
Total	29	14.728	
		7	

Fits and Diagnostics for Unusual Observations

Obs	GPA	Fit	Resid	Std Resid	
14	4.14	3.29	0.85	1.46	X
	3	1	2		
24	2.00	3.38	-1.38	-2.02	R
	0	8	8		
26	2.00	3.38	-1.38	-2.02	R
	0	8	8		

R Large residual

X Unusual X

Fitted Line Plot Without Outliers

Without outliers

Regression Analysis: GPA versus TIME(MINUTES)

Regression Equation

GP
A = 3.674 - 0.00539
A TIME(MINUTES)

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.674	0.340	10.79	0.000	
TIME(MINUTES)	-0.00539	0.00930	-0.58	0.567	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.67795	1.28	0.00%	0.00%
2	%		

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.1547	0.1547	0.34	0.567
TIME(MINUTES)	1	0.1547	0.1547	0.34	0.567
Error	26	11.950	0.4596		
Lack-of-Fit	7	2.7474	0.3925	0.81	0.590
Pure Error	19	9.2027	0.4844		
Total	27	12.104			

Fits and Diagnostics for Unusual Observations

Obs	GPA	Fit	Resid	Std Resid
24	2.00	3.35	-1.35	-2.18 R
	0	0	0	
25	4.81	3.43	1.37	2.10 R
	0	1	9	

R Large residual

Histogram With Outliers

Worksheet 1

Regression Analysis: GPA versus TIME(MINUTES)

Regression Equation

$$\text{GPA} = 3.580 - 0.00321 \text{ TIME(MINUTES)}$$

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.580	0.310	11.54	0.000	
TIME(MINUTES)	-0.00321	0.00766	-0.42	0.678	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.723005	0.63%	0.00%	0.00%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.0921	0.09212	0.18	0.678
TIME(MINUTES)	1	0.0921	0.09212	0.18	0.678
Error	28	14.6366	0.52274		
Lack-of-Fit	8	4.6331	0.57914	1.16	0.370
Pure Error	20	10.0035	0.50018		
Total	29	14.7287			

Fits and Diagnostics for Unusual Observations

Obs	GPA	Fit	Resid	Std Resid	
14	4.143	3.291	0.852	1.46	X
24	2.000	3.388	-1.388	-2.02	R

26 2.00 3.38 -1.38 -2.02 R
 0 8 8

R Large residual

X Unusual X

Histogram without Outliers

Without outliers

Regression Analysis: GPA versus TIME(MINUTES)

Regression Equation

GP = 3.674 - 0.00539
 A TIME(MINUTES)

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.674	0.340	10.79	0.000	
TIME(MINUTES)	-0.00539	0.00930	-0.58	0.567	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.677952	1.28%	0.00%	0.00%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
--------	----	--------	--------	---------	---------

Regression	1	0.1547	0.1547	0.34	0.567
	1	0.1547	0.1547	0.34	0.567
TIME(MINUTE S)					
Error	26	11.950	0.4596		
		1			
Lack-of-Fit	7	2.7474	0.3925	0.81	0.590
Pure Error	19	9.2027	0.4844		
Total	27	12.104			
		8			

Fits and Diagnostics for Unusual Observations

Obs	GPA	Fit	Resid	Std Resid
24	2.00	3.35	-1.35	-2.18 R
	0	0	0	
25	4.81	3.43	1.37	2.10 R
	0	1	9	

R Large residual

When asked to complete this project I thought about how I could have a strong relationship and what I could benefit from after I completed the project. After some time I came up with the question “Is there a relationship between a highschool student's GPA and how long it takes for them to get ready in the morning?”.

I decided that I wanted to collect my own data and began this project by creating a google survey which I sent to 30 people who I either shared a class with or who I knew was a senior at North Olmsted. After collecting the data I then put it into Minitab and was shocked when I found out that that correlation was extremely weak at just 1.28. In my data I found that there were 2 outliers, which I then created another page to analyze the data without them. The “R” was 1.1314, the “R-Sq” was 1.28, and the regression equation was “GPA=3,674-0.005393 Time(Minutes). In this analysis I saw that there was no change with outliers and without outliers in the “R”, regression equation, and “R-Sq” because of the extremely weak correlation in the data set.

When creating this project I wanted to survey just seniors at North Olmsted to see if being more mature would give me better results. When I look back at my results I find that I did not collect data in the most efficient way and I believe that if there was a better way to find students from all different classes and programs at our school, there may have been a stronger correlation. Also, if I surveyed 50 students from each grade randomly selected, I could have broadened my data.

Descriptive Statistics: TIME(MINUTES), GPA

Statistics

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
TIME(MINUTES)	6.67	7.53	0.00	8.75	0.00	5.00	9.00
GPA	3.463	0.713	2.000	2.900	3.564	3.013	4.810

Data collected:

GPA Time

2.9 45

2.9 40

3.5 60

4.051 45

3.3 10

3.1 30

3.4 20

3.6 60

2.8 30

3.35 5

3.65 30

4 30

2.5 25

4.143 90

3.9 45

2.7 45

3.529 30

3.6 40

3.6 25

3.67 30

4.4 45

3.6 30

2.576 30

2 60

4.42 20

2	60
4.81	45
3.1	30
4.4	30
4.379	15