Project Assignment 2 - Regression Analysis

Team Member

Masudul Islam

Ismail Eyamba

Project Objective

The purpose of this assignment is to find the missing data of 27 cars and build four regression models, one for each response variable i.e.

- ❑ Acceleration
- ❑ Max speed
- ❑ Fuel consumption
- ❑ CO2 emission

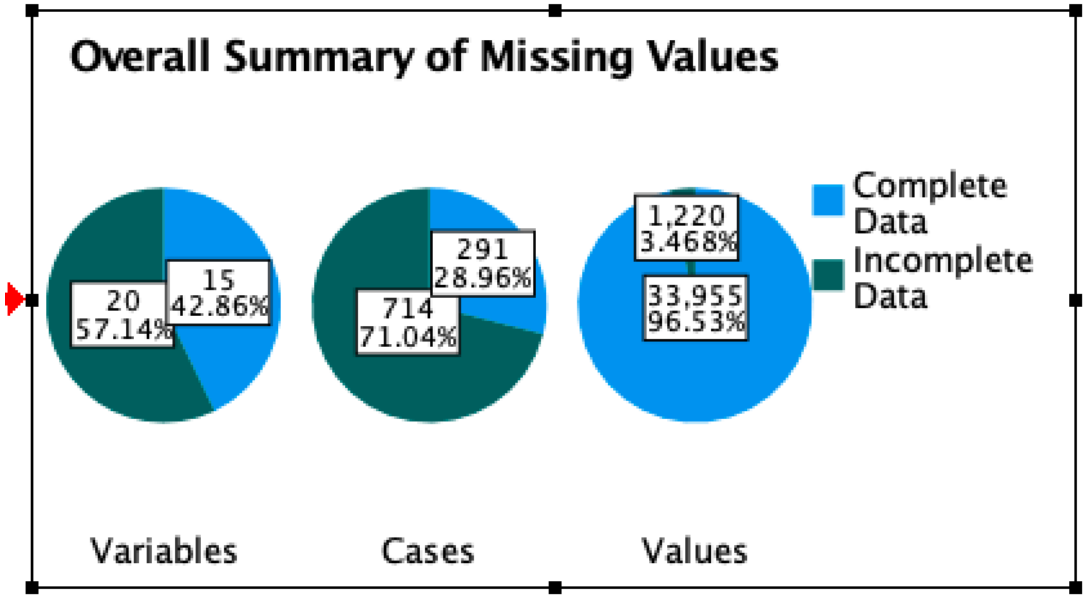
Removing Variable

Variable Summary[a,b]

	Missing N	Missing Percent	Valid N	Mean	Std. Deviation
TrunkVolume	666	66.3%	339	426.89	110.594
Number of seats	133	13.2%	872		
Acceleration 0-100 km/h	52	5.2%	953	7.365	2.4002
Number of gears	45	4.5%	960		
Automatic	42	4.2%	963		
Top speed in km/h	41	4.1%	964	221.65	34.380
Minimum trunk volume	35	3.5%	970	484.36	127.388
CO2 emissions in g/km	33	3.3%	972	148.59	59.992
MaxWeight	32	3.2%	973	2264.61	379.440
Max Torque in Nm	25	2.5%	980	411.8245	164.47723
Engine volume in cc	23	2.3%	982	2147.04	777.929
Weight	23	2.3%	982	1734.66	305.746
Maximum trunk volume	21	2.1%	984	1215.29	723.074
Number of doors	15	1.5%	990		
Max Power in kW	14	1.4%	991	178.9919	89.44314
TankVolume	9	0.9%	996	56.82	12.811
Distance between wheels	3	0.3%	1002	2805.70	180.070
Height	3	0.3%	1002	1525.81	173.293
Width	3	0.3%	1002	1856.80	111.251
Length	2	0.2%	1003	4678.81	280.080

a. Maximum number of variables shown: 25

b. Minimum percentage of missing values for variable to be included: 0.1%



Trunk volume and Number of seats variables contain large percentage of missing data which we are going to avoid to predict the asking missing data.

Acceleration

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.889[a]	.790	.789	1.0678	1.252

a. Predictors: (Constant), Max Power in kW, TurboCharged, Height, FrontWheelDrive, Number of doors

b. Dependent Variable: Acceleration 0-100 km/h

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3958.041	5	791.608	694.243	.000[b]
	Residual	1053.588	924	1.140		
	Total	5011.629	929			

a. Dependent Variable: Acceleration 0-100 km/h

b. Predictors: (Constant), Max Power in kW, TurboCharged, Height, FrontWheelDrive, Number of doors

Descriptive Statistics

	Mean	Std. Deviation	N
Acceleration 0-100 km/h	7.314	2.3226	930
Height	1514.57	129.643	930
FrontWheelDrive	.85	.362	930
Number of doors	3.71	.705	930
TurboCharged	.95	.221	930
Max Power in kW	183.8140	89.68651	930

Correlations

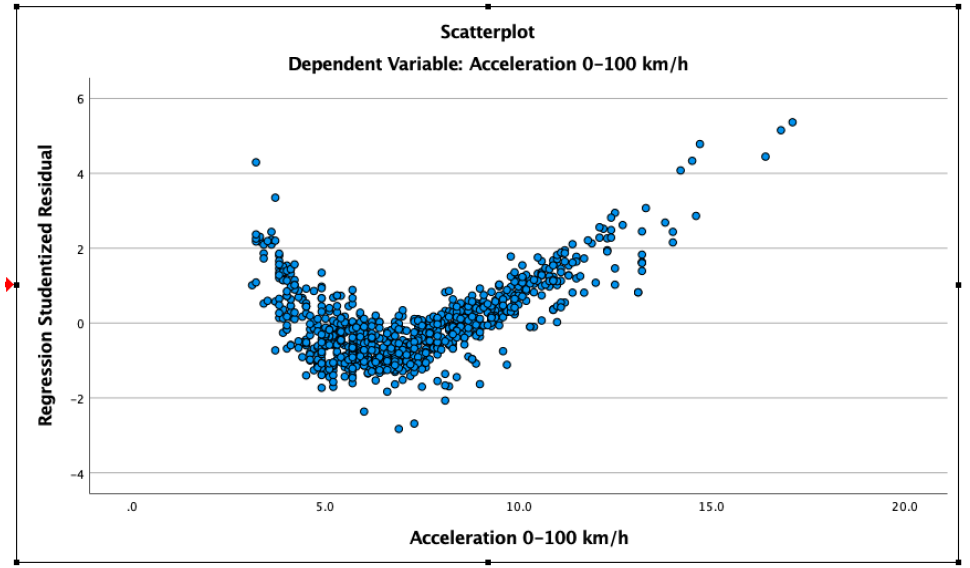
		Acceleration 0-100 km/h	Height	FrontWheelDrive	Number of doors	TurboCharged	Max Power in kW
Pearson Correlation	Acceleration 0-100 km/h	1.000	.227	.152	.170	-.164	-.849
	Height	.227	1.000	.259	.367	-.058	-.019
	FrontWheelDrive	.152	.259	1.000	.305	.021	-.092
	Number of doors	.170	.367	.305	1.000	.028	-.156
	TurboCharged	-.164	-.058	.021	.028	1.000	.006
	Max Power in kW	-.849	-.019	-.092	-.156	.006	1.000
Sig. (1-tailed)	Acceleration 0-100 km/h	.	<.001	<.001	<.001	<.001	<.001
	Height	.000	.	.000	.000	.038	.281
	FrontWheelDrive	.000	.000	.	.000	.261	.003
	Number of doors	.000	.000	.000	.	.193	.000
	TurboCharged	.000	.038	.261	.193	.	.423
	Max Power in kW	.000	.281	.003	.000	.423	.

In this model summary R sqare (0.790) indicates our set of five predictors account for 79% of the varriance in acceleration. ANOVA table shows our predticors are significantly important and it works. Durbin Watson value 1.252 is not less than 1 or grater than 3 which means the assumption of independent of observation has met. In correlation matrix table some of our predictors positively and some are negatively related to Acceleration.

Predicted Acceleration (Y)

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-1.331	12.237	7.314	2.0641	930
Std. Predicted Value	-4.188	2.385	.000	1.000	930
Standard Error of Predicted Value	.040	.225	.078	.035	930
Adjusted Predicted Value	-1.446	12.208	7.312	2.0659	930
Residual	-2.9746	5.6593	.0000	1.0649	930
Std. Residual	-2.786	5.300	.000	.997	930
Stud. Residual	-2.827	5.367	.001	1.002	930



Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	7.479	.456		16.394	<.001	6.584	8.374
	Height	.004	.000	.210	12.701	<.001	.003	.004
	FrontWheelDrive	.241	.103	.038	2.331	.020	.038	.443
	Number of doors	-.154	.056	-.047	-2.759	.006	-.263	-.044
	TurboCharged	-1.532	.159	-.146	-9.647	<.001	-1.844	-1.221
	Max Power in kW	-.022	.000	-.848	-55.407	<.001	-.023	-.021

a. Dependent Variable: Acceleration 0-100 km/h

Using coefficient table we will calculate the acccleration.

Equation : $Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots$

Y = dependent variable

β0 = constant of Y intercept

β1… = coefficient

X1… = independent variables

Y= 7.479 + (0.004*Height) + (0.241*FrontWheelDrive) + (-0.154*Door) + (-1.532*TurboCharge) + (-0.022*MaxPower)

Max Speed

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Weight, TurboCharged, Max Power in kW[b]	.	Enter

a. Dependent Variable: Top speed in km/h

b. All requested variables entered.

Correlations

		Top speed in km/h	Max Power in kW	TurboCharged	Weight
Pearson Correlation	Top speed in km/h	1.000	.593	.253	.278
	Max Power in kW	.593	1.000	.018	.684
	TurboCharged	.253	.018	1.000	-.010
	Weight	.278	.684	-.010	1.000
Sig. (1-tailed)	Top speed in km/h	.	<.001	<.001	<.001

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.663[a]	.439	.437	25.454	1.122

a. Predictors: (Constant), Weight, TurboCharged, Max Power in kW

b. Dependent Variable: Top speed in km/h

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	474057.156	3	158019.052	243.895	<.001[b]
	Residual	605785.696	935	647.899		
	Total	1079842.852	938			

a. Dependent Variable: Top speed in km/h

b. Predictors: (Constant), Weight, TurboCharged, Max Power in kW

In this model summary R sqare (0.439) indicates our set of 3 predictors contribute for 43% of the varriance in top speed. ANOVA table shows our predticors are significantly important and it works. Durbin Watson value is not less than 1 or grater than 3 which means the assumption of independent of observation has met. In correlation matrix table all our predictors are positively correlated to max speed.

Predicted Max Speed (Y)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	181.588	6.409		28.332	<.001	169.010	194.166
	Max Power in kW	.284	.013	.746	22.219	<.001	.259	.309
	TurboCharged	35.161	3.634	.237	9.675	<.001	28.029	42.293
	Weight	-.026	.004	-.230	-6.848	<.001	-.033	-.018

a. Dependent Variable: Top speed in km/h

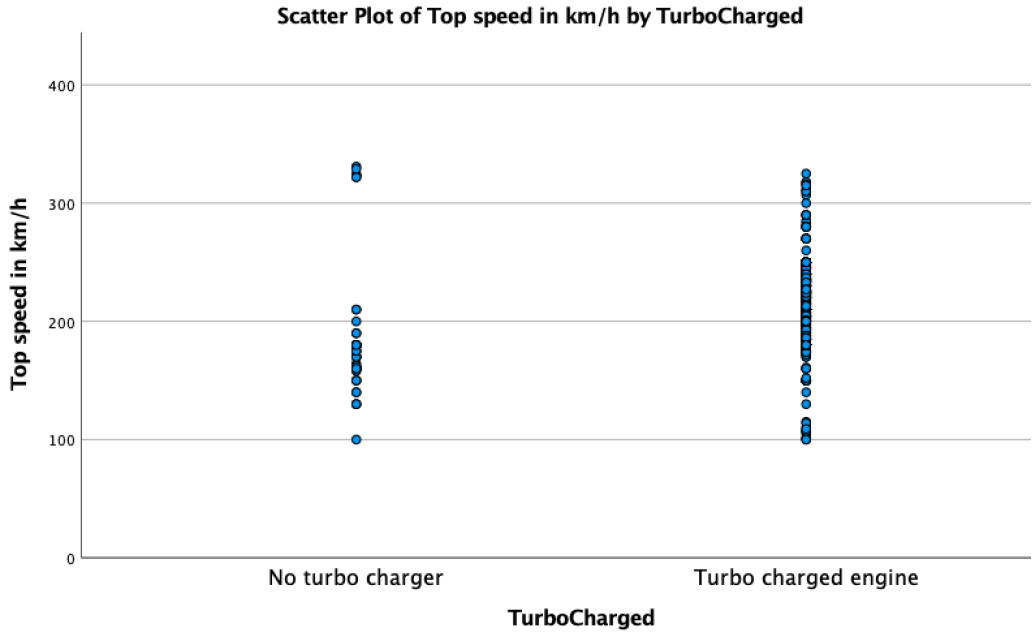
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	151.58	306.57	221.44	22.481	939
Residual	-120.027	72.889	.000	25.413	939
Std. Predicted Value	-3.107	3.787	.000	1.000	939
Std. Residual	-4.715	2.864	.000	.998	939

a. Dependent Variable: Top speed in km/h

Using coefficient table we will calculate the max speed.
Equation : $Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots$
Y = dependent variable
β0 = constant of Y intercept
β1… = coefficient
X1… = independent variables

Y=181.588+(0.284 * MaxPower)+(35.161 * TurboCharged)+(-0.026 * Weight)



Scatter plot indicates turbo charge has a big Impact on top speed

Fuel Consumption

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.697[a]	.486	.483	1.4586	1.074

a. Predictors: (Constant), Weight, Electric, PredictedMaxSpeed, Engine volume in cc, Max Power in kW

b. Dependent Variable: Fuel consumption in l/100 km / Fuel type

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1698.058	5	339.612	159.628	<.001[b]
	Residual	1797.750	845	2.128		
	Total	3495.807	850			

a. Dependent Variable: Fuel consumption in l/100 km / Fuel type

b. Predictors: (Constant), Weight, Electric, PredictedMaxSpeed, Engine volume in cc, Max Power in kW

Correlations

		Fuel consumption in l/100 km / Fuel type	Electric	Max Power in kW	PredictedMax Speed	Engine volume in cc	Weight
Pearson Correlation	Fuel consumption in l/100 km / Fuel type	1.000	-.218	.511	.532	.585	.261
	Electric	-.218	1.000	.313	.170	.145	.431
	Max Power in kW	.511	.313	1.000	.923	.844	.753
	PredictedMaxSpeed	.532	.170	.923	1.000	.723	.544
	Engine volume in cc	.585	.145	.844	.723	1.000	.734
	Weight	.261	.431	.753	.544	.734	1.000
Sig. (1-tailed)	Fuel consumption in l/100 km / Fuel type	.	<.001	<.001	<.001	<.001	<.001
	Electric	.000	.	.000	.000	.000	.000
	Max Power in kW	.000	.000	.	.000	.000	.000
	PredictedMaxSpeed	.000	.000	.000	.	.000	.000
	Engine volume in cc	.000	.000	.000	.000	.	.000
	Weight	.000	.000	.000	.000	.000	.

In this model summary R sqare (0.486) indicates our set of five predictors account for 48% of the varriance in acceleration. ANOVA table shows our predticors are significantly important and it works. Durbin Watson value 1.074 is not less than 1 or grater than 3 which means the assumption of independent of observation has met. In correlation matrix table all our predictors positively correlated to fuel consumption except electric.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Weight, Electric, PredictedMax Speed, Engine volume in cc, Max Power in kW[b]	.	Enter

a. Dependent Variable: Fuel consumption in l/100 km / Fuel type

b. All requested variables entered.

Predicted Fuel Consumption (Y)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	1.123	1.883		.597	.551	-2.573	4.820
	Electric	-1.237	.132	-.282	-9.365	<.001	-1.497	-.978
	Max Power in kW	.005	.003	.200	1.692	.091	-.001	.011
	PredictedMaxSpeed	.015	.009	.143	1.709	.088	-.002	.033
	Engine volume in cc	.002	.000	.523	9.984	<.001	.001	.002
	Weight	-.002	.000	-.230	-4.716	<.001	-.002	-.001

a. Dependent Variable: Fuel consumption in l/100 km / Fuel type

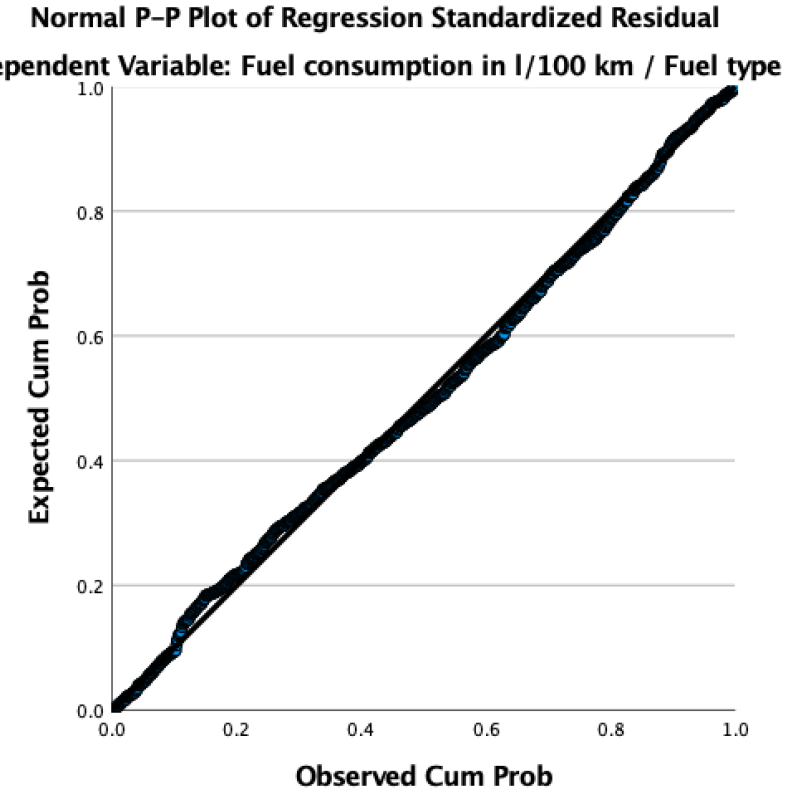
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2.825	11.752	5.502	1.4134	851
Residual	-4.6389	4.7482	.0000	1.4543	851
Std. Predicted Value	-1.894	4.422	.000	1.000	851
Std. Residual	-3.180	3.255	.000	.997	851

a. Dependent Variable: Fuel consumption in l/100 km / Fuel type

Using coefficient table we will calculate the fuel consumption.
Equation : $Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots$
Y = dependent variable
β0 = constant of Y intercept
β1… = coefficient
X1… = independent variables

Y=1.123+(1.237*Electric)+(0.005*MaxPower)+(0.015*PredictedMaxSpeed)+(0.002*Engine volume in cc)+(-0.002*Weight)



CO2 Emission

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Engine volume in cc, Electric, PredictedMax Speed, Max Power in kW[b]	.	Enter

a. Dependent Variable: CO2 emissions in g/km

b. All requested variables entered.

Correlations

		CO2 emissions in g/km	Electric	PredictedMax Speed	Max Power in kW	Engine volume in cc
Pearson Correlation	CO2 emissions in g/km	1.000	-.299	.537	.549	.670
	Electric	-.299	1.000	.084	.210	.074
	PredictedMaxSpeed	.537	.084	1.000	.937	.759
	Max Power in kW	.549	.210	.937	1.000	.868
	Engine volume in cc	.670	.074	.759	.868	1.000
Sig. (1-tailed)	CO2 emissions in g/km	.	<.001	<.001	<.001	<.001

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.759[a]	.576	.574	36.537	1.061

a. Predictors: (Constant), Engine volume in cc, Electric, PredictedMaxSpeed, Max Power in kW

b. Dependent Variable: CO2 emissions in g/km

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1660192.080	4	415048.020	310.907	<.001[b]
	Residual	1222822.672	916	1334.959		
	Total	2883014.751	920			

a. Dependent Variable: CO2 emissions in g/km

b. Predictors: (Constant), Engine volume in cc, Electric, PredictedMaxSpeed, Max Power in kW

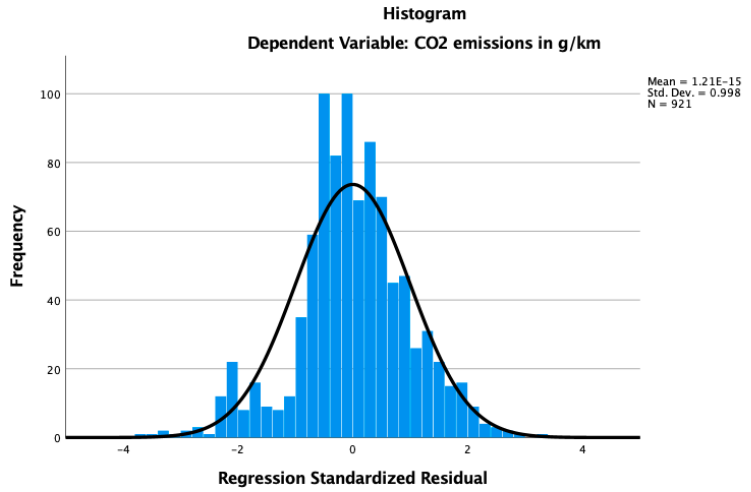
ANOVA table shows our model is working and has sogificatnt important on predicors. Durbin Watson value is not less than 1 or grater than 3 which means the assumption of independent of observation has met. R square represents 57% of CO2 emissions accounted by the combination of our predictors.

Predicted CO2 Emission (Y)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	-33.243	35.614		-.933	.351	-103.137	36.652
	Electric	-41.248	3.094	-.334	-13.332	<.001	-47.319	-35.176
	PredictedMaxSpeed	.479	.188	.184	2.542	.011	.109	.849
	Max Power in kW	-.091	.063	-.145	-1.454	.146	-.214	.032
	Engine volume in cc	.049	.004	.681	13.825	<.001	.042	.056

a. Dependent Variable: CO2 emissions in g/km



Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	65.35	310.43	151.62	42.480	921
Residual	-132.149	118.338	.000	36.458	921
Std. Predicted Value	-2.031	3.738	.000	1.000	921
Std. Residual	-3.617	3.239	.000	.998	921

a. Dependent Variable: CO2 emissions in g/km

Using coefficient table we will calculate the CO2 emission.
Equation : $Y = \beta0 + \beta1X1 + \beta2X2 + \ldots$
Y = dependent variable
β0 = constant of Y intercept
β1… = coefficient
X1… = independent variables

Y=-33.243+(-41.248 * Electric)+(0.479 * PredictedMaxSpeed)+(0.091*MaxPower)+(0.049*Displacement)

Thank You!