REPORT

On

Multiple Linear Regression And Factor Analysis

Submitted by

EKANSH BARI

Registration No 11615973

(MGN909) Section KOE11 - A22

Programme Name: B.Tech Computer Science

School of Computer Science & Engineering
Lovely Professional University, Phagwara

Multiple Linear Regression

&

Factor Analysis

Multiple Regression

The dataset

examscore	hoursspentrevising	anxiety	Alevelentrypoints
62	40	40	24
58	31	65	20
52	35	34	22
55	26	91	22
75	51	46	28
82	48	52	28
38	25	48	18
55	37	61	20
48	30	34	18
68	44	74	26
62	32	54	24
62	40	61	24
72	61	26	26
58	35	13	24
65	45	54	20
42	30	58	20
68	39	62	24
68	47	39	26
58	41	57	22
72	46	17	28

Null Hypothesis (Ho): Exam is normally distributed.

Alternative Hypothesis (Ha): Exam is not normally distributed.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
exam score	.092	20	.200*	.982	20	.957

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

By doing a normality test we found out that the p-value is 0.957; we are rejecting the alternative hypothesis in favour of the null hypothesis.

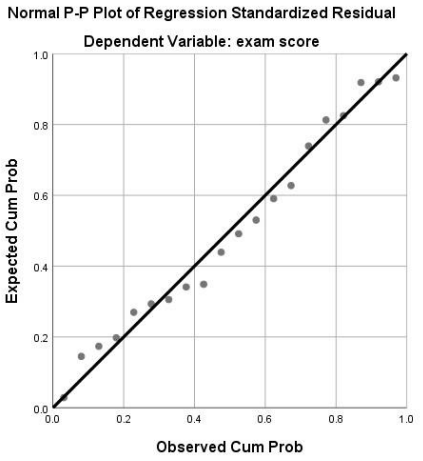
Since exam score is normally distributed we can do multiple linear regression on the dependent variable.

Correlations

		exam score	hours spent revising	anxiety	A-level entry points
Pearson Correlation	exam score	1.000	.821	-.118	.872
	hours spent revising	.821	1.000	-.340	.732
	anxiety	-.118	-.340	1.000	-.244
	A-level entry points	.872	.732	-.244	1.000
Sig. (1-tailed)	exam score	.	.000	.310	.000
	hours spent revising	.000	.	.072	.000
	anxiety	.310	.072	.	.150
	A-level entry points	.000	.000	.150	.
N	exam score	20	20	20	20
	hours spent revising	20	20	20	20
	anxiety	20	20	20	20
	A-level entry points	20	20	20	20

There is no multicollinearity between the independent variables.

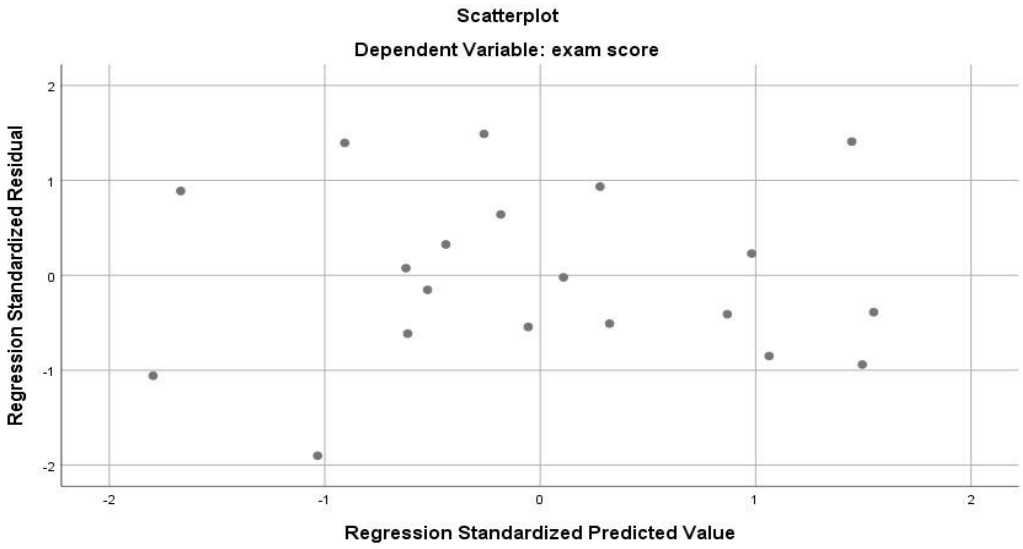
The exam score is dependent on hours spent revising and A-level entry point.



Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	42.72	76.74	61.00	10.169	20
Std. Predicted Value	-1.797	1.547	.000	1.000	20
Standard Error of Predicted Value	1.155	3.012	1.922	.559	20
Adjusted Predicted Value	42.56	79.70	60.98	10.522	20
Residual	-8.493	6.654	.000	4.100	20
Std. Residual	-1.901	1.489	.000	.918	20
Stud. Residual	-2.017	1.820	.003	1.032	20
Deleted Residual	-9.565	9.939	.021	5.244	20
Stud. Deleted Residual	-2.262	1.979	.005	1.085	20
Mahal. Distance	.320	7.684	2.850	2.186	20
Cook's Distance	.000	.409	.074	.112	20
Centered Leverage Value	.017	.404	.150	.115	20

a. Dependent Variable: exam score



Exam score follows the line in P-P plot and in scatterplot non-of, the value is greater than 3 and less than -3 on both the axis. Cook's distance is less than 1.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.927[a]	.860	.834	4.468	.860	32.811	3	16	.000

a. Predictors: (Constant), A-level entry points, anxiety, hours spent revising

b. Dependent Variable: exam score

Since R Square value is 0.860; our model explains 86% of the values.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1964.654	3	654.885	32.811	.000[b]
	Residual	319.346	16	19.959		
	Total	2284.000	19			

a. Dependent Variable: exam score

b. Predictors: (Constant), A-level entry points, anxiety, hours spent revising

The slope of the line is zero. So, we are rejecting the null hypothesis.

We are obtaining statistical significant value.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations		
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
1	(Constant)	-11.823	8.806		-1.343	.198			
	hours spent revising	.551	.171	.456	3.226	.005	.821	.628	.302
	anxiety	.104	.058	.179	1.796	.091	-.118	.410	.168
	A-level entry points	1.989	.469	.581	4.239	.001	.872	.727	.396

a. Dependent Variable: exam score

From standardized coefficients, we can see than A-level entry point and hours spent revising have the most contribution.

For A-level entry point and hours spent revising we have p-value 0.001 and 0.005 which is statistically significant. But, for anxiety p-value is 0.091; not statistically significant.

A-level entry point and hours spent revising contribute to most of the unique values in the dataset.

Examsore = 1.989*Alevelentrypoints + 0.551* hourspentrevising – 11.823 Computing the model or predicted value.

examscore	multiple
62	57.95
58	45.04
52	51.22
55	46.26
75	71.97
82	70.32
38	37.75
55	48.34
48	40.51
68	64.14
62	53.55
62	57.95
72	73.50
58	55.20
65	52.75
42	44.49
68	57.40
68	65.79
58	54.53
72	69.22

Factor Analysis

Correlation Matrix:

Correlation Matrix[a]

		PQ1	PQ2	PQ3	PQ4	PQ5	PQ6	PQ7	PQ8	PQ9	PQ10	PQ11	PQ12	PQ13	PQ14	PQ15	PQ16	PQ17	PQ18	PQ19	PQ20	PQ21	PQ22	PQ23
Correlation	PQ1	1.000	-0.059	-0.360	0.476	0.419	0.187	0.291	0.281	-0.070	0.207	0.294	0.359	0.303	0.280	0.199	0.483	0.319	0.368	-0.160	0.207	0.347	-0.069	-0.045
	PQ2	-0.059	1.000	0.263	-0.063	-0.094	-0.167	-0.211	-0.087	0.319	-0.072	-0.152	-0.201	-0.185	-0.187	-0.151	-0.119	-0.064	-0.166	0.246	-0.180	-0.195	0.235	0.044
	PQ3	-0.360	0.263	1.000	-0.379	-0.335	-0.322	-0.437	-0.289	0.329	-0.265	-0.354	-0.374	-0.363	-0.379	-0.362	-0.435	-0.305	-0.412	0.380	-0.306	-0.415	0.256	0.155
	PQ4	0.476	-0.063	-0.379	1.000	0.433	0.316	0.438	0.383	-0.186	0.250	0.432	0.430	0.359	0.358	0.401	0.437	0.429	0.437	-0.278	0.289	0.443	-0.109	-0.040
	PQ5	0.419	-0.094	-0.335	0.433	1.000	0.242	0.347	0.292	-0.085	0.275	0.300	0.383	0.289	0.264	0.250	0.442	0.324	0.328	-0.243	0.195	0.385	-0.137	-0.078
	PQ6	0.187	-0.167	-0.322	0.316	0.242	1.000	0.559	0.200	-0.127	0.439	0.369	0.336	0.450	0.369	0.397	0.205	0.276	0.547	-0.210	0.100	0.275	-0.240	-0.064
	PQ7	0.291	-0.211	-0.437	0.438	0.347	0.559	1.000	0.299	-0.141	0.331	0.423	0.473	0.476	0.435	0.418	0.376	0.362	0.586	-0.351	0.251	0.484	-0.200	-0.068
	PQ8	0.281	-0.087	-0.289	0.383	0.292	0.200	0.299	1.000	-0.039	0.236	0.696	0.223	0.322	0.246	0.273	0.312	0.568	0.287	-0.189	0.206	0.329	-0.105	-0.014
	PQ9	-0.070	0.319	0.329	-0.186	-0.085	-0.127	-0.141	-0.039	1.000	-0.161	-0.143	-0.196	-0.188	-0.112	-0.248	-0.169	-0.066	-0.122	0.249	-0.187	-0.175	0.237	0.103
	PQ10	0.207	-0.072	-0.265	0.250	0.275	0.439	0.331	0.236	-0.161	1.000	0.318	0.247	0.310	0.305	0.298	0.334	0.261	0.391	-0.158	0.048	0.195	-0.219	-0.121
	PQ11	0.294	-0.152	-0.354	0.432	0.300	0.369	0.423	0.696	-0.143	0.318	1.000	0.295	0.444	0.316	0.365	0.376	0.614	0.385	-0.280	0.297	0.387	-0.148	-0.029
	PQ12	0.359	-0.201	-0.374	0.430	0.383	0.336	0.473	0.223	-0.196	0.247	0.295	1.000	0.477	0.389	0.392	0.441	0.291	0.534	-0.309	0.309	0.497	-0.216	-0.067
	PQ13	0.303	-0.185	-0.363	0.359	0.289	0.450	0.476	0.322	-0.188	0.310	0.444	0.477	1.000	0.438	0.362	0.357	0.409	0.560	-0.314	0.201	0.358	-0.275	-0.105
	PQ14	0.280	-0.187	-0.379	0.358	0.264	0.369	0.435	0.246	-0.112	0.305	0.316	0.389	0.438	1.000	0.393	0.391	0.290	0.482	-0.315	0.192	0.357	-0.259	-0.124
	PQ15	0.199	-0.151	-0.362	0.401	0.250	0.397	0.418	0.273	-0.248	0.298	0.365	0.392	0.362	0.393	1.000	0.474	0.284	0.408	-0.277	0.215	0.323	-0.232	-0.069
	PQ16	0.483	-0.119	-0.435	0.437	0.442	0.205	0.376	0.312	-0.169	0.334	0.376	0.441	0.357	0.391	0.474	1.000	0.376	0.427	-0.263	0.200	0.399	-0.152	-0.120
	PQ17	0.319	-0.064	-0.305	0.429	0.324	0.276	0.362	0.568	-0.066	0.261	0.614	0.291	0.409	0.290	0.284	0.376	1.000	0.351	-0.164	0.186	0.360	-0.171	-0.155
	PQ18	0.368	-0.166	-0.412	0.437	0.328	0.547	0.586	0.287	-0.122	0.391	0.385	0.534	0.560	0.482	0.408	0.427	0.351	1.000	-0.325	0.226	0.431	-0.168	-0.143
	PQ19	-0.160	0.246	0.380	-0.278	-0.243	-0.210	-0.351	-0.189	0.249	-0.158	-0.280	-0.309	-0.314	-0.315	-0.277	-0.263	-0.164	-0.325	1.000	-0.264	-0.319	0.188	0.085
	PQ20	0.207	-0.180	-0.306	0.289	0.195	0.100	0.251	0.206	-0.187	0.048	0.297	0.309	0.201	0.192	0.215	0.200	0.186	0.226	-0.264	1.000	0.516	-0.069	0.018
	PQ21	0.347	-0.195	-0.415	0.443	0.385	0.275	0.484	0.329	-0.175	0.195	0.387	0.497	0.358	0.357	0.323	0.399	0.360	0.431	-0.319	0.516	1.000	-0.167	-0.065
	PQ22	-0.069	0.235	0.256	-0.109	-0.137	-0.240	-0.200	-0.105	0.237	-0.219	-0.148	-0.216	-0.275	-0.259	-0.232	-0.152	-0.171	-0.168	0.188	-0.069	-0.167	1.000	0.250
	PQ23	-0.045	0.044	0.155	-0.040	-0.078	-0.064	-0.068	-0.014	0.103	-0.121	-0.029	-0.067	-0.105	-0.124	-0.069	-0.120	-0.155	-0.143	0.085	0.018	-0.065	0.250	1.000

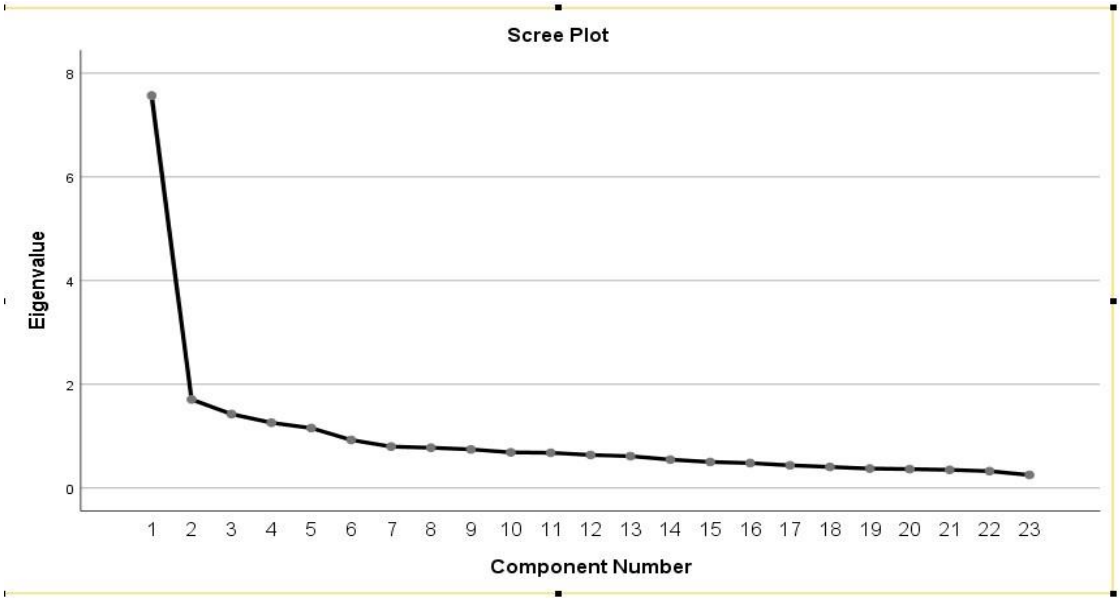
All the values of the correlation matrix are less than 0.8. So, there is no multicollinearity.

Determinant value is 0.00001 than means items are weakly related.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.916
Bartlett's Test of Sphericity	Approx. Chi-Square	4205.946
	df	253
	Sig.	.000

Statistical significance is there.



Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	7.567	32.901	32.901	7.567	32.901	32.901	5.542
2	1.707	7.422	40.323	1.707	7.422	40.323	3.262
3	1.424	6.192	46.515	1.424	6.192	46.515	1.333
4	1.258	5.471	51.985	1.258	5.471	51.985	4.382
5	1.155	5.022	57.008	1.155	5.022	57.008	4.822
6	.926	4.027	61.034				
7	.798	3.468	64.502				
8	.775	3.368	67.871				
9	.744	3.234	71.104				
10	.687	2.989	74.093				

We have 5 new components having eigenvalue greater than 1 Components

from 1 to 5 explains 57% of the variance.

Communalities

	Initial	Extraction
PQ1	1.000	.595
PQ2	1.000	.479
PQ3	1.000	.514
PQ4	1.000	.550
PQ5	1.000	.500
PQ6	1.000	.698
PQ7	1.000	.634
PQ8	1.000	.779
PQ9	1.000	.535
PQ10	1.000	.456
PQ11	1.000	.809
PQ12	1.000	.562
PQ13	1.000	.527
PQ14	1.000	.448
PQ15	1.000	.403
PQ16	1.000	.600
PQ17	1.000	.696
PQ18	1.000	.675
PQ19	1.000	.388
PQ20	1.000	.583
PQ21	1.000	.592
PQ22	1.000	.543
PQ23	1.000	.546

Extraction Method: Principal Component Analysis.

All the extraction is above 0.3

PQ11 factor has 80% eluded the variance

Component Correlation Matrix

Component	1	2	3	4	5
1	1.000	-.309	.088	-.413	-.407
2	-.309	1.000	-.027	.205	.247
3	.088	-.027	1.000	-.001	.036
4	-.413	.205	-.001	1.000	.406
5	-.407	.247	.036	.406	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Our new factors are orthogonally related because all the values are less than 0.5. All the new components are weakly related. So, we should choose the method as varimax in rotation.

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
PQ6	.822				
PQ18	.724				
PQ7	.686				
PQ13	.609				
PQ14	.557				
PQ10	.556				
PQ15	.491				
PQ1		.747			
PQ16		.679			
PQ5		.658			
PQ4		.598			
PQ21		.521		-.412	
PQ12	.465	.511			
PQ8			.849		
PQ11			.809		
PQ17			.741		
PQ2				.667	
PQ9				.662	
PQ20				-.554	
PQ19				.519	
PQ3		-.421		.450	
PQ23					.723
PQ22					.577

Extraction Method: Principal Component Analysis

Questions fall under same components are related means questions are asking similar data from the user.

Checking Reliability of new factors:

Component 1: (PQ6, PQ7, PQ10, PQ13, PQ14, PQ15, PQ18)

Reliability Statistics

Cronbach's Alpha	N of Items
.839	7

Since, it is 0.839 highly reliable. **Component 2:** (PQ1, PQ4, PQ5, PQ12, PQ16, PQ21)

Reliability Statistics

Cronbach's Alpha	N of Items
.815	6

Since, it is 0.815; highly reliable.

Component 3: (PQ8, PQ11, PQ17)

Reliability Statistics

Cronbach's Alpha	N of Items
.834	3

Since, it is 0.834; highly reliable.

Component 4: (PQ2, PQ3, PQ9, PQ19, PQ20)

Reliability Statistics

Cronbach's Alpha	N of Items
.319	5

Reliability is low, that's a problem.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
PQ2	11.50	5.825	.319	.151
PQ3	10.48	5.258	.325	.112
PQ9	10.24	4.501	.335	.061
PQ19	10.89	5.394	.288	.147
PQ20	9.64	9.058	-.339	.623

If we delete question 20 then reliability increases to 0.623.

Question 20 doesn't have much effect on the survey.

Component 5: (PQ22, PQ23)

Reliability Statistics

Cronbach's Alpha	N of Items
.400	2

Reliability is low; Question 22 and 23 doesn't have much effect on the survey.

It depends on who is conducting the survey whether to drop question which doesn't have much effect, and which combine questions having a similarity.

We cannot do regression analysis on factors since they are all independent and dependent variable is not provided.