DS203 – ROLL NO.22B2153

SIDDHARTH VERMA

Part A

y	x	xy	xsq	ycap	e	abs(`e)	esq	y-ybar	y_ybarsq
7.238462	0.025641	0.185602	0.000657	6.062387	1.176074	1.176074	1.383151	-5.67418	32.19629
6.310256	0.051282	0.323603	0.00263	6.202188	0.108068	0.108068	0.011679	-6.60238	43.59146
8.315385	0.076923	0.639645	0.005917	6.341989	1.973395	1.973395	3.89429	-4.59725	21.13475
4.787179	0.102564	0.490993	0.010519	6.48179	-1.69461	1.694611	2.871705	-8.12546	66.0231
5.592308	0.128205	0.716963	0.016437	6.621591	-1.02928	1.029284	1.059425	-7.32033	53.58725
7.830769	0.153846	1.204734	0.023669	6.761392	1.069377	1.069377	1.143567	-5.08187	25.8254
9.902564	0.179487	1.777383	0.032216	6.901193	3.001371	3.001371	9.008226	-3.01008	9.060552
5.607692	0.205128	1.150296	0.042078	7.040994	-1.4333	1.433302	2.054355	-7.30495	53.36225
5.146154	0.230769	1.187574	0.053254	7.180795	-2.03464	2.034642	4.139767	-7.76649	60.31829
4.784615	0.25641	1.226824	0.065746	7.320597	-2.53598	2.535981	6.4312	-8.12802	66.06477
7.05641	0.282051	1.99027	0.079553	7.460398	-0.40399	0.403987	0.163206	-5.85623	34.29542
9.394872	0.307692	2.89073	0.094675	7.600199	1.794673	1.794673	3.220852	-3.51777	12.37469
6.8	0.333333	2.266667	0.111111	7.74	-0.94	0.94	0.883599	-6.11264	37.36436
4.871795	0.358974	1.748849	0.128863	7.879801	-3.00801	3.008006	9.048099	-8.04084	64.65518
5.376923	0.384615	2.068047	0.147929	8.019602	-2.64268	2.642679	6.983751	-7.53572	56.78702
10.71538	0.410256	4.396055	0.16831	8.159403	2.555982	2.555982	6.533043	-2.19725	4.827928
11.55385	0.435897	5.036292	0.190007	8.299204	3.254642	3.254642	10.5927	-1.35879	1.846319
9.258974	0.461538	4.273373	0.213018	8.439005	0.819969	0.819969	0.67235	-3.65366	13.34927
8.097436	0.487179	3.944905	0.237344	8.578806	-0.48137	0.48137	0.231717	-4.8152	23.18618
12.10256	0.512821	6.206443	0.262985	8.718607	3.383957	3.383957	11.45117	-0.81008	0.656222
8.774359	0.538462	4.724655	0.289941	8.858408	-0.08405	0.084049	0.007064	-4.13828	17.12536
9.612821	0.564103	5.422617	0.318212	8.998209	0.614611	0.614611	0.377747	-3.29982	10.8888
6.451282	0.589744	3.804602	0.347798	9.13801	-2.68673	2.686728	7.218508	-6.46136	41.74914
11.55641	0.615385	7.111637	0.378698	9.277811	2.278599	2.278599	5.192014	-1.35623	1.839357
9.928205	0.641026	6.364234	0.410914	9.417612	0.510593	0.510593	0.260705	-2.98443	8.906847
7.9	0.666667	5.266667	0.444444	9.557413	-1.65741	1.657413	2.747019	-5.01264	25.12655
12.77179	0.692308	8.842012	0.47929	9.697214	3.07458	3.07458	9.453045	-0.14084	0.019837
6.74359	0.717949	4.841552	0.51545	9.837015	-3.09343	3.093426	9.569283	-6.16905	38.05717
8.882051	0.74359	6.604602	0.552926	9.976816	-1.09477	1.094765	1.198511	-4.03059	16.24564
9.520513	0.769231	7.323471	0.591716	10.11662	-0.5961	0.596105	0.355341	-3.39213	11.50652
7.225641	0.794872	5.743458	0.631821	10.25642	-3.03078	3.030778	9.185613	-5.687	32.34195
13.59744	0.820513	11.15687	0.673241	10.39622	3.201216	3.201216	10.24779	0.684797	0.468947
11.10256	0.846154	9.394477	0.715976	10.53602	0.566543	0.566543	0.320971	-1.81008	3.276372
10.64103	0.871795	9.276792	0.760026	10.67582	-0.0348	0.034796	0.001211	-2.27161	5.160228
9.679487	0.897436	8.686719	0.805391	10.81562	-1.13614	1.136136	1.290804	-3.23315	10.45327
8.251282	0.923077	7.616568	0.852071	10.95542	-2.70414	2.704142	7.312383	-4.66136	21.72825
7.989744	0.948718	7.580013	0.900066	11.09522	-3.10548	3.105481	9.644014	-4.9229	24.2349
10.92821	0.974359	10.64799	0.949375	11.23503	-0.30682	0.306821	0.094139	-1.98443	3.937979
11.8	1	11.8	1	11.37483	0.425173	0.425173	0.180772	-1.11264	1.237966
11.47179	1.025641	11.76594	1.05194	11.51463	-0.04283	0.042833	0.001835	-1.44084	2.076032
8.74359	1.051282	9.191979	1.105194	11.65443	-2.91084	2.910839	8.472986	-4.16905	17.38097

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.904994
R Square	0.819015
Adjusted R Square	0.817129
Standard Error	1.888324
Observations	98

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1549.075	1549.075	434.4297	2.08E-37
Residual	96	342.3137	3.565767		
Total	97	1891.389			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	5.873099	0.390306	15.04741	5.36E-27	5.098347	6.64785
0.025641026	5.481191	0.262975	20.84298	2.08E-37	4.959189	6.003194

mse_sim	19.43347				
rmse_sim	4.408341				

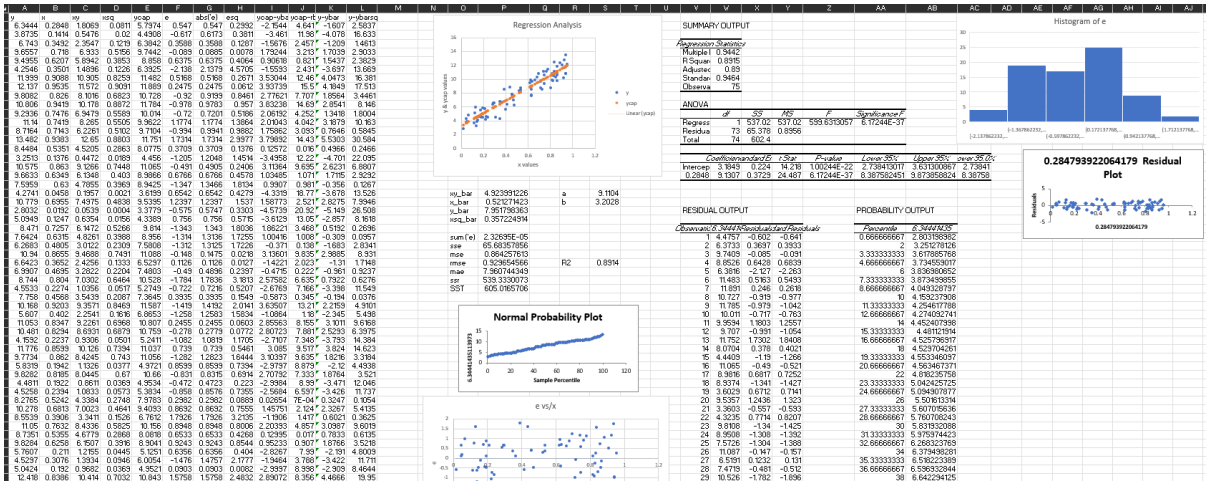
RESIDUAL OUTPUT

PROBABILITY OUTPUT

Observation	Predicted	Residuals	Standard Residuals	Percentile	y
1	6.154185	0.156071	0.08308	0.510204	4.784615
2	6.294729	2.020656	1.075638	1.530612	4.787179
3	6.435272	-1.64809	-0.87731	2.55102	4.871795
4	6.575816	-0.98351	-0.52354	3.571429	5.146154
5	6.716359	1.11441	0.593224	4.591837	5.376923
6	6.856902	3.045662	1.62127	5.612245	5.592308
7	6.997446	-1.38975	-0.7398	6.632653	5.607692
8	7.137989	-1.99184	-1.0603	7.653061	6.310256
9	7.278532	-2.49392	-1.32756	8.673469	6.451282
10	7.419076	-0.36267	-0.19305	9.693878	6.74359
11	7.559619	1.835253	0.976944	10.71429	6.8
12	7.700163	-0.90016	-0.47918	11.73469	7.05641
13	7.840706	-2.96891	-1.58041	12.7551	7.225641
14	7.981249	-2.60433	-1.38634	13.77551	7.830769
15	8.121793	2.593592	1.380624	14.79592	7.9
16	8.262336	3.29151	1.752141	15.81633	7.989744
17	8.402879	0.856095	0.455717	16.83673	8.097436
18	8.543423	-0.44599	-0.23741	17.85714	8.251282

PART B

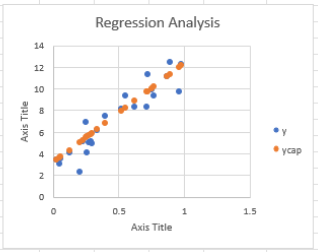
Work on Training Data

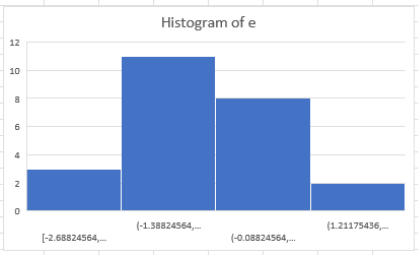


TESTING DATA WORK

y	x	ycap	e	e_sq	abs_e	y-ybar	y-ybarsq	ycap-ybar	ycap-ybarsq
3.397539	0.027206	3.450675	-0.05314	0.002823	0.053136	-3.75481	14.0986	-3.70167	13.70238951
7.467513	0.396252	6.812824	0.654689	0.428618	0.654689	0.315164	0.099328	-0.33952	0.115277182
8.315798	0.713349	9.701698	-1.3859	1.920718	1.3859	1.163449	1.353614	2.549349	6.499179587
8.309891	0.619632	8.847901	-0.53801	0.289454	0.53801	1.157542	1.339904	1.695552	2.87489533
12.43841	0.890606	11.31658	1.121835	1.258513	1.121835	5.286063	27.94246	4.164228	17.34079261
6.948992	0.248428	5.466089	1.482903	2.199	1.482903	-0.20336	0.041354	-1.68626	2.843472451
2.351169	0.201594	5.039415	-2.68825	7.226665	2.688246	-4.80118	23.05133	-2.11293	4.464489958
3.522185	0.055537	3.708782	-0.1866	0.034818	0.186597	-3.63016	13.17809	-3.44357	11.85815181
9.35813	0.547374	8.189605	1.168525	1.365451	1.168525	2.205781	4.865472	1.037256	1.075900718
11.11345	0.870359	11.13211	-0.01866	0.000348	0.018664	3.961099	15.6903	3.979763	15.83851102
5.033997	0.269361	5.656804	-0.62281	0.387889	0.622807	-2.11835	4.487417	-1.49555	2.236655529
6.137698	0.332331	6.230481	-0.09278	0.008609	0.092783	-1.01465	1.029517	-0.92187	0.849839814
3.08212	0.046149	3.623256	-0.54114	0.292828	0.541135	-4.07023	16.56676	-3.52909	12.45449815
5.170329	0.288463	5.830827	-0.6605	0.436259	0.660499	-1.98202	3.928405	-1.32152	1.746419193
4.062765	0.259699	5.568773	-1.50601	2.268059	1.506008	-3.08958	9.545528	-1.58358	2.50771346
11.30703	0.721935	9.779917	1.527113	2.332074	1.527113	4.154681	17.26137	2.627568	6.904111823
8.095555	0.522158	7.959875	0.13568	0.018409	0.13568	0.943206	0.889638	0.807526	0.652097874
9.359504	0.770366	10.22114	-0.86164	0.742418	0.861637	2.207155	4.871533	3.068792	9.417482096
9.90627	0.750346	10.03876	-0.13249	0.017553	0.132486	2.753921	7.58408	2.886407	8.331343876
5.164621	0.223309	5.237254	-0.07263	0.005276	0.072633	-1.98773	3.951063	-1.91509	3.66758883
4.130518	0.12105	4.305629	-0.17511	0.030664	0.175111	-3.02183	9.131461	-2.84672	8.103812993
12.2971	0.980275	12.13349	0.16361	0.026768	0.16361	5.14475	26.46845	4.98114	24.81175673
9.703482	0.962763	11.97395	-2.27047	5.15503	2.270469	2.551133	6.508282	4.821603	23.24785219
4.982304	0.292246	5.865289	-0.88299	0.779664	0.882986	-2.17005	4.709096	-1.28706	1.656522257

	SSR	183.2008	
	SST	218.593	
		y_bar	7.152349
a	9.110371		
b	3.202822	R2	0.838091
MAE	0.789311		
sse	27.22791		
mse	1.134496		
rmse	1.065127		





The findings of data indicate the model has worked finely with the testing data

The error in both Testing and Training Data is form of Standard Normal Distribution as visible from Histogram and therefore we can say our model has worked quite well in predicting the target label y in testing data.

Although Note the value of R2 HAS REDUCED SIGNIFICANTLY FROM 0.89 to 0.838

i.e relevancy of model has decreased over training data although it has performed well in examining errors but failed in replicating our population