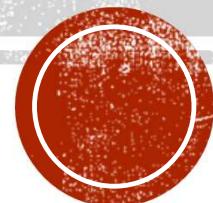


LINEAR REGRESSION MODEL

TED TALK

Aunchana Pimpisal

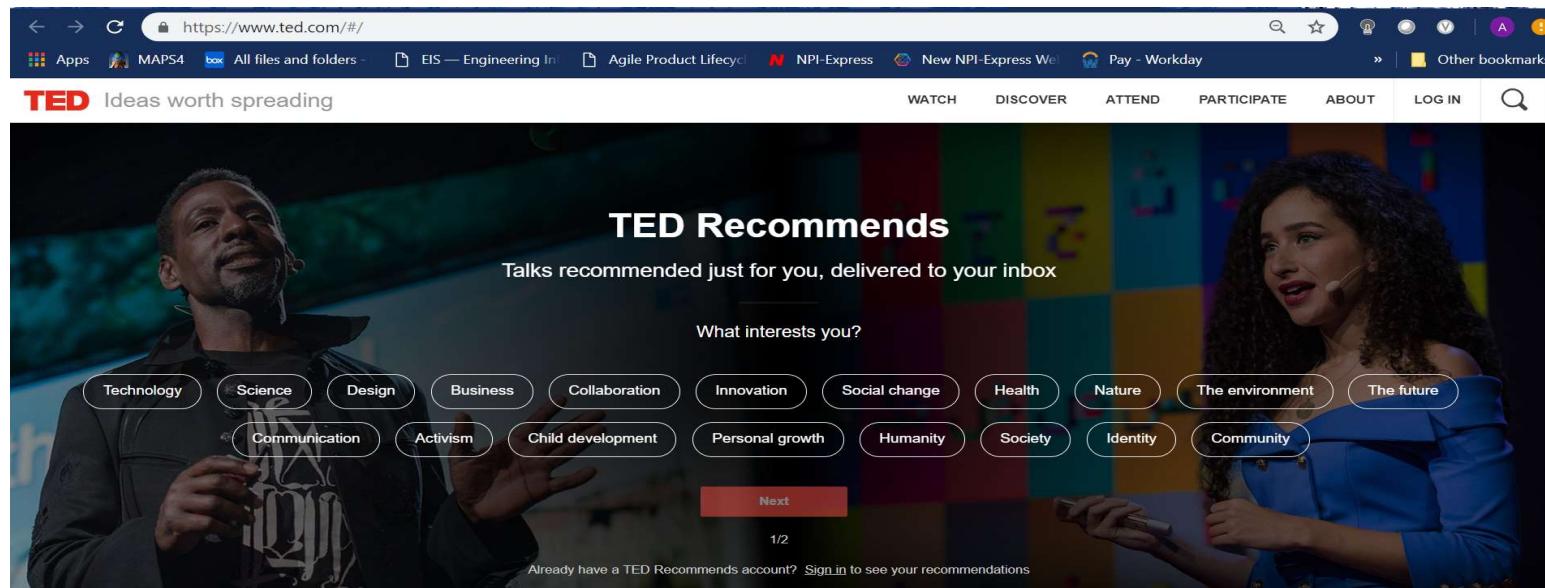
16 Dec'18



TED TALK

TED Conferences LLC (Technology, Entertainment, Design) is a media organization that posts talks online for free distribution under the slogan "ideas worth spreading."

TED events are also held throughout North America and in Europe and Asia, offering live streaming of the talks. They address a wide range of topics within the research and practice of science and culture, often through storytelling. The speakers are given a maximum of 18 minutes to present their ideas in the most innovative and engaging ways they can. TED Talks given by academics tend to be watched more online while art and design videos tend to be watched less than average.^[20]



Reference: [https://en.wikipedia.org/wiki/TED_\(conference\)#TEDGlobal](https://en.wikipedia.org/wiki/TED_(conference)#TEDGlobal)



DATASET :TED TALK

views	comments	duration	num Speaker	languages	event	event1	event2	event3	event4	event5	event6	event7	event8	event9	main Speaker	Name	Speaker occupation	Title
1106561	84	1201	1	32	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Jimmy Wales	Jimmy Wales: The Founder of Wikipedia	The birth of Wikipedia	
1096862	184	1140	1	29	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	David Deutsch	David Deutsch: Ch	Quantum physicist	
2885999	507	1316	1	36	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Richard Dawkins	Richard Dawkins: Evolutionary biologist	Why the universe seems so strange	
10000702	991	1177	1	45	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Barry Schwartz	Barry Schwartz: Th	Psychologist	
3277740	612	1365	1	29	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Aubrey de Grey	Aubrey de Grey: A Crusader against aging	A roadmap to end aging	
503517	74	952	1	21	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Iqbal Quadir	Iqbal Quadir: How	Founder, GrameenPhone	
705389	58	773	1	26	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Jacqueline Novogratz	Jacqueline Novogratz	Investor and advocate for	
580891	43	1080	1	21	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Sasa Vucinic	Sasa Vucinic: Why	Nonprofit venture capital	
809226	67	1125	1	24	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Ashraf Ghani	Ashraf Ghani: How	President-elect of Afghanistan	
1053758	105	1280	1	20	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Peter Donnelly	Peter Donnelly: Ho	Mathematician; statistician	
673036	85	843	1	22	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Robert Neuwirth	Robert Neuwirth: Author	The hidden world of shadow cities	
2121177	203	1046	1	29	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Martin Rees	Martin Rees: Is thi	Astrophysicist	
1409327	91	1141	1	24	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Charles Leadbeater	Charles Leadbeater	Innovation consultant	
2441805	226	1155	1	30	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Carl Honors	Carl Honors: In p	Journalist	
1392010	107	1054	1	32	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Alex Steffen	Alex Steffen: The r	Planetary futurist	
761930	27	850	1	24	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Paul Bennett	Paul Bennett: Desi	Designer; creative director	
762264	131	1012	1	24	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Nick Bostrom	Nick Bostrom: A p	Philosopher	
560904	34	1011	1	21	TEDGlobal 2005	1	0	0	0	0	0	0	0	0	Craig Venter	Craig Venter: Sam	Biologist, genetics pioneer	
648234	97	1070	1	20	TEDGlobal 2007	0	1	0	0	0	0	0	0	0	George Ayittey	George Ayittey: Afr	Economist	
524049	108	1330	1	21	TEDGlobal 2007	0	1	0	0	0	0	0	0	0	Ngozi Okonjo-Iweala	Ngozi Okonjo-Iweala	Economist	
1543596	174	252	1	44	TEDGlobal 2007	0	1	0	0	0	0	0	0	0	William Kamkwamba	William Kamkwamba	Inventor	
335086	31	1141	1	21	TEDGlobal 2007	0	1	0	0	0	0	0	0	0	Euvin Naidoo	Euvin Naidoo: Wh	Investment banker	
1216429	150	1051	1	27	TEDGlobal 2007	0	1	0	0	0	0	0	0	0	Patrick Awuah	Patrick Awuah: Ho	University founder	
																	How to educate leaders? Liberal arts	

Reference: <https://www.kaggle.com/rounakbanik/ted-talks>



Microsoft Excel
Worksheet



MODEL CHART

Ted talk: Regressors Factor effect to views

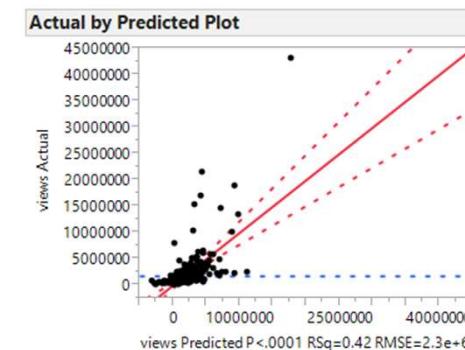


Define Variable as respond (views) and Regressors (4 Quantitative and 9 Qualitative Variables)



Fitting model

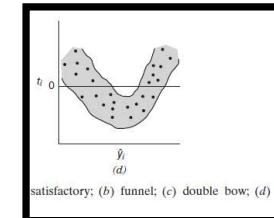
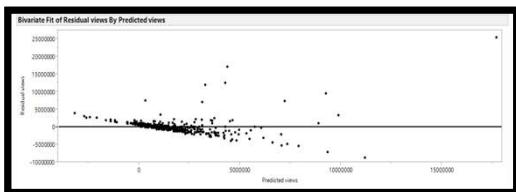
- Hypothesis testing : Model Adequacy
 - Residual plot Assumption:
 - ϵ has no zero mean and non constant variance.
 - Errors are uncorrelated
 - Errors are not normally distributed.
- 2/3 are not following assumption.



MODEL CHART

Transformation

As Variance-stabilizing Transformations , transform Y into $Y' = \log Y$.

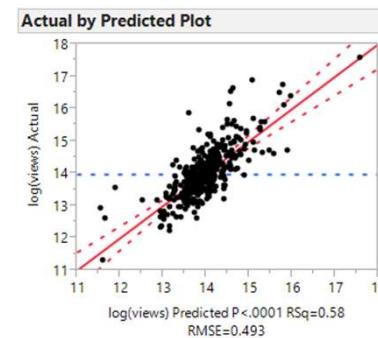


Stepwise : Fitting the model



Fitting model

- Hypothesis testing : Model Adequacy
 - Residual plot Assumption:
 - ε has no zero mean and non constant variance.
 - Errors are uncorrelated
 - **Errors are not normally distributed.**
- 1/3 are not following assumption.

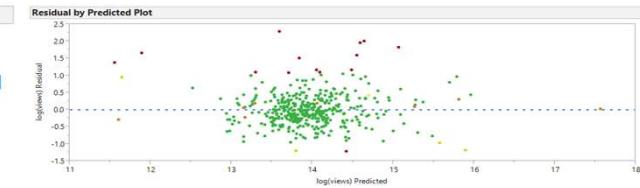
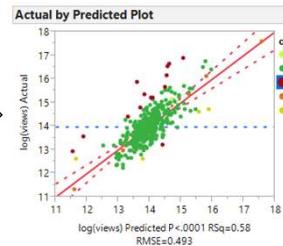
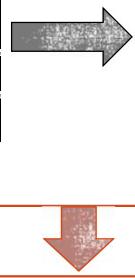


MODEL CHART

Diagnostics for leverage and influence

cutting group “Risk” from R program and “Degrade mode” from Cov.r about 14 data point out.

Statistics for detect influential	Cut off	Cut off Number
cook.d	1	1
dov.r	improve:1+3p/n	1.083526682
	Degrade:1+3p/n	0.916473318
dfb_3		
dfb_X1		
dfb_X10		
dfb_X11		
dfb_X12		
dfb_X13		
dfb_X2		
dfb_X4		
dfb_X6		
dfb_X7		
dfb_X8		
dfb_X9		
dfit	$2\sqrt{\frac{p}{n}}$	0.333719807
hat	$\frac{2p}{n}$	0.055684455



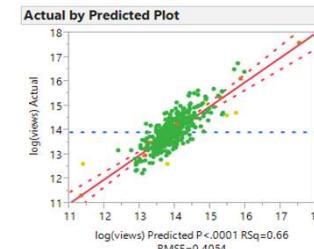
Stepwise : Fitting the model



Fitting model

- Hypothesis testing : Model Adequacy
- Residual plot Assumption:
 - ϵ has no zero mean and non constant variance.
 - Errors are uncorrelated
 - Errors are not normally distributed.

Model: $\text{Log(views)} = 12.7208 + 0.0010 * \text{comments} + 0.0005 * \text{durations} + 0.0542 * \text{languages} - 0.8908 * \text{event1} - 1.3408 * \text{event2} - 0.9396 * \text{event3} - 1.2772 * \text{event4} - 1.2927 * \text{event5} - 0.7942 * \text{event6} - 0.6708 * \text{event7} - 0.7128 * \text{event8}$



TED TALK: REGRESSOR FACTOR EFFECT TO VIEWS

- **Response:**

Views => The number of views on the talk

- **Regressor :**

4 Quantitative Variable

Comments => The number of comments made on the talk

Duration => The duration of the talk in seconds

num_speaker => The number of speakers in the talk

Languages => The number of languages in which the talk is available

9 Qualitative Variable

Event => The TED event where the talk took place

event	event1	event2	event3	event4	event5	event6	event7	event8	event9	
TEDGlobal 2005	1	0	0	0	0	0	0	0	0	0 if the observation is from TEDGlobal 2005
TEDGlobal 2007	0	1	0	0	0	0	0	0	0	0 if the observation is from TEDGlobal 2007
TEDGlobal 2009	0	0	1	0	0	0	0	0	0	0 if the observation is from TEDGlobal 2009
TEDGlobal 2010	0	0	0	1	0	0	0	0	0	0 if the observation is from TEDGlobal 2010
TEDGlobal 2011	0	0	0	0	1	0	0	0	0	0 if the observation is from TEDGlobal 2011
TEDGlobal 2012	0	0	0	0	0	1	0	0	0	0 if the observation is from TEDGlobal 2012
TEDGlobal 2013	0	0	0	0	0	0	1	0	0	0 if the observation is from TEDGlobal 2013
TEDGlobal 2014	0	0	0	0	0	0	0	1	0	0 if the observation is from TEDGlobal 2014
TEDGlobal 2017	0	0	0	0	0	0	0	0	0	0 if the observation is from TEDGlobal 2017



MODEL: HYPOTHESIS TESTING

Prediction Expression :

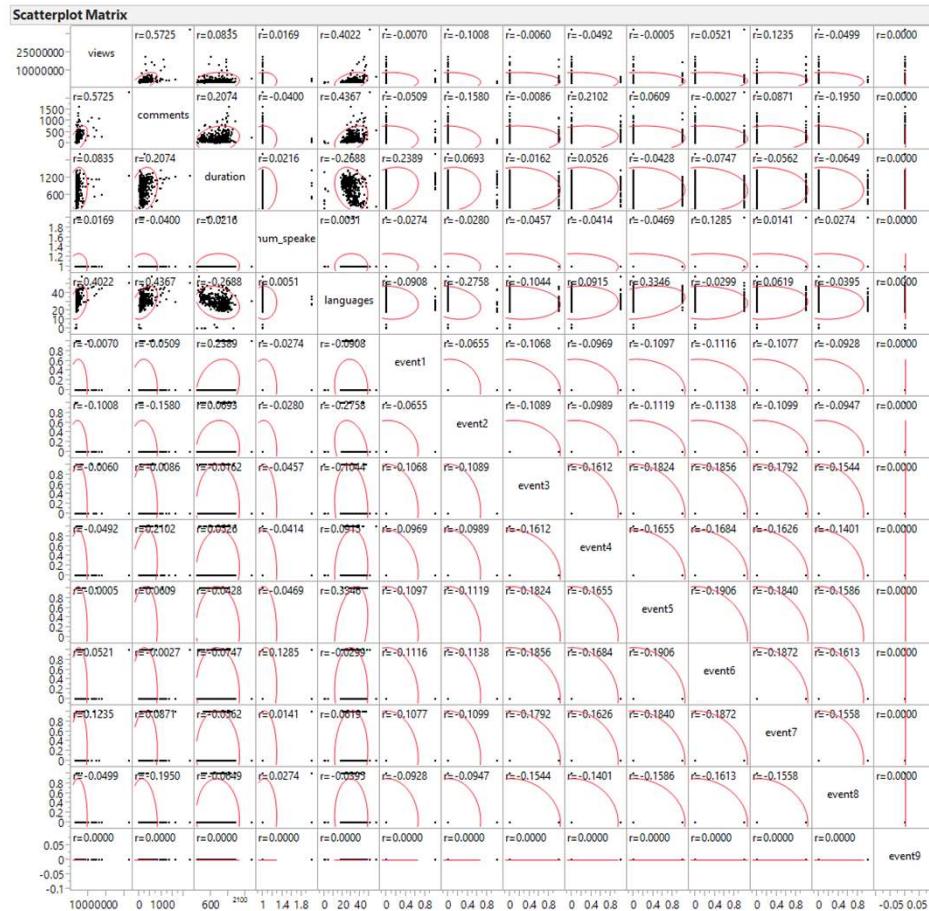
```

Prediction Expression
-732974.19427922
+ 6100.97660846304 * comments
+ 700.814732083342 * duration
+ 458442.635518517 * num_speaker
+ 122036.911869021 * languages
+ -3177534.2194861 * event1
+ -2863817.7599929 * event2
+ -3291497.7531205 * event3
+ -4942207.6634355 * event4
+ -4415397.5990058 * events5
+ -3076091.9180759 * events6
+ -3061976.1165143 * events7
+ -3023238.3826778 * events8
+ 0 * event9

```

Graph	Hypothesis testing	Result																																																																																																
<p>Actual by Predicted Plot</p> <p>views Actual vs views Predicted P<.0001 RSq=0.42 RMSE=2.3e+6</p>	<p>1. Test for significant of Regression(F-Test)</p> <p>Hypothesis $H_0: \beta_0 = \beta_1 = \dots = \beta_i = 0, i=0,1,2,\dots,14$ $H_1: \beta_0 \neq 0$ for at least one $i, i=0,1,2,\dots,14$, $\alpha = 0.05$</p> <p>Result from JMP</p> <table border="1"> <thead> <tr> <th colspan="2">Summary of Fit</th> <th colspan="3">Analysis of Variance</th> </tr> </thead> <tbody> <tr> <td>RSquare</td> <td>0.415576</td> <td>Source</td> <td>DF</td> <td>Sum of Squares</td> </tr> <tr> <td>RSquare Adj</td> <td>0.398798</td> <td>Model</td> <td>12</td> <td>1.5817e+15</td> </tr> <tr> <td>Root Mean Square Error</td> <td>2306835</td> <td>Error</td> <td>418</td> <td>2.2244e+15</td> </tr> <tr> <td>Mean of Response</td> <td>1721104</td> <td>C. Total</td> <td>430</td> <td>3.8061e+15</td> </tr> <tr> <td>Observations (or Sum Wgts)</td> <td>431</td> <td></td> <td></td> <td>F Ratio</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td>24.7695</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td>Prob > F</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td><.0001*</td> </tr> </tbody> </table>	Summary of Fit		Analysis of Variance			RSquare	0.415576	Source	DF	Sum of Squares	RSquare Adj	0.398798	Model	12	1.5817e+15	Root Mean Square Error	2306835	Error	418	2.2244e+15	Mean of Response	1721104	C. Total	430	3.8061e+15	Observations (or Sum Wgts)	431			F Ratio					24.7695					Prob > F					<.0001*	<p>Analysis Of Variance</p> <ul style="list-style-type: none"> • P-Value ≤ 0.05 : Reject H_0 \Rightarrow Model Adequacy • MSE = 5.321×10^{12} <p>Summary of Fit</p> <ul style="list-style-type: none"> • Rsquare = 0.4156 , Rsquare Adj=0.3987 \Rightarrow All of regressors effect to views about 41.56%. Other factors effect about 58.44% 																																																			
Summary of Fit		Analysis of Variance																																																																																																
RSquare	0.415576	Source	DF	Sum of Squares																																																																																														
RSquare Adj	0.398798	Model	12	1.5817e+15																																																																																														
Root Mean Square Error	2306835	Error	418	2.2244e+15																																																																																														
Mean of Response	1721104	C. Total	430	3.8061e+15																																																																																														
Observations (or Sum Wgts)	431			F Ratio																																																																																														
				24.7695																																																																																														
				Prob > F																																																																																														
				<.0001*																																																																																														
	<p>2. Test on Individual Regression coefficients(T-test)</p> <p>Hypothesis $H_0: \beta_i = 0, i=0,1,2,\dots,14$ $H_1: \beta_i \neq 0, i=0,1,2,\dots,14, \alpha = 0.05$</p> <p>Result from JMP</p> <table border="1"> <thead> <tr> <th colspan="6">Parameter Estimates</th> </tr> <tr> <th>Term</th> <th>Estimate</th> <th>Std Error</th> <th>t Ratio</th> <th>Prob> t </th> <th></th> </tr> </thead> <tbody> <tr> <td>Intercept</td> <td>-732974.2</td> <td>1718628</td> <td>-0.43</td> <td>0.6700</td> <td></td> </tr> <tr> <td>comments</td> <td>6100.9766</td> <td>595.8317</td> <td>10.24</td> <td><.0001*</td> <td></td> </tr> <tr> <td>duration</td> <td>700.81473</td> <td>432.352</td> <td>1.62</td> <td>0.1058</td> <td></td> </tr> <tr> <td>num_speaker</td> <td>458442.63551</td> <td>1463.731</td> <td>314.4</td> <td><.0001*</td> <td></td> </tr> <tr> <td>languages</td> <td>122036.91</td> <td>21044.3</td> <td>5.80</td> <td><.0001*</td> <td></td> </tr> <tr> <td>event1</td> <td>-3177534</td> <td>1510138</td> <td>-2.10</td> <td>0.0360*</td> <td></td> </tr> <tr> <td>event2</td> <td>-2863817</td> <td>1463.731</td> <td>-1.96</td> <td>0.0511</td> <td></td> </tr> <tr> <td>event3</td> <td>-3291497</td> <td>1448.631</td> <td>-2.27</td> <td>0.0236*</td> <td></td> </tr> <tr> <td>event4</td> <td>-4942207</td> <td>1476093</td> <td>-3.35</td> <td>0.0009*</td> <td></td> </tr> <tr> <td>event5</td> <td>-4415397</td> <td>1504533</td> <td>-2.93</td> <td>0.0035*</td> <td></td> </tr> <tr> <td>event6</td> <td>-3076091</td> <td>1453.824</td> <td>-2.12</td> <td>0.0349*</td> <td></td> </tr> <tr> <td>event7</td> <td>-3061976</td> <td>1464.624</td> <td>-2.09</td> <td>0.0372*</td> <td></td> </tr> <tr> <td>event8</td> <td>-3023238</td> <td>1470.654</td> <td>-2.06</td> <td>0.0405*</td> <td></td> </tr> <tr> <td>event9</td> <td>Zeroed</td> <td>0</td> <td>.</td> <td>.</td> <td></td> </tr> </tbody> </table>	Parameter Estimates						Term	Estimate	Std Error	t Ratio	Prob> t		Intercept	-732974.2	1718628	-0.43	0.6700		comments	6100.9766	595.8317	10.24	<.0001*		duration	700.81473	432.352	1.62	0.1058		num_speaker	458442.63551	1463.731	314.4	<.0001*		languages	122036.91	21044.3	5.80	<.0001*		event1	-3177534	1510138	-2.10	0.0360*		event2	-2863817	1463.731	-1.96	0.0511		event3	-3291497	1448.631	-2.27	0.0236*		event4	-4942207	1476093	-3.35	0.0009*		event5	-4415397	1504533	-2.93	0.0035*		event6	-3076091	1453.824	-2.12	0.0349*		event7	-3061976	1464.624	-2.09	0.0372*		event8	-3023238	1470.654	-2.06	0.0405*		event9	Zeroed	0	.	.		<p>Parameters Estimates</p> <ul style="list-style-type: none"> • All of Parameters have P-Value ≤ 0.05 : Reject H_0 of All parameters \Rightarrow Some parameter are not on equation.
Parameter Estimates																																																																																																		
Term	Estimate	Std Error	t Ratio	Prob> t																																																																																														
Intercept	-732974.2	1718628	-0.43	0.6700																																																																																														
comments	6100.9766	595.8317	10.24	<.0001*																																																																																														
duration	700.81473	432.352	1.62	0.1058																																																																																														
num_speaker	458442.63551	1463.731	314.4	<.0001*																																																																																														
languages	122036.91	21044.3	5.80	<.0001*																																																																																														
event1	-3177534	1510138	-2.10	0.0360*																																																																																														
event2	-2863817	1463.731	-1.96	0.0511																																																																																														
event3	-3291497	1448.631	-2.27	0.0236*																																																																																														
event4	-4942207	1476093	-3.35	0.0009*																																																																																														
event5	-4415397	1504533	-2.93	0.0035*																																																																																														
event6	-3076091	1453.824	-2.12	0.0349*																																																																																														
event7	-3061976	1464.624	-2.09	0.0372*																																																																																														
event8	-3023238	1470.654	-2.06	0.0405*																																																																																														
event9	Zeroed	0	.	.																																																																																														

MODEL : MATRIX SCATTER PLOT



Correlations

	views	comments	duration	num_speaker	languages	event1	event2	event3	event4	event5	event6	event7	event8	event9
views	1.0000	0.5725	0.0835	0.0169	0.4022	-0.0070	-0.1008	-0.0060	-0.0492	-0.0005	0.0521	0.1235	-0.0499	0.0000
comments	0.5725	1.0000	0.2074	-0.0400	0.4367	-0.0509	-0.1580	-0.0086	0.2102	0.0609	-0.0027	0.0871	-0.1950	0.0000
duration	0.0835	0.2074	1.0000	0.0216	0.2389	0.0693	-0.0162	0.0526	-0.0428	-0.0747	-0.0428	-0.0562	-0.0649	0.0000
num_speaker	0.0169	-0.0400	0.0216	1.0000	0.0051	-0.0274	-0.0280	-0.0457	-0.0469	0.1285	0.141	0.0274	0.0274	0.0000
languages	0.4022	0.4367	-0.2688	0.0051	1.0000	-0.0908	-0.2758	-0.1044	0.0915	0.3346	-0.0299	0.0619	-0.0395	0.0000
event1	-0.0070	r=-0.0509	r=0.2389	r=-0.0274	r=-0.0908	r=-0.0655	r=-0.1068	r=-0.0969	r=-0.1097	r=-0.1116	r=-0.1077	r=-0.0928	r=0.0000	
event2	r=-0.1008	r=-0.1580	r=0.0693	r=-0.0280	r=-0.2758	r=-0.0655	r=-0.1089	r=-0.0989	r=-0.1119	r=-0.1138	r=-0.1099	r=-0.0947	r=0.0000	
event3	r=-0.0060	r=-0.0086	r=-0.0162	r=-0.0457	r=-0.1068	r=-0.0655	r=-0.1612	r=-0.1824	r=-0.1856	r=-0.1792	r=-0.1544	r=0.0000		
event4	r=-0.0492	r=0.2102	r=-0.0526	r=-0.0414	r=-0.0914	r=-0.1068	r=-0.1655	r=-0.1684	r=-0.1626	r=-0.1401	r=0.0000			
event5	r=-0.0005	r=0.0609	r=-0.0428	r=-0.0469	r=-0.0969	r=-0.0989	r=-0.1612	r=-0.1824	r=-0.1856	r=-0.1792	r=-0.1626	r=0.0000		
event6	r=0.0521	r=-0.0027	r=-0.0747	r=0.1285	r=-0.0289	r=-0.1116	r=-0.1138	r=-0.1824	r=-0.1655	r=-0.1906	r=-0.1840	r=-0.1586	r=0.0000	
event7	r=0.1235	r=-0.1950	r=0.0871	r=0.0502	r=0.0141	r=0.0183	r=0.0177	r=0.1099	r=0.1792	r=0.1626	r=0.1840	r=0.1872	r=0.1558	r=0.0000
event8	r=-0.0499	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000
event9	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=0.0000	r=1.0000

Result Matrix Scatter plot as correlation

- The number of views is correlated with the number of comments about 0.5725.
- Other correlation each variable are not more than 0.5



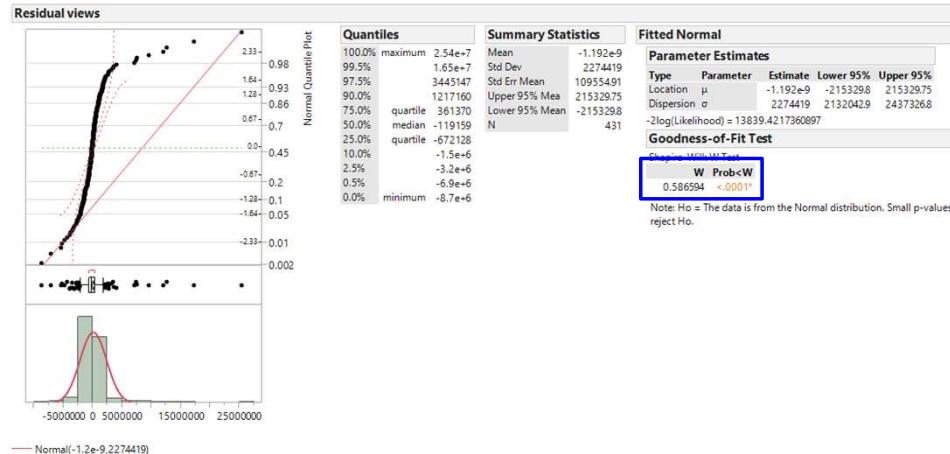
RESIDUAL PLOT

- Normal Probability plot of Residual => Check the normality assumption

Result : Residual Distribution and Studentized Residual Distribution are **not** normally distributed.

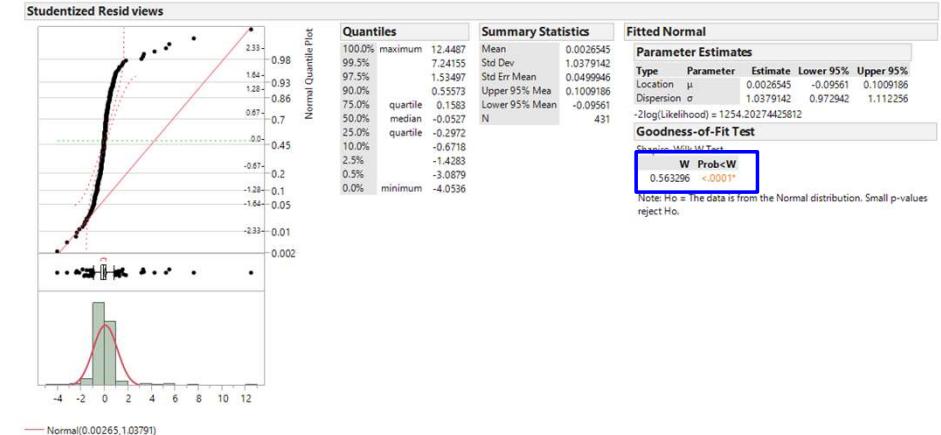
Residual

Hypothesis H_0 : Distribution is normally distributed
 H_1 : Distribution is **not** normally distributed



Studentized Residual

Hypothesis H_0 : Distribution is normally distributed
 H_1 : Distribution is **not** normally distributed



- P-Value ≤ 0.05 : Reject H_0
 \Rightarrow Residual Distribution is **not** normally distributed.

- P-Value ≤ 0.05 : Reject H_0
 \Rightarrow Studentized Residual Distribution is **not** normally Distributed.

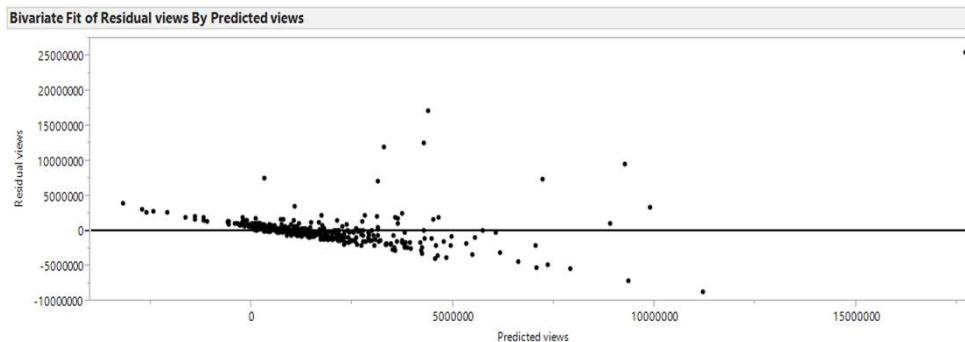


RESIDUAL PLOT

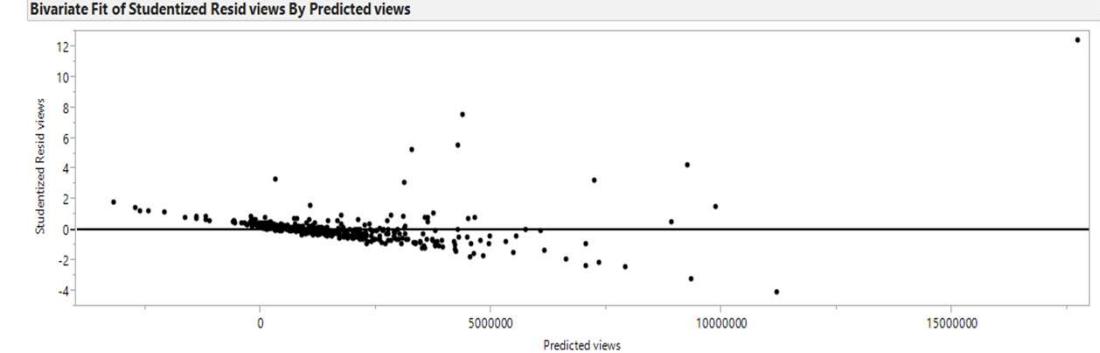
2. Residuals against Fitted values, \hat{Y}_i

Result : Residuals against Fitted values and Studentized Residual against Fitted values have nonconstant variance, non linearity and potential outlier.

Residual



Studentized Residual



Checking for

Non constant variance

Non linearity

Potential outlier

Residual

Yes

Yes

Yes

Studentized Residual

Yes

Yes

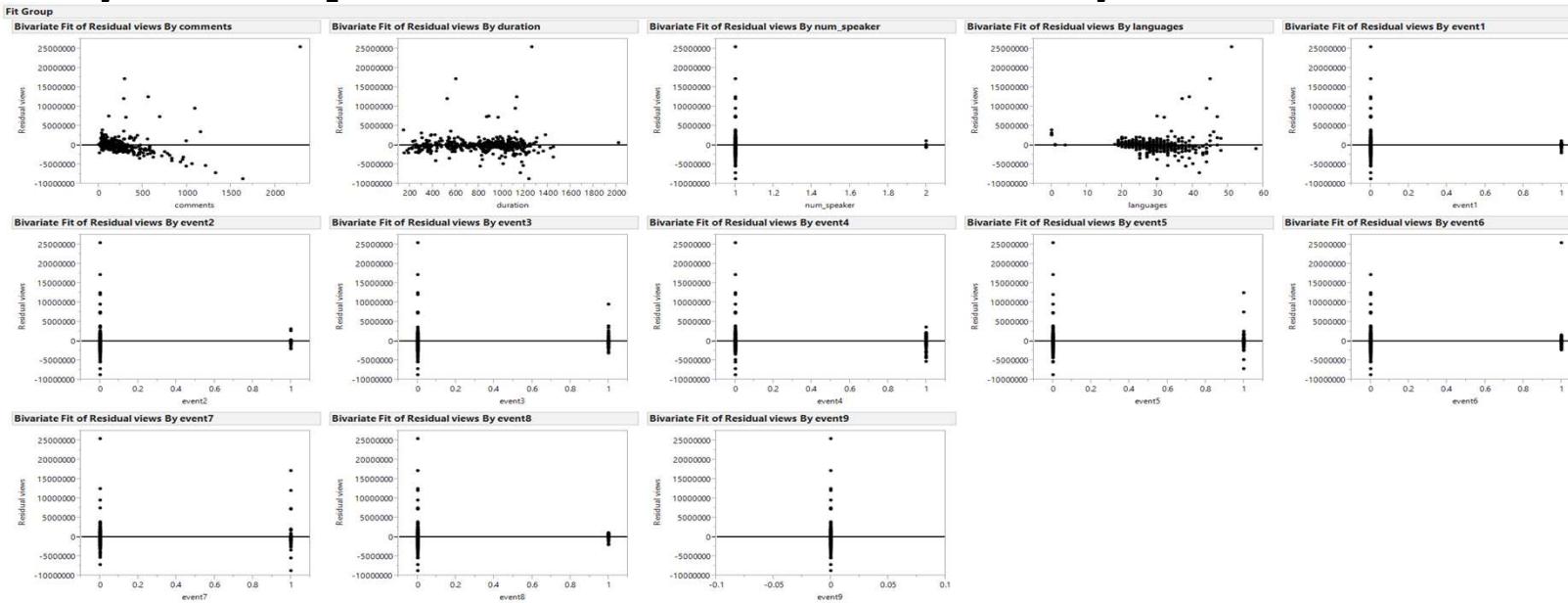
Yes



RESIDUAL PLOT

3. Residuals against Regressors in the model

Result : Comments , durations and languages that are quantitative Variable are non constant variance and non linearity. But num_speaker is constant variance and linearity.



Checking for	comments	durations	num_speaker	language
Non constant variance	Yes	Yes	No	Yes
Non linearity	Yes	Yes	No	Yes

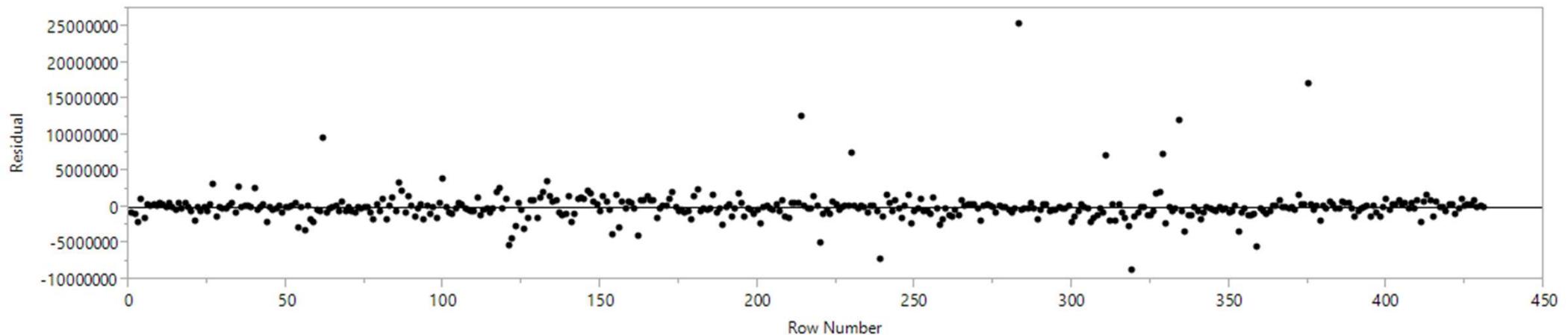


RESIDUAL PLOT ?

4. Residuals against Time order => check for Correlate errors

Result : Residuals are uncorrelated.

Residual by Row Plot



RESIDUAL PLOT SUMMARY

Assumption

1 Error term , ε has zero mean and constant variance

Result ε has no zero mean and non constant variance.

2. Errors are uncorrelated.

Result Errors are uncorrelated.

3 Errors are normally distributed.

Result Errors are not normally distributed.

2/3 are not following assumption.

SO, GOING TO TRANSFORMATION !!!!!!!!!!!!!!!



TRANSFORMATION



PARAMETER TRANSFORMATION

- As model on above , ε has no zero mean and non constant variance that not follow assumption.
- Residual views plot is similar like pattern for residual plot (d) non linear.

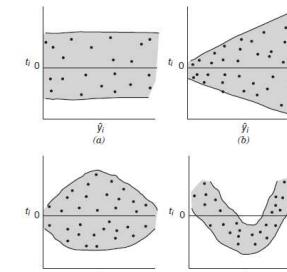
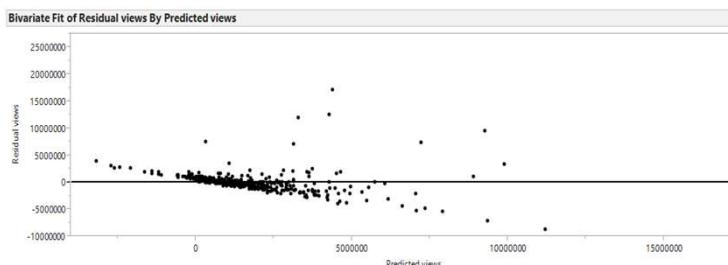


Figure 4.5 Patterns for residual plots: (a) satisfactory; (b) funnel; (c) double bow; (d) nonlinear.

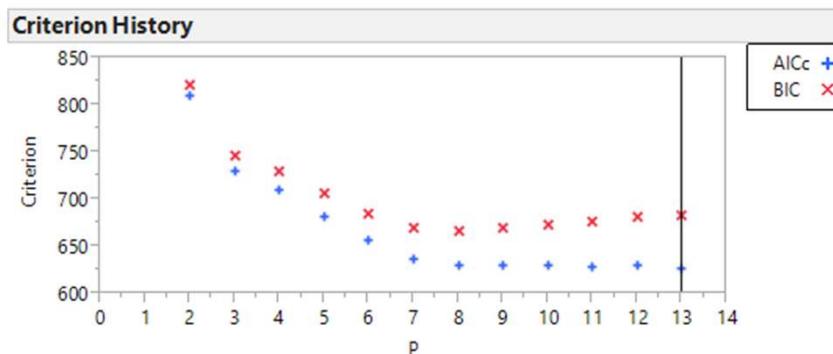
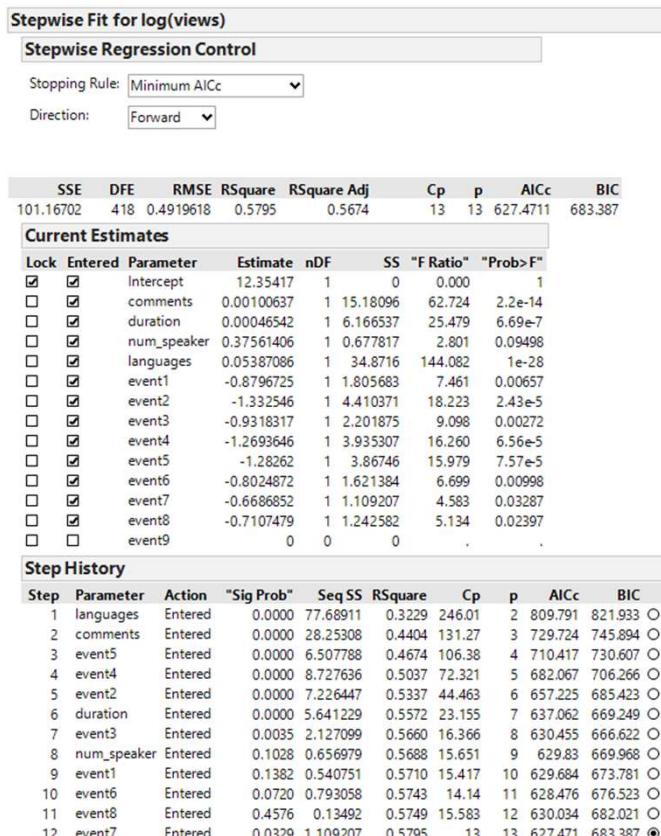
- As Variance-stabilizing Transformations , transform Y into $Y' = \log Y$.

TABLE 5.1 Useful Variance-Stabilizing Transformations

Relationship of σ^2 to $E(y)$	Transformation
$\sigma^2 \propto$ constant	$y' = y$ (no transformation)
$\sigma^2 \propto E(y)$	$y' = \sqrt{y}$ (square root; Poisson data)
$\sigma^2 \propto E(y)[1 - E(y)]$	$y' = \sin^{-1}(\sqrt{y})$ (arcsin; binomial proportions $0 \leq y_i \leq 1$)
$\sigma^2 \propto [E(y)]^2$	$y' = \ln(y)(\text{log})$
$\sigma^2 \propto [E(y)]^3$	$y' = y^{-1/2}$ (reciprocal square root)
$\sigma^2 \propto [E(y)]^4$	$y' = y^{-1}$ (reciprocal)



STEPWISE : FITTING THE MODEL



Result Fitting the model by least AICc =627.471.

$\text{Log}(views) = 12.3541 + 0.0010 * \text{comments} + 0.0005 * \text{durations} + 0.3756 * \text{num_speaker} + 0.0538 * \text{languages} - 0.8797 * \text{event1} - 1.3325 * \text{event2} - 0.9318 * \text{event3} - 1.2694 * \text{event4} - 1.2826 * \text{event5} - 0.8024 * \text{event6} - 0.6687 * \text{event7} - 0.7107 * \text{event8} + \varepsilon$



MODEL: HYPOTHESIS TESTING

Prediction Expression :

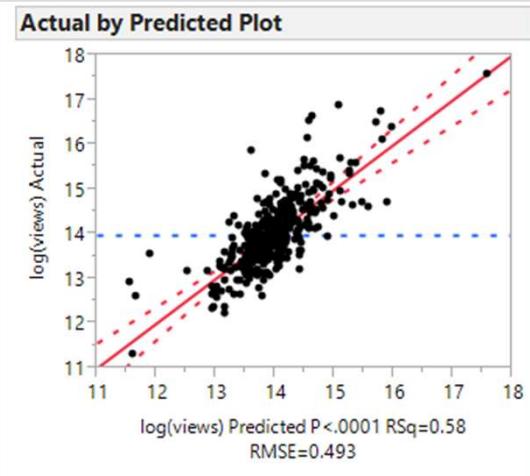
$$\text{Log}(views) = 12.3541 + 0.0010 * \text{comments} + 0.0005 * \text{durations} + 0.3756 * \text{num_speaker} + 0.0538 * \text{languages} - 0.8797 * \text{event1} - 1.3325 * \text{event2} - 0.9318 * \text{event3} - 1.2694 * \text{event4} - 1.2826 * \text{event5} - 0.8024 * \text{event6} - 0.6687 * \text{event7} - 0.7107 * \text{event8}$$

Graph	Hypothesis testing	Result																																																																												
	<p>1. Test for significant of Regression(F-Test)</p> <p>Hypothesis $H_0: \beta_0 = \beta_1 = \dots = \beta_{13} = 0, i=0,1,2,\dots,13$ $H_1: \beta_i \neq 0$ for at least one $i, i=0,1,2,\dots,13$, $\alpha = 0.05$</p> <p>Result from JMP</p> <table border="1"> <thead> <tr> <th colspan="2">Summary of Fit</th> </tr> </thead> <tbody> <tr> <td>RSquare</td> <td>0.579477</td> </tr> <tr> <td>RSquare Adj</td> <td>0.567405</td> </tr> <tr> <td>Root Mean Square Error</td> <td>0.491962</td> </tr> <tr> <td>Mean of Response</td> <td>13.96628</td> </tr> <tr> <td>Observations (or Sum Wgts)</td> <td>431</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="5">Analysis of Variance</th> </tr> <tr> <th>Source</th> <th>DF</th> <th>Sum of Squares</th> <th>Mean Square</th> <th>F Ratio</th> </tr> </thead> <tbody> <tr> <td>Model</td> <td>12</td> <td>139.40730</td> <td>11.6173</td> <td>48.0000</td> </tr> <tr> <td>Error</td> <td>418</td> <td>101.16702</td> <td>0.2420</td> <td>Prob > F</td> </tr> <tr> <td>C. Total</td> <td>430</td> <td>240.57432</td> <td></td> <td><.0001*</td> </tr> </tbody> </table>	Summary of Fit		RSquare	0.579477	RSquare Adj	0.567405	Root Mean Square Error	0.491962	Mean of Response	13.96628	Observations (or Sum Wgts)	431	Analysis of Variance					Source	DF	Sum of Squares	Mean Square	F Ratio	Model	12	139.40730	11.6173	48.0000	Error	418	101.16702	0.2420	Prob > F	C. Total	430	240.57432		<.0001*	<p>Analysis Of Variance</p> <ul style="list-style-type: none"> P-Value ≤ 0.05 : Reject H_0 \Rightarrow Model Adequacy MSE = 0.2420 <p>Summary of Fit</p> <ul style="list-style-type: none"> Rsquare = 0.5795 , Rsquare Adj=0.5674 \Rightarrow All of regressors effect to views about effect to log(views) about 57.95%. Other factors effect about 42.05% 																																							
Summary of Fit																																																																														
RSquare	0.579477																																																																													
RSquare Adj	0.567405																																																																													
Root Mean Square Error	0.491962																																																																													
Mean of Response	13.96628																																																																													
Observations (or Sum Wgts)	431																																																																													
Analysis of Variance																																																																														
Source	DF	Sum of Squares	Mean Square	F Ratio																																																																										
Model	12	139.40730	11.6173	48.0000																																																																										
Error	418	101.16702	0.2420	Prob > F																																																																										
C. Total	430	240.57432		<.0001*																																																																										
	<p>2. Test on Individual Regression coefficients(T-test)</p> <p>Hypothesis $H_0: \beta_i = 0, i=0,1,2,\dots,14$ $H_1: \beta_i \neq 0, i=0,1,2,\dots,14, \alpha = 0.05$</p> <p>Result from JMP</p> <table border="1"> <thead> <tr> <th colspan="6">Parameter Estimates</th> </tr> <tr> <th>Term</th> <th>Estimate</th> <th>Std Error</th> <th>t Ratio</th> <th>Prob> t </th> </tr> </thead> <tbody> <tr> <td>Intercept</td> <td>12.35417</td> <td>0.366519</td> <td>33.71</td> <td><.0001*</td> </tr> <tr> <td>comments</td> <td>0.001064</td> <td>0.000127</td> <td>7.92</td> <td><.0001*</td> </tr> <tr> <td>duration</td> <td>0.0004654</td> <td>9.22e-5</td> <td>5.05</td> <td><.0001*</td> </tr> <tr> <td>num_speaker</td> <td>0.3756141</td> <td>0.224449</td> <td>1.67</td> <td>0.0950</td> </tr> <tr> <td>languages</td> <td>0.0538709</td> <td>0.004488</td> <td>12.00</td> <td><.0001*</td> </tr> <tr> <td>event1</td> <td>-0.879672</td> <td>0.322056</td> <td>-2.73</td> <td>0.0066*</td> </tr> <tr> <td>event2</td> <td>-1.332546</td> <td>0.312159</td> <td>-4.27</td> <td><.0001*</td> </tr> <tr> <td>event3</td> <td>-0.931832</td> <td>0.309899</td> <td>-3.02</td> <td>0.0027*</td> </tr> <tr> <td>event4</td> <td>-1.269965</td> <td>0.314795</td> <td>-4.03</td> <td><.0001*</td> </tr> <tr> <td>event5</td> <td>-1.28262</td> <td>0.320861</td> <td>-4.00</td> <td><.0001*</td> </tr> <tr> <td>event6</td> <td>-0.802487</td> <td>0.310046</td> <td>-2.59</td> <td>0.0100*</td> </tr> <tr> <td>event7</td> <td>-0.668665</td> <td>0.312354</td> <td>-2.14</td> <td>0.0329*</td> </tr> <tr> <td>event8</td> <td>-0.710748</td> <td>0.313678</td> <td>-2.27</td> <td>0.0240*</td> </tr> </tbody> </table>	Parameter Estimates						Term	Estimate	Std Error	t Ratio	Prob> t	Intercept	12.35417	0.366519	33.71	<.0001*	comments	0.001064	0.000127	7.92	<.0001*	duration	0.0004654	9.22e-5	5.05	<.0001*	num_speaker	0.3756141	0.224449	1.67	0.0950	languages	0.0538709	0.004488	12.00	<.0001*	event1	-0.879672	0.322056	-2.73	0.0066*	event2	-1.332546	0.312159	-4.27	<.0001*	event3	-0.931832	0.309899	-3.02	0.0027*	event4	-1.269965	0.314795	-4.03	<.0001*	event5	-1.28262	0.320861	-4.00	<.0001*	event6	-0.802487	0.310046	-2.59	0.0100*	event7	-0.668665	0.312354	-2.14	0.0329*	event8	-0.710748	0.313678	-2.27	0.0240*	<p>Parameters Estimates</p> <ul style="list-style-type: none"> All of Parameters have P-Value ≤ 0.05 : Reject H_0 of All parameters \Rightarrow All of parameter except number of speakers are on equation.
Parameter Estimates																																																																														
Term	Estimate	Std Error	t Ratio	Prob> t																																																																										
Intercept	12.35417	0.366519	33.71	<.0001*																																																																										
comments	0.001064	0.000127	7.92	<.0001*																																																																										
duration	0.0004654	9.22e-5	5.05	<.0001*																																																																										
num_speaker	0.3756141	0.224449	1.67	0.0950																																																																										
languages	0.0538709	0.004488	12.00	<.0001*																																																																										
event1	-0.879672	0.322056	-2.73	0.0066*																																																																										
event2	-1.332546	0.312159	-4.27	<.0001*																																																																										
event3	-0.931832	0.309899	-3.02	0.0027*																																																																										
event4	-1.269965	0.314795	-4.03	<.0001*																																																																										
event5	-1.28262	0.320861	-4.00	<.0001*																																																																										
event6	-0.802487	0.310046	-2.59	0.0100*																																																																										
event7	-0.668665	0.312354	-2.14	0.0329*																																																																										
event8	-0.710748	0.313678	-2.27	0.0240*																																																																										

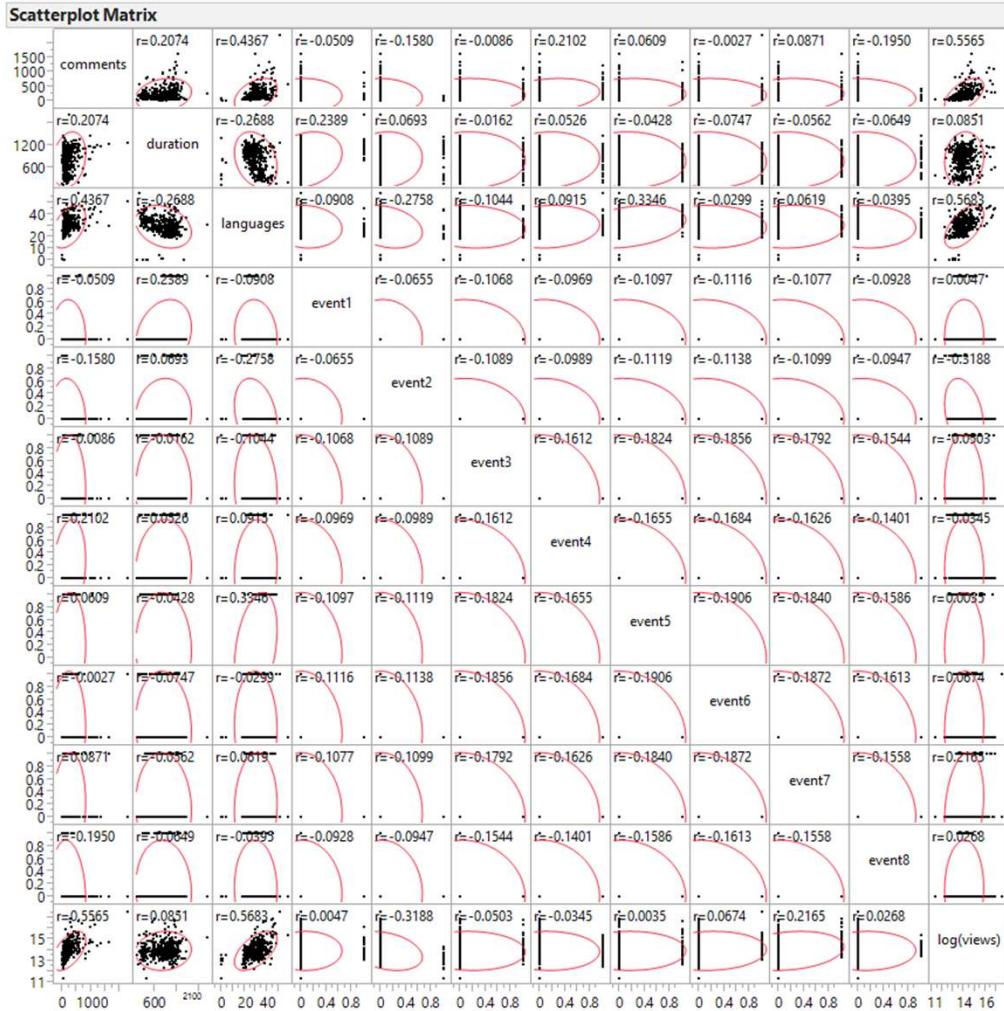
MODEL

Prediction Expression :

$$\text{Log}(views) = 12.7208 + 0.0001 * \text{comments} + 0.0005 * \text{durations} + 0.0542 * \text{languages} - 0.8909 * \text{event1} - 1.3408 * \text{event2} - 0.9396 * \text{event3} - 1.2772 * \text{event4} - 1.2927 * \text{event5} - 0.7942 * \text{event6} - 0.6709 * \text{event7} - 0.7128 * \text{event8}$$

Graph	Hypothesis testing	Result																																																																																				
 <p>Actual by Predicted Plot log(views) Actual vs Predicted P<.0001 RSq=0.58 RMSE=0.493</p>	<p>1. Test for significant of Regression(F-Test)</p> <p>Hypothesis $H_0: \beta_0 = \beta_1 = \dots = \beta_{12} = 0, i=0,1,2,\dots,12$ $H_1: \beta_i \neq 0$ for at least one $i, i=0,1,2,\dots,12$, $\alpha = 0.05$</p> <p>Result from JMP</p> <table border="1" data-bbox="646 747 968 917"> <tr><th colspan="2">Summary of Fit</th></tr> <tr><td>RSquare</td><td>0.57666</td></tr> <tr><td>RSquare Adj</td><td>0.565546</td></tr> <tr><td>Root Mean Square Error</td><td>0.493018</td></tr> <tr><td>Mean of Response</td><td>13.96628</td></tr> <tr><td>Observations (or Sum Wgts)</td><td>431</td></tr> </table> <table border="1" data-bbox="990 747 1365 890"> <tr><th colspan="5">Analysis of Variance</th></tr> <tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Ratio</th></tr> <tr><td>Model</td><td>11</td><td>138.72948</td><td>12.6118</td><td>51.8861</td></tr> <tr><td>Error</td><td>419</td><td>101.84484</td><td>0.2431</td><td>Prob > F <.0001*</td></tr> <tr><td>C. Total</td><td>430</td><td>240.57432</td><td></td><td></td></tr> </table>	Summary of Fit		RSquare	0.57666	RSquare Adj	0.565546	Root Mean Square Error	0.493018	Mean of Response	13.96628	Observations (or Sum Wgts)	431	Analysis of Variance					Source	DF	Sum of Squares	Mean Square	F Ratio	Model	11	138.72948	12.6118	51.8861	Error	419	101.84484	0.2431	Prob > F <.0001*	C. Total	430	240.57432			<p>Analysis Of Variance</p> <ul style="list-style-type: none"> P-Value ≤ 0.05 : Reject H_0 \Rightarrow Model Adequacy MSE = 0.2431 <p>Summary of Fit</p> <ul style="list-style-type: none"> Rsquare = 0.5767 , Rsquare Adj=0.5655 \Rightarrow All of parameter except number of speakers effect to log(views) about 57.67%. Other factors effect about 42.33% 																																															
Summary of Fit																																																																																						
RSquare	0.57666																																																																																					
RSquare Adj	0.565546																																																																																					
Root Mean Square Error	0.493018																																																																																					
Mean of Response	13.96628																																																																																					
Observations (or Sum Wgts)	431																																																																																					
Analysis of Variance																																																																																						
Source	DF	Sum of Squares	Mean Square	F Ratio																																																																																		
Model	11	138.72948	12.6118	51.8861																																																																																		
Error	419	101.84484	0.2431	Prob > F <.0001*																																																																																		
C. Total	430	240.57432																																																																																				
	<p>2. Test on Individual Regression coefficients(T-test)</p> <p>Hypothesis $H_0: \beta_i = 0, i=0,1,2,\dots,12$ $H_1: \beta_i \neq 0, i=0,1,2,\dots,12, \alpha = 0.05$</p> <p>Result from JMP</p> <table border="1" data-bbox="804 1171 1057 1383"> <tr><th colspan="6">Parameter Estimates</th></tr> <tr><th>Term</th><th>Estimate</th><th>Std Error</th><th>t Ratio</th><th>Prob> t </th><th></th></tr> <tr><td>Intercept</td><td>12.720813</td><td>0.294462</td><td>43.20</td><td><.0001*</td><td></td></tr> <tr><td>comments</td><td>0.0009922</td><td>0.000127</td><td>7.81</td><td><.0001*</td><td></td></tr> <tr><td>duration</td><td>0.0004761</td><td>9.218e-5</td><td>5.16</td><td><.0001*</td><td></td></tr> <tr><td>languages</td><td>0.0542746</td><td>0.004491</td><td>12.08</td><td><.0001*</td><td></td></tr> <tr><td>event1</td><td>-0.890861</td><td>0.322678</td><td>-2.76</td><td>0.0060*</td><td></td></tr> <tr><td>event2</td><td>-1.340764</td><td>0.31279</td><td>-4.29</td><td><.0001*</td><td></td></tr> <tr><td>event3</td><td>-0.939572</td><td>0.309567</td><td>-3.04</td><td>0.0026*</td><td></td></tr> <tr><td>event4</td><td>-1.277243</td><td>0.315486</td><td>-4.05</td><td><.0001*</td><td></td></tr> <tr><td>event5</td><td>-1.29274</td><td>0.321492</td><td>-4.02</td><td><.0001*</td><td></td></tr> <tr><td>event6</td><td>-0.794205</td><td>0.310672</td><td>-2.56</td><td>0.0109*</td><td></td></tr> <tr><td>event7</td><td>-0.670873</td><td>0.313021</td><td>-2.14</td><td>0.0327*</td><td></td></tr> <tr><td>event8</td><td>-0.712824</td><td>0.314349</td><td>-2.27</td><td>0.0239*</td><td></td></tr> </table>	Parameter Estimates						Term	Estimate	Std Error	t Ratio	Prob> t		Intercept	12.720813	0.294462	43.20	<.0001*		comments	0.0009922	0.000127	7.81	<.0001*		duration	0.0004761	9.218e-5	5.16	<.0001*		languages	0.0542746	0.004491	12.08	<.0001*		event1	-0.890861	0.322678	-2.76	0.0060*		event2	-1.340764	0.31279	-4.29	<.0001*		event3	-0.939572	0.309567	-3.04	0.0026*		event4	-1.277243	0.315486	-4.05	<.0001*		event5	-1.29274	0.321492	-4.02	<.0001*		event6	-0.794205	0.310672	-2.56	0.0109*		event7	-0.670873	0.313021	-2.14	0.0327*		event8	-0.712824	0.314349	-2.27	0.0239*		<p>Parameters Estimates</p> <ul style="list-style-type: none"> All of Parameters have P-Value ≤ 0.05 \therefore Reject H_0 of All parameters \Rightarrow All of parameter except number of speakers are on equation.
Parameter Estimates																																																																																						
Term	Estimate	Std Error	t Ratio	Prob> t																																																																																		
Intercept	12.720813	0.294462	43.20	<.0001*																																																																																		
comments	0.0009922	0.000127	7.81	<.0001*																																																																																		
duration	0.0004761	9.218e-5	5.16	<.0001*																																																																																		
languages	0.0542746	0.004491	12.08	<.0001*																																																																																		
event1	-0.890861	0.322678	-2.76	0.0060*																																																																																		
event2	-1.340764	0.31279	-4.29	<.0001*																																																																																		
event3	-0.939572	0.309567	-3.04	0.0026*																																																																																		
event4	-1.277243	0.315486	-4.05	<.0001*																																																																																		
event5	-1.29274	0.321492	-4.02	<.0001*																																																																																		
event6	-0.794205	0.310672	-2.56	0.0109*																																																																																		
event7	-0.670873	0.313021	-2.14	0.0327*																																																																																		
event8	-0.712824	0.314349	-2.27	0.0239*																																																																																		

MODEL : MATRIX SCATTER PLOT



Correlations

	comments	duration	languages	event1	event2	event3	event4	event5	event6	event7	event8	log(views)
comments	1.0000	0.2074	0.4367	-0.0509	-0.1580	-0.0086	0.2102	0.0609	-0.0027	0.0871	-0.1950	0.5565
duration	0.2074	1.0000	-0.2688	0.2389	0.0693	-0.0162	0.0526	-0.0428	-0.0747	-0.0562	-0.0649	0.0851
languages	0.4367	-0.2688	1.0000	-0.0908	-0.2758	-0.1044	0.0915	0.3346	-0.0299	0.0619	-0.0395	0.5683
event1	-0.0509	0.2389	-0.0908	1.0000	-0.0655	-0.1068	-0.0969	-0.1097	-0.1116	-0.1077	-0.0928	0.0047
event2	-0.1580	0.0693	-0.2758	-0.0655	1.0000	-0.1089	-0.0989	-0.1119	-0.1138	-0.1099	-0.0947	-0.3188
event3	-0.0086	-0.0162	-0.1044	-0.1068	-0.1089	1.0000	-0.1068	-0.1089	-0.1612	-0.1824	-0.1792	-0.1544
event4	0.2102	0.0526	0.0915	-0.0969	-0.0989	-0.1612	1.0000	-0.1655	-0.1684	-0.1626	-0.1401	-0.0345
event5	0.0609	-0.0428	0.3346	-0.1097	-0.1119	-0.1824	-0.1655	1.0000	-0.1906	-0.1840	-0.1586	0.0035
event6	-0.0027	-0.0747	-0.0299	-0.1116	-0.1138	-0.1856	-0.1684	-0.1906	1.0000	-0.1872	-0.1613	0.0674
event7	0.0871	-0.0562	0.0619	-0.1077	-0.1116	-0.1792	-0.1626	-0.1792	-0.1626	1.0000	-0.1558	0.2165
event8	-0.1950	-0.0649	-0.0395	-0.0947	-0.1099	-0.1792	-0.1626	-0.1544	-0.1401	-0.1586	1.0000	0.0268
log(views)	0.5565	0.0851	0.5683	0.0047	-0.3188	-0.0503	-0.0345	0.0035	-0.0345	0.0674	0.2165	1.0000

Result Matrix Scatter plot as correlation

- Log(views) are correlated with the number of comments about 0.5565 and the number of languages about 0.5683.
- Other correlation each variable are not more than 0.5



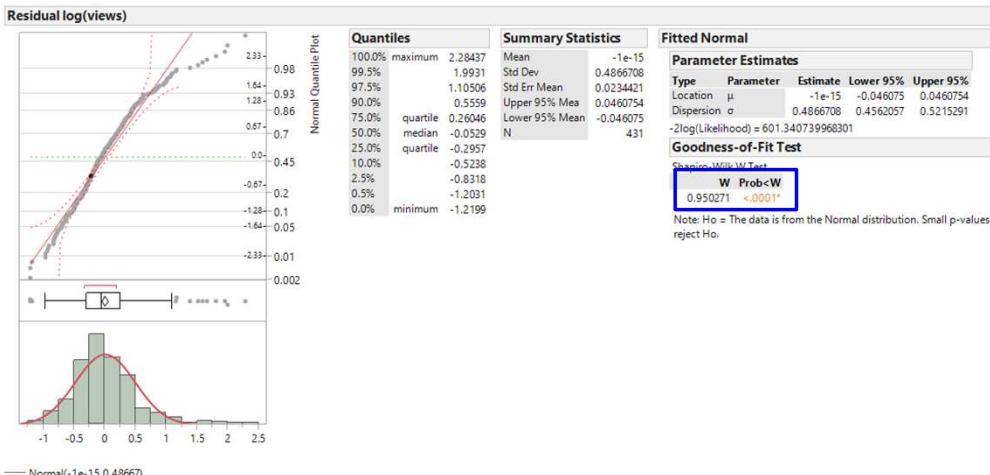
RESIDUAL PLOT

- Normal Probability plot of Residual => Check the normality assumption

Result : Residual Distribution and Studentized Residual Distribution are **not** normally distributed.

Residual

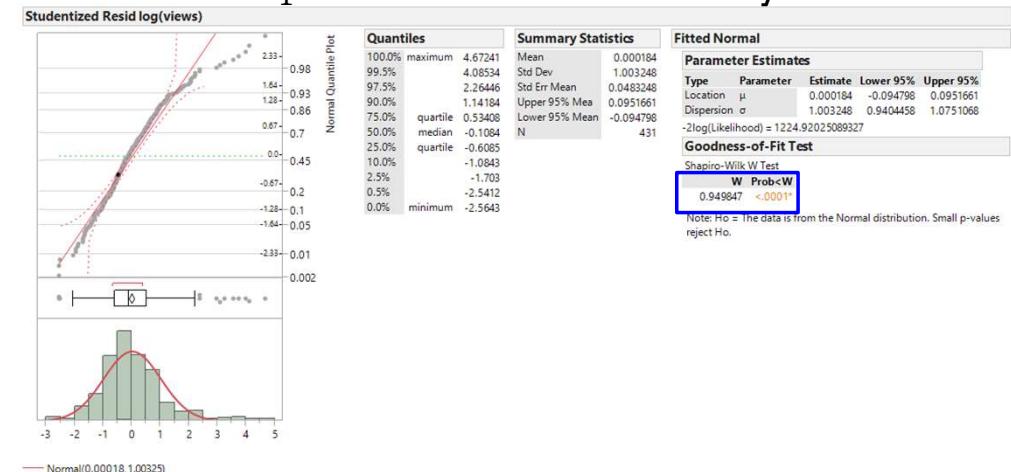
Hypothesis H_0 : Distribution is normally distributed
 H_1 : Distribution is **not** normally distributed



- P-Value ≤ 0.05 : Reject H_0
 \Rightarrow Residual Distribution is **not** normally distributed.

Studentized Residual

Hypothesis H_0 : Distribution is normally distributed
 H_1 : Distribution is **not** normally distributed



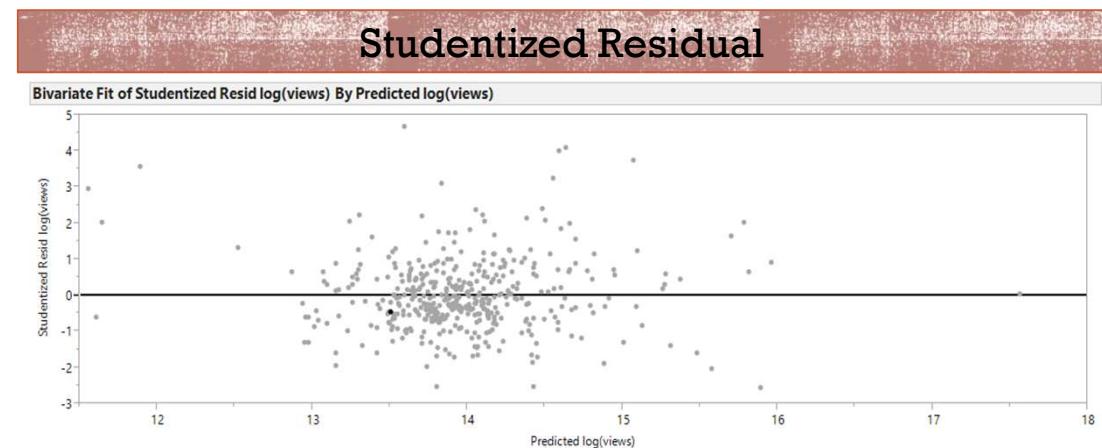
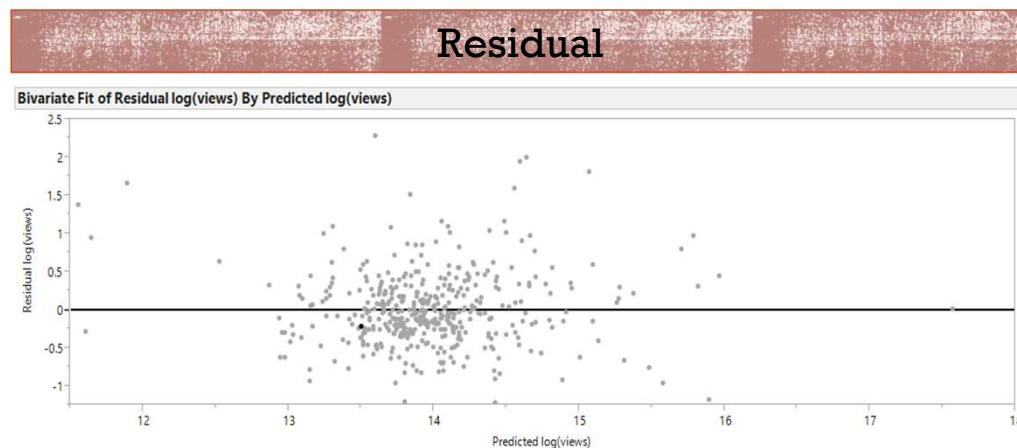
- P-Value ≤ 0.05 : Reject H_0
 \Rightarrow Studentized Residual Distribution is **not** normally Distributed.



RESIDUAL PLOT

2. Residuals against Fitted values, \hat{Y}_i

Result : Residuals against Fitted values and Studentized Residual against Fitted values are constant variance, linearity and potential outlier.



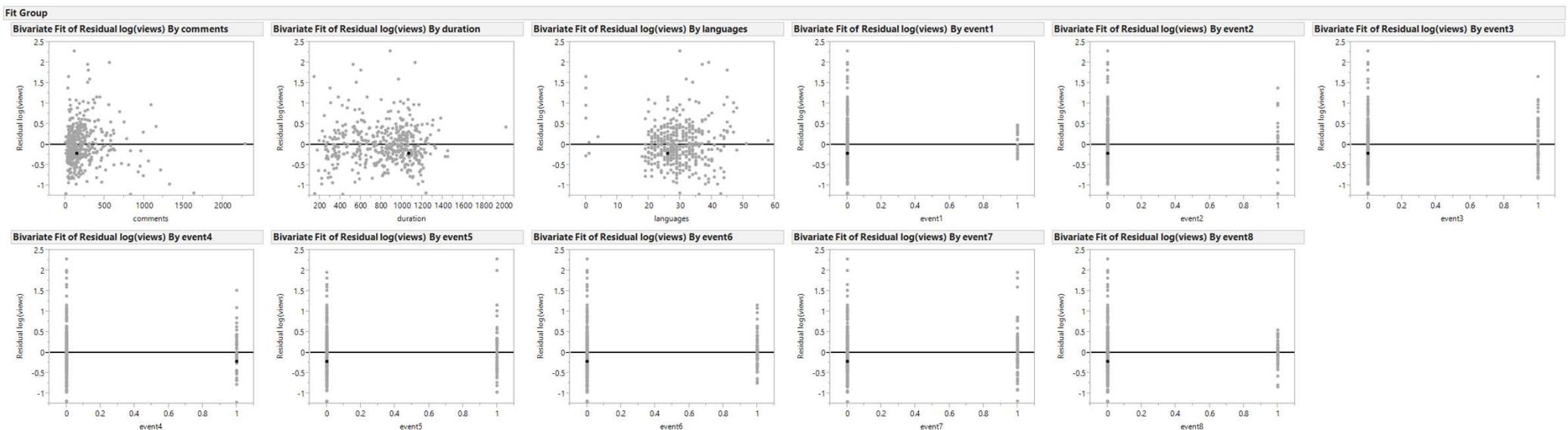
Checking for	Residual	Studentized Residual
Non constant variance	No	No
Non linearity	No	No
Potential outlier	Yes	Yes



RESIDUAL PLOT

3. Residuals against Regressors **in** the model

Result : Comments , durations and languages that are quantitative Variable are constant variance and linearity.



Checking for	comments	durations	language
Non constant variance	No	No	No
Non linearity	No	No	No

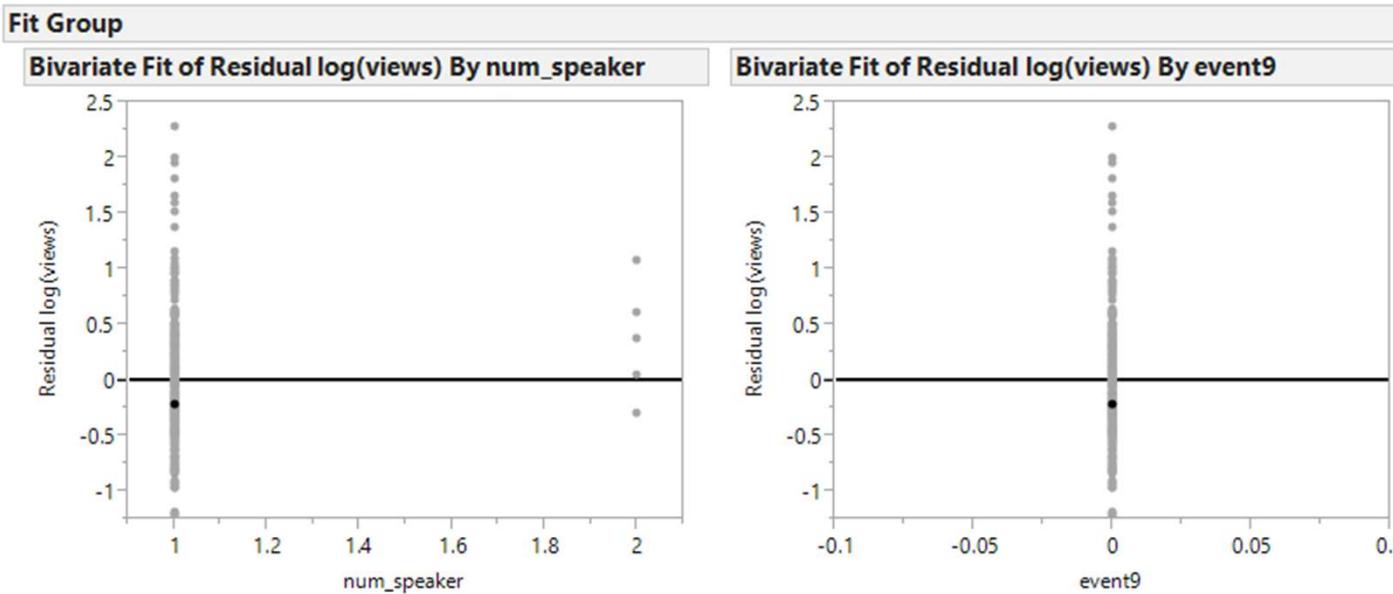


RESIDUAL PLOT ?

4. Residuals against Regressors are **not in** the model

If a pattern appears, could indicate that adding regressor might improve the model fit

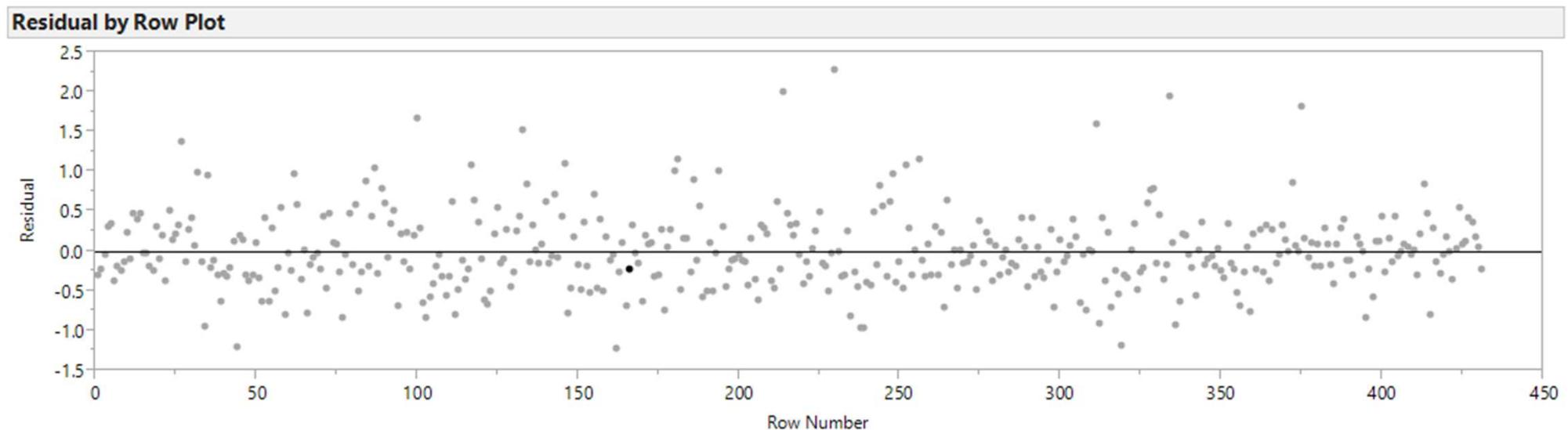
Result : num_speaker are constant variance and linearity. Don't have pattern, so no improve model.



RESIDUAL PLOT ?

5. Residuals against Time order => check for Correlate errors

Result : Residuals are uncorrelated.



RESIDUAL PLOT SUMMARY

Assumption

1 Error term , ε has zero mean and constant variance

Result ε has zero mean and constant variance.

2. Errors are uncorrelated.

Result Errors are uncorrelated.

3 Errors are normally distributed.

Result Errors are not normally distributed.

1/3 are not following assumption.

SO, GOING TO DIAGNOSTICS FOR LEVERAGE AND INFLUENCE !!!!!!!!!!!!!!!

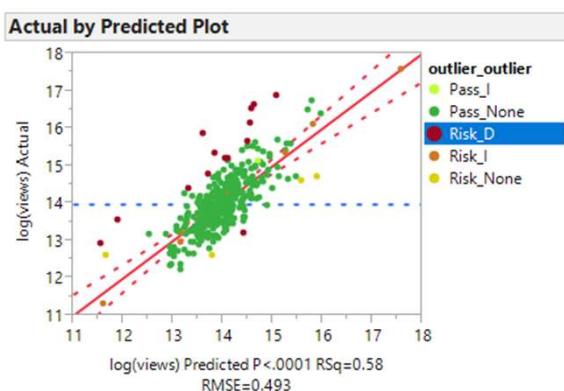
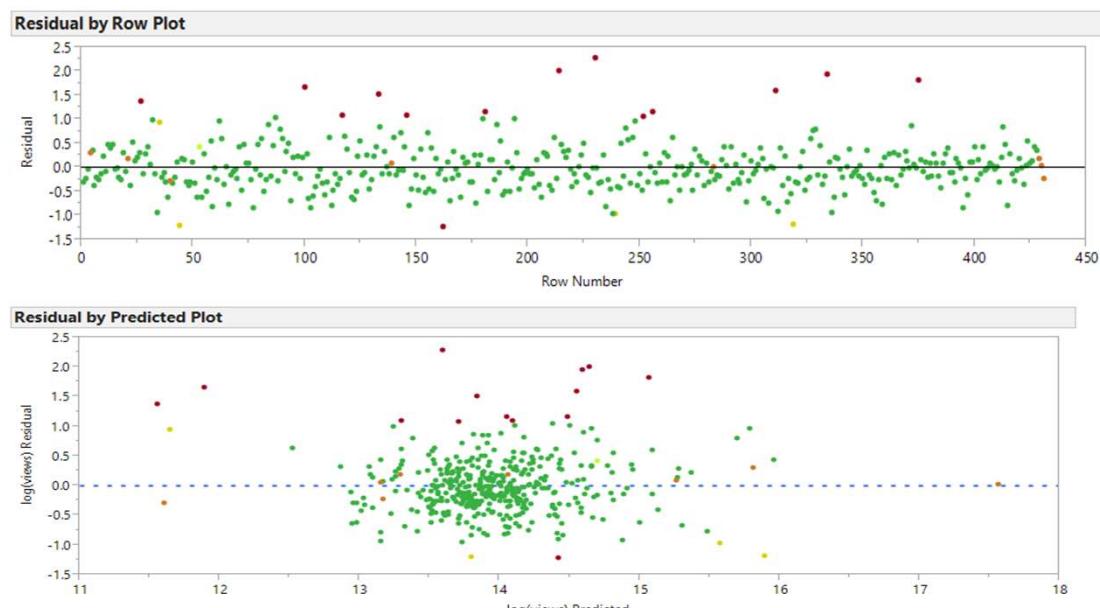


DIAGNOSTICS FOR LEVERAGE AND INFLUENCE



DIAGNOSTICS FOR LEVERAGE AND INFLUENCE

Statistics for detect influential	Cut off	Cut off Number
cook.d	1	1
cov.r	improve:1+3p/n Degrade:1-3p/n	1.083526682 0.916473318
dfb.1	$\frac{2}{\sqrt{n}}$	0.09633661
dfb.X1		
dfb.X10		
dfb.X11		
dfb.X12		
dfb.X13		
dfb.X2		
dfb.X4		
dfb.X6		
dfb.X7		
dfb.X8		
dfb.X9		
dffit	$2\sqrt{\frac{p}{n}}$	0.333719807
hat	$\frac{2p}{n}$	0.055684455



Result: cutting group “Risk” and “Degrade mode” about 14 data point out.
 Separate into 2 group by R program as
 “*”= Risk and “ ” = Pass
 and also cut off criteria by Cov.r as
 “I” =improve model (1+3p/n)
 “D” =Degrade model(1-3p/n)

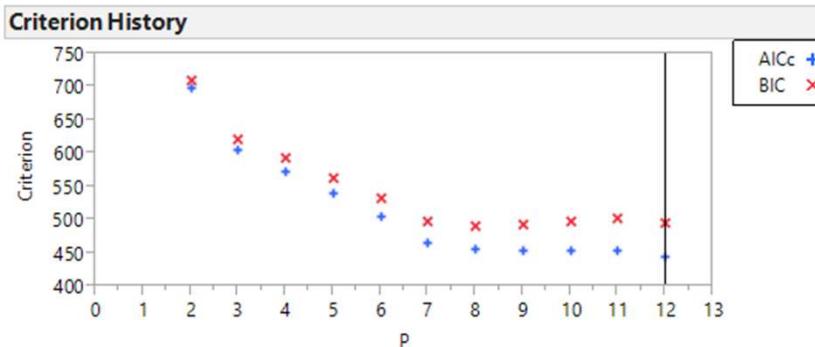
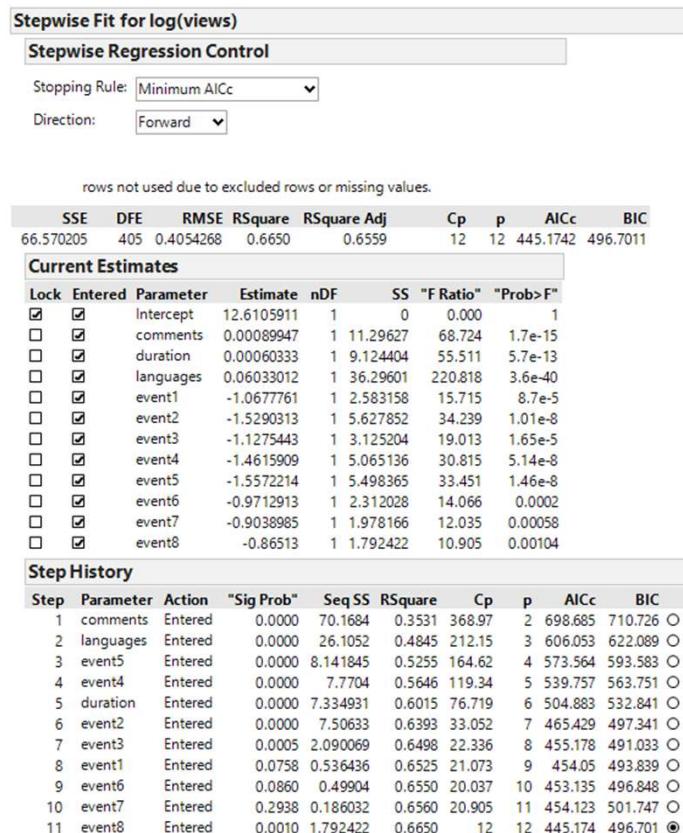


DIAGNOSTICS FOR LEVERAGE AND INFLUENCE

ID	dfb.1_	dfb.X1	dfb.X2	dfb.X4	dfb.X6	dfb.X7	dfb.X8	dfb.X9	dfb.X10	dfb.X11	dfb.X12	dfb.X13	dffit	cov.r	cook.d	hat	inf	cov_outlier	inf_outlier	outlier_out
4	0.00141	0.08060	0.00059	0.04300	0.01890	0.01420	0.02280	0.02900	0.02780	0.02370	0.02620	0.01920	0.18132	1.10000	0.00274	0.07490	*	I	Risk	Risk_I
21	0.00272	0.01250	0.01840	0.05830	0.01750	0.00889	0.02040	0.02210	0.02580	0.02160	0.02250	0.02180	0.10965	1.10300	0.00100	0.07090	*	I	Risk	Risk_I
27	0.16900	0.42200	0.60600	0.75000	0.28800	0.40400	0.24900	0.27000	0.31100	0.25500	0.26400	0.26900	1.01628	0.89100	0.08450	0.10370	*	D	Risk	Risk_D
35	0.10400	0.28400	0.36900	0.49700	0.18800	0.26900	0.16400	0.17700	0.20500	0.16800	0.17400	0.17800	0.66863	1.01300	0.03700	0.09760	*	None	Risk	Risk_None
40	0.03180	0.08290	0.11300	0.14900	0.05670	0.08090	0.04940	0.05350	0.06170	0.05070	0.05240	0.05350	0.20142	1.12900	0.00339	0.09820	*	I	Risk	Risk_I
44	0.02140	0.15800	0.13500	0.38600	0.11000	0.05730	0.12800	0.13600	0.16400	0.13700	0.14100	0.14200	0.72206	0.92200	0.04290	0.07370	*	None	Risk	Risk_None
100	0.23100	0.51200	0.82300	1.08000	0.41800	0.31800	0.46000	0.40100	0.45700	0.37600	0.39000	0.39200	1.25572	0.79600	0.12800	0.10740	*	D	Risk	Risk_D
117	0.01200	0.01040	0.03530	0.11200	0.04260	0.03270	0.09820	0.04900	0.05320	0.04440	0.04700	0.04290	0.31177	0.90900	0.00802	0.01900	*	D	Risk	Risk_D
133	0.03080	0.01820	0.13000	0.02710	0.00424	0.00212	0.00714	0.08470	0.01020	0.00916	0.00892	0.00994	0.45789	0.79500	0.01710	0.02090	*	D	Risk	Risk_D
139	0.00124	0.00272	0.00908	0.03190	0.00995	0.00846	0.01150	0.00702	0.01450	0.01220	0.01280	0.01210	0.04968	1.09600	0.00021	0.06210	*	I	Risk	Risk_I
146	0.02590	0.04700	0.09830	0.06960	0.03030	0.02180	0.02280	0.03950	0.02830	0.02290	0.02370	0.02460	0.32314	0.90900	0.00862	0.02020	*	D	Risk	Risk_D
162	0.07540	0.35300	0.29700	0.14300	0.05460	0.04130	0.02610	0.09240	0.03370	0.02480	0.02180	0.04150	0.53413	0.89300	0.02350	0.04220	*	D	Risk	Risk_D
181	0.03230	0.08320	0.14300	0.07570	0.01780	0.01790	0.03370	0.03930	0.01220	0.03690	0.03960	0.03200	0.37389	0.89400	0.01150	0.02370	*	D	Risk	Risk_D
214	0.05010	0.15700	0.19300	0.08620	0.05830	0.03560	0.04740	0.06260	0.03910	0.04680	0.05230	0.03790	0.61588	0.64300	0.03040	0.02120	*	D	Risk	Risk_D
230	0.00924	0.08220	0.05320	0.11300	0.03900	0.03160	0.04820	0.05690	0.16900	0.05100	0.05470	0.04580	0.62294	0.55100	0.03070	0.01660	*	D	Risk	Risk_D
239	0.01010	0.46600	0.03390	0.10700	0.00956	0.01800	0.00008	0.01910	0.05240	0.00113	0.00575	0.02410	0.56828	0.98200	0.02670	0.07110	*	None	Risk	Risk_None
252	0.03340	0.03280	0.12700	0.08740	0.04010	0.02820	0.03100	0.03520	0.03810	0.08150	0.03260	0.03180	0.29948	0.91200	0.00741	0.01810	*	D	Risk	Risk_D
256	0.03370	0.16800	0.12300	0.14400	0.05020	0.03990	0.04030	0.03990	0.05110	0.09590	0.04130	0.04850	0.34294	0.89300	0.00969	0.02020	*	D	Risk	Risk_D
283	0.00028	0.01320	0.00119	0.00101	0.00040	0.00004	0.00074	0.00137	0.00081	0.00001	0.00098	0.00008	0.01548	1.31300	0.00002	0.21600	*	I	Risk	Risk_I
311	0.03130	0.03350	0.12000	0.07530	0.03520	0.02460	0.02630	0.02990	0.03240	0.02630	0.05170	0.02720	0.42792	0.76900	0.01490	0.01660	*	D	Risk	Risk_D
319	0.04880	0.88900	0.16100	0.40800	0.09790	0.09490	0.07710	0.05130	0.10400	0.08170	0.13900	0.12400	0.97483	0.97300	0.07810	0.12480	*	None	Risk	Risk_None
334	0.02000	0.02640	0.10100	0.14900	0.04000	0.03690	0.05200	0.05580	0.06630	0.05590	0.03990	0.05600	0.55726	0.65800	0.02500	0.01840	*	D	Risk	Risk_D
375	0.03010	0.17600	0.07130	0.43700	0.14600	0.11900	0.14800	0.15900	0.18600	0.15500	0.06840	0.16000	0.65749	0.70500	0.03490	0.02910	*	D	Risk	Risk_D
429	0.30500	0.00274	0.00124	0.00759	0.28100	0.28900	0.29300	0.28800	0.28300	0.29200	0.29000	0.28900	0.31589	1.53500	0.00833	0.33360	*	I	Risk	Risk_I
430	0.07910	0.00031	0.00001	0.00100	0.07190	0.07420	0.07490	0.07350	0.07200	0.07460	0.07410	0.07370	0.08186	1.54300	0.00056	0.33340	*	I	Risk	Risk_I
431	0.38400	0.00196	0.00150	0.00468	0.34900	0.36100	0.36400	0.35700	0.35000	0.36300	0.36000	0.35800	0.39764	1.53000	0.01320	0.33340	*	I	Risk	Risk_I



STEPWISE : FITTING THE MODEL



Result Fitting the model by least AICc =445.174.

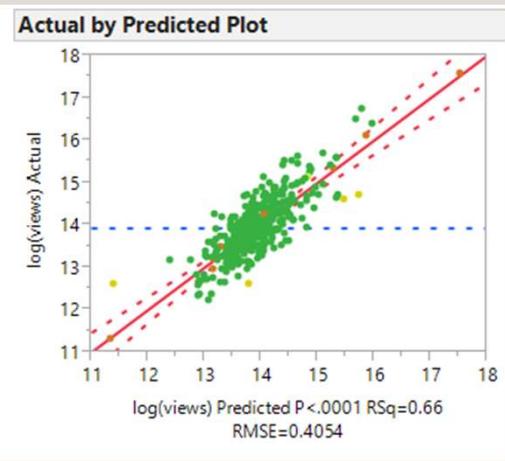
$\text{Log(views)} = 12.7208 + 0.0010 * \text{comments} + 0.0005 * \text{durations}$
 $+ 0.0542 * \text{languages} - 0.8908 * \text{event1} - 1.3408 * \text{event2} -$
 $0.9396 * \text{event3} - 1.2772 * \text{event4} - 1.2927 * \text{event5} -$
 $0.7942 * \text{event6} - 0.6708 * \text{event7} - 0.7128 * \text{event8}$



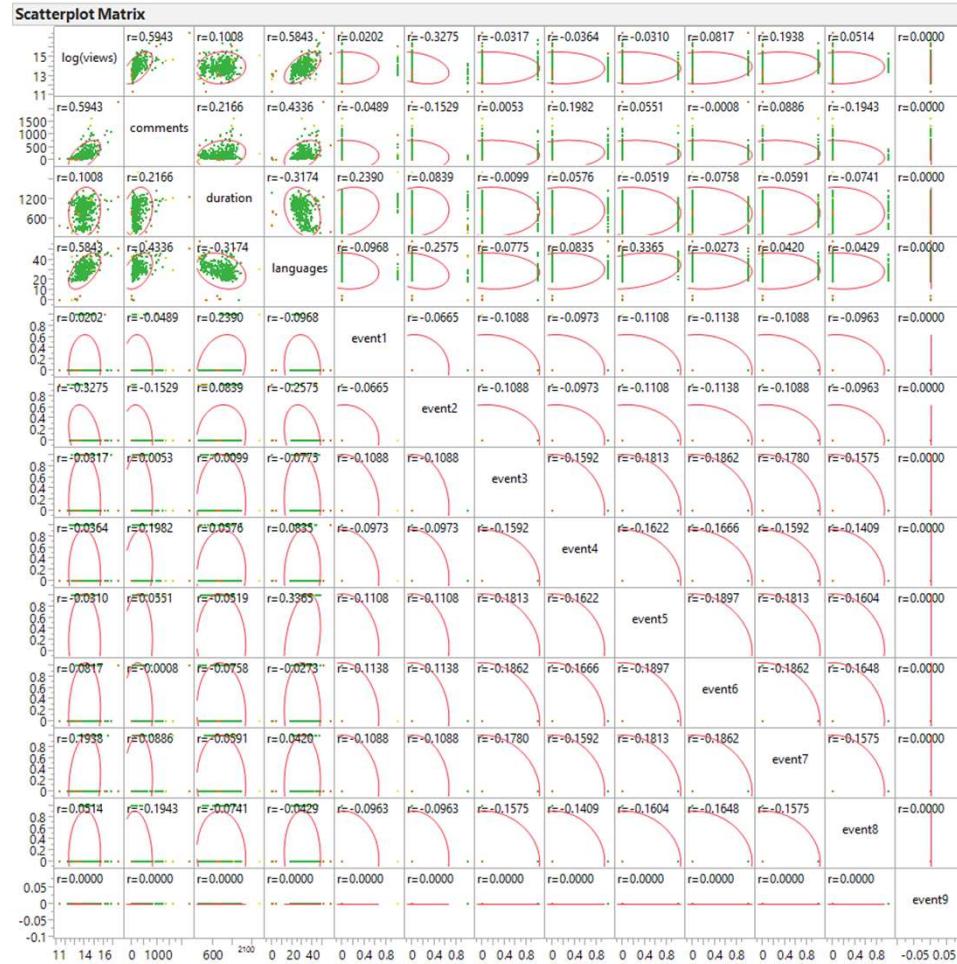
MODEL :HYPOTHESIS TESTING

Prediction Expression :

$$\log(\text{views}) = 12.7208 + 0.0010 * \text{comments} + 0.0005 * \text{durations} + 0.0542 * \text{languages} - 0.8908 * \text{event1} - 1.3408 * \text{event2} - 0.9396 * \text{event3} - 1.2772 * \text{event4} - 1.2927 * \text{event5} - 0.7942 * \text{event6} - 0.6708 * \text{event7} - 0.7128 * \text{event8}$$

Graph	Hypothesis testing	Result																																																																																				
 <p>Actual by Predicted Plot log(views) Actual vs log(views) Predicted P<.0001 RSq=0.66 RMSE=0.4054</p>	<p>1. Test for significant of Regression(F-Test)</p> <p>Hypothesis $H_0: \beta_0 = \beta_1 = \dots = \beta_{12} = 0, i=0,1,2,\dots,12$ $H_1: \beta_i \neq 0$ for at least one $i, i=0,1,2,\dots,12$, $\alpha = 0.05$</p> <p><u>Result from JMP</u></p> <table border="1" data-bbox="662 762 954 913"> <tr><th colspan="2">Summary of Fit</th></tr> <tr><td>RSquare</td><td>0.664973</td></tr> <tr><td>RSquare Adj</td><td>0.655874</td></tr> <tr><td>Root Mean Square Error</td><td>0.405427</td></tr> <tr><td>Mean of Response</td><td>13.92587</td></tr> <tr><td>Observations (or Sum Wgts)</td><td>417</td></tr> </table> <table border="1" data-bbox="999 762 1403 913"> <tr><th colspan="5">Analysis of Variance</th></tr> <tr> <th>Source</th> <th>DF</th> <th>Sum of Squares</th> <th>Mean Square</th> <th>F Ratio</th> </tr> <tr> <td>Model</td> <td>11</td> <td>132.13110</td> <td>12.0119</td> <td>73.0781</td> </tr> <tr> <td>Error</td> <td>405</td> <td>66.57021</td> <td>0.1644</td> <td>Prob > F</td> </tr> <tr> <td>C. Total</td> <td>416</td> <td>198.70130</td> <td></td> <td><.0001*</td> </tr> </table>	Summary of Fit		RSquare	0.664973	RSquare Adj	0.655874	Root Mean Square Error	0.405427	Mean of Response	13.92587	Observations (or Sum Wgts)	417	Analysis of Variance					Source	DF	Sum of Squares	Mean Square	F Ratio	Model	11	132.13110	12.0119	73.0781	Error	405	66.57021	0.1644	Prob > F	C. Total	416	198.70130		<.0001*	<p><u>Analysis Of Variance</u></p> <ul style="list-style-type: none"> • P-Value ≤ 0.05 : Reject H_0 \Rightarrow Model Adequacy • MSE = 0.1644 <p><u>Summary of Fit</u></p> <ul style="list-style-type: none"> • Rsquare = 0.6650 , Rsquare Adj=0.6559 \Rightarrow All of parameter except number of speakers effect to log(views) about 66.50%. Other factors effect about 33.50% 																																															
Summary of Fit																																																																																						
RSquare	0.664973																																																																																					
RSquare Adj	0.655874																																																																																					
Root Mean Square Error	0.405427																																																																																					
Mean of Response	13.92587																																																																																					
Observations (or Sum Wgts)	417																																																																																					
Analysis of Variance																																																																																						
Source	DF	Sum of Squares	Mean Square	F Ratio																																																																																		
Model	11	132.13110	12.0119	73.0781																																																																																		
Error	405	66.57021	0.1644	Prob > F																																																																																		
C. Total	416	198.70130		<.0001*																																																																																		
	<p>2. Test on Individual Regression coefficients(T-test)</p> <p>Hypothesis $H_0: \beta_i = 0, i=0,1,2,\dots,12$ $H_1: \beta_i \neq 0, i=0,1,2,\dots,12, \alpha = 0.05$</p> <p><u>Result from JMP</u></p> <table border="1" data-bbox="797 1167 1066 1381"> <tr><th colspan="6">Parameter Estimates</th></tr> <tr> <th>Term</th> <th>Estimate</th> <th>Std Error</th> <th>t Ratio</th> <th>Prob> t </th> <th></th> </tr> <tr> <td>Intercept</td> <td>12.610591</td> <td>0.243433</td> <td>51.80</td> <td><.0001*</td> <td></td> </tr> <tr> <td>comments</td> <td>0.0008995</td> <td>0.000109</td> <td>8.29</td> <td><.0001*</td> <td></td> </tr> <tr> <td>duration</td> <td>0.0006033</td> <td>0.000081</td> <td>7.45</td> <td><.0001*</td> <td></td> </tr> <tr> <td>languages</td> <td>0.0603301</td> <td>0.00406</td> <td>14.86</td> <td><.0001*</td> <td></td> </tr> <tr> <td>event1</td> <td>-1.067776</td> <td>0.26935</td> <td>-3.96</td> <td><.0001*</td> <td></td> </tr> <tr> <td>event2</td> <td>-1.52903</td> <td>0.26131</td> <td>-5.85</td> <td><.0001*</td> <td></td> </tr> <tr> <td>event3</td> <td>-1.127544</td> <td>0.258587</td> <td>-4.36</td> <td><.0001*</td> <td></td> </tr> <tr> <td>event4</td> <td>-1.46159</td> <td>0.263395</td> <td>-5.55</td> <td><.0001*</td> <td></td> </tr> <tr> <td>event5</td> <td>-1.557221</td> <td>0.269244</td> <td>-5.78</td> <td><.0001*</td> <td></td> </tr> <tr> <td>event6</td> <td>-0.971291</td> <td>0.25698</td> <td>-3.75</td> <td>0.0002*</td> <td></td> </tr> <tr> <td>event7</td> <td>-0.903899</td> <td>0.260556</td> <td>-3.47</td> <td>0.0006*</td> <td></td> </tr> <tr> <td>event8</td> <td>-0.86513</td> <td>0.261983</td> <td>-3.30</td> <td>0.0010*</td> <td></td> </tr> </table>	Parameter Estimates						Term	Estimate	Std Error	t Ratio	Prob> t		Intercept	12.610591	0.243433	51.80	<.0001*		comments	0.0008995	0.000109	8.29	<.0001*		duration	0.0006033	0.000081	7.45	<.0001*		languages	0.0603301	0.00406	14.86	<.0001*		event1	-1.067776	0.26935	-3.96	<.0001*		event2	-1.52903	0.26131	-5.85	<.0001*		event3	-1.127544	0.258587	-4.36	<.0001*		event4	-1.46159	0.263395	-5.55	<.0001*		event5	-1.557221	0.269244	-5.78	<.0001*		event6	-0.971291	0.25698	-3.75	0.0002*		event7	-0.903899	0.260556	-3.47	0.0006*		event8	-0.86513	0.261983	-3.30	0.0010*		<p><u>Parameters Estimates</u></p> <ul style="list-style-type: none"> • All of Parameters have P-Value ≤ 0.05 : Reject H_0 of All parameters \Rightarrow All of parameter except number of speakers are on equation.
Parameter Estimates																																																																																						
Term	Estimate	Std Error	t Ratio	Prob> t																																																																																		
Intercept	12.610591	0.243433	51.80	<.0001*																																																																																		
comments	0.0008995	0.000109	8.29	<.0001*																																																																																		
duration	0.0006033	0.000081	7.45	<.0001*																																																																																		
languages	0.0603301	0.00406	14.86	<.0001*																																																																																		
event1	-1.067776	0.26935	-3.96	<.0001*																																																																																		
event2	-1.52903	0.26131	-5.85	<.0001*																																																																																		
event3	-1.127544	0.258587	-4.36	<.0001*																																																																																		
event4	-1.46159	0.263395	-5.55	<.0001*																																																																																		
event5	-1.557221	0.269244	-5.78	<.0001*																																																																																		
event6	-0.971291	0.25698	-3.75	0.0002*																																																																																		
event7	-0.903899	0.260556	-3.47	0.0006*																																																																																		
event8	-0.86513	0.261983	-3.30	0.0010*																																																																																		

MODEL : MATRIX SCATTER PLOT



Correlations

	log(views)	comments	duration	languages	event1	event2	event3	event4	event5	event6	event7	event8	event9
log(views)	1.0000	0.5943	0.1008	0.5843	0.0202	-0.3275	-0.0317	-0.0364	-0.0310	0.0817	0.1938	0.0514	0.0000
comments	0.5943	1.0000	0.2166	0.4336	-0.0489	-0.1529	0.0053	0.1982	0.0551	-0.0008	0.0886	-0.1943	0.0000
duration	0.1008	0.2166	1.0000	-0.3174	-0.2390	-0.0839	-0.0099	0.0576	-0.0519	-0.0758	-0.0591	-0.0741	0.0000
languages	0.5843	0.4336	-0.3174	1.0000	-0.0519	0.0551	-0.0008	0.0866	-0.1943	0.0514	-0.0000	-0.0429	0.0000
event1	0.0202	-0.0489	0.2390	-0.0519	1.0000	-0.0665	-0.1088	-0.0973	-0.1108	-0.1138	-0.1088	-0.0963	0.0000
event2	-0.3275	-0.1529	0.0839	-0.2575	-0.0665	1.0000	-0.1088	-0.1088	-0.1088	-0.1088	-0.1138	-0.1088	-0.0963
event3	-0.0317	0.0053	-0.0999	-0.0775	-0.1088	-0.0973	1.0000	-0.1592	-0.1813	-0.1862	-0.1780	-0.1575	0.0000
event4	-0.0364	0.1982	0.0576	0.0835	-0.1592	-0.1813	-0.1862	1.0000	-0.1592	-0.1622	-0.1666	-0.1592	0.0000
event5	-0.0310	0.0551	-0.0519	0.3365	-0.1108	-0.1108	-0.1813	-0.1622	1.0000	-0.1897	-0.1813	-0.1604	0.0000
event6	0.0817	-0.0008	-0.0758	-0.0273	-0.1138	-0.1088	-0.1897	-0.1813	-0.1897	1.0000	-0.1862	-0.1648	0.0000
event7	0.1938	0.0886	0.0591	0.0420	-0.1088	-0.1780	-0.1592	-0.1813	-0.1862	-0.1575	1.0000	-0.1575	0.0000
event8	0.0514	0.1943	-0.0741	-0.0429	-0.0963	-0.0963	-0.1575	-0.1409	-0.1604	-0.1648	-0.1575	1.0000	0.0000
event9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Result Matrix Scatter plot as correlation

- Log(views) are correlated with the number of comments about 0.5943 and the number of languages about 0.5843.
- Other correlation each variable are not more than 0.5



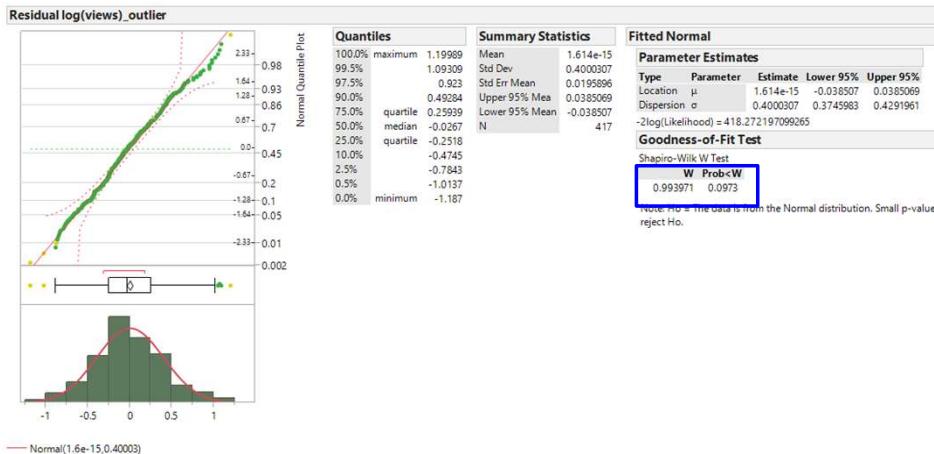
RESIDUAL PLOT

- Normal Probability plot of Residual => Check the normality assumption

Result : Residual Distribution and Studentized Residual Distribution are normally distributed.

Residual

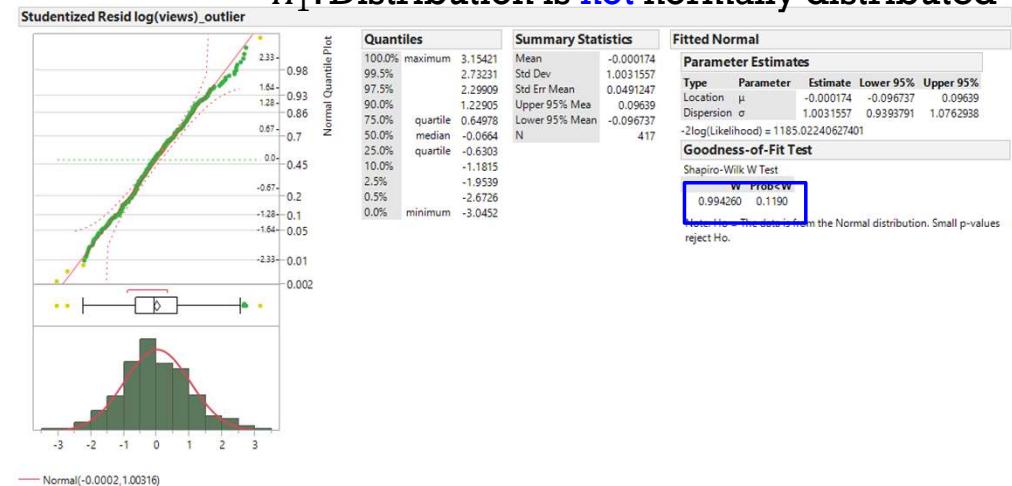
Hypothesis H_0 : Distribution is normally distributed
 H_1 : Distribution is **not** normally distributed



- P-Value > 0.05 : Not Reject H_0
=> Residual Distribution is normally distributed.

Studentized Residual

Hypothesis H_0 : Distribution is normally distributed
 H_1 : Distribution is **not** normally distributed



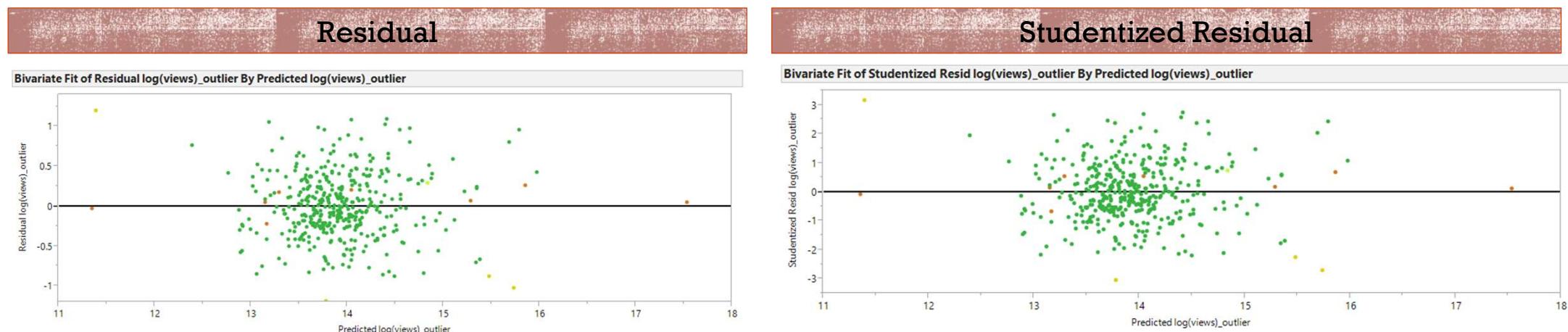
- P-Value > 0.05 : Not Reject H_0
=> Studentized Residual Distribution is normally Distributed.



RESIDUAL PLOT

2. Residuals against Fitted values, \hat{Y}_i

Result : Residuals against Fitted values and Studentized Residual against Fitted values are constant variance, linearity and potential outlier but accept to improve model.



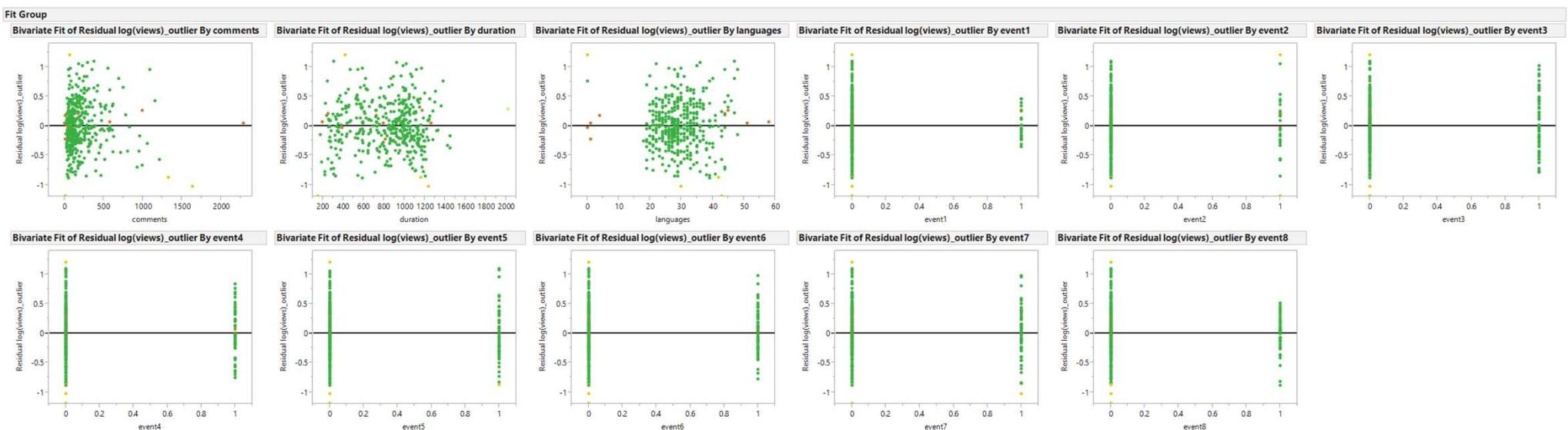
Checking for	Residual	Studentized Residual
Non constant variance	No	No
Non linearity	No	No
Potential outlier	Yes. but accept to improve model	Yes. but accept to improve model



RESIDUAL PLOT

3. Residuals against Regressors **in** the model

Result : Comments and languages that are quantitative Variable are non constant variance and non linearity. Event4 and event5 are dummy variable. So, They are constant variance and linearity.



Checking for	comments	durations	language
Non constant variance	No	No	No
Non linearity	No	No	No



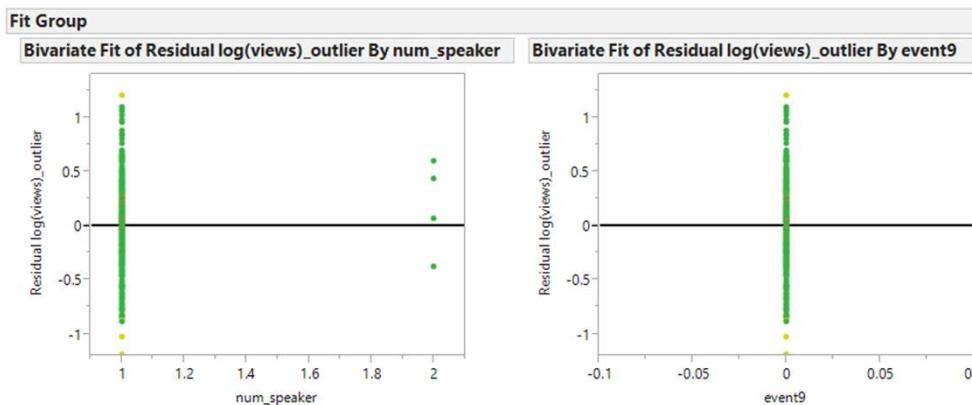
RESIDUAL PLOT ?

4. Residuals against Regressors are **not in** the model

If a pattern appears, could indicate that adding regressor might improve the model fit

Result : Comments and languages that are quantitative Variable are non constant variance and non linearity.

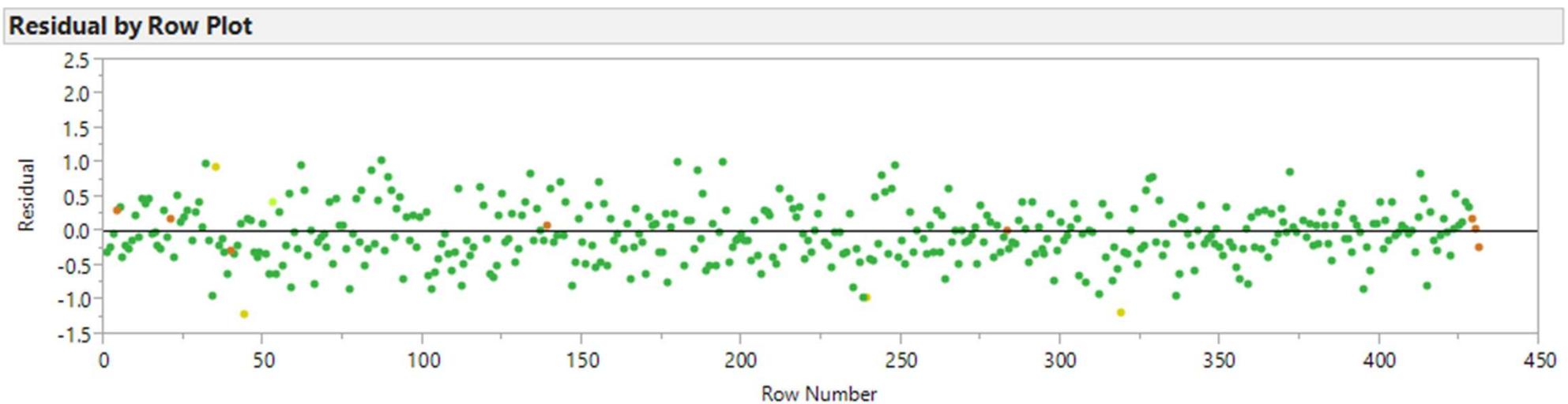
Event4 and event5 are dummy variable. So, They are constant variance and linearity.



RESIDUAL PLOT (WRONG) CHECK BY HYPOTHESIS THAT DEPENDENT OR NOT?

5. Residuals against Time order => check for Correlate errors

Result :Residuals are uncorrelated.



RESIDUAL PLOT SUMMARY

Assumption

1 Error term , ε has zero mean and constant variance

Result ε has zero mean and constant variance.

2. Errors are uncorrelated.

Result Errors are uncorrelated.

3 Errors are normally distributed.

Result Errors are normally distributed.

All are following assumption.

SO, MODEL !!!!!!!

$\text{Log}(views) = 12.7208 + 0.0010 * \text{comments} + 0.0005 * \text{durations} + 0.0542 * \text{languages} - 0.8908 * \text{event1} - 1.3408 * \text{event2} - 0.9396 * \text{event3} - 1.2772 * \text{event4} - 1.2927 * \text{event5} - 0.7942 * \text{event6} - 0.6708 * \text{event7} - 0.7128 * \text{event8}$



SUMMARY

- Model that effect to views in linear Regression model is

$$\text{Log}(views) = 12.7208 + 0.0010 * \text{comments} + 0.0005 * \text{durations} + 0.0542 * \text{languages} - 0.8908 * \text{event1} - 1.3408 * \text{event2} - 0.9396 * \text{event3} - 1.2772 * \text{event4} - 1.2927 * \text{event5} - 0.7942 * \text{event6} - 0.6708 * \text{event7} - 0.7128 * \text{event8}$$

By the number of comments , the duration of talk in second, the number of language that available , Event: TEDGlobal 2005, Event: TEDGlobal 2009, Event: TEDGlobal 2010, Event: TEDGlobal 2011, Event: TEDGlobal 2012, Event: TEDGlobal 2013, Event: TEDGlobal 2014, Event: TEDGlobal 2017 are effect to log(vlews) about 66.50% from R² in Hypothesis test. Other factors effect about 33.50%. MSE =0.1644

