MATH 5345 / Regression Analysis: Final Report

Dr. Sun

Gabriela Lara and Joshua Mitchell

Contents

List of Tables

List of Figures

Abstract

Introduction

Models and Analysis Results

Conclusion and Discussion

References

	Estimate	Std. Error	t value	Pr(>|t|)	Significance
(Intercept)	-18.3106	4.6933	-3.90	0.0001	***
wgt_c	-0.0067	0.0007	-10.23	0.0000	***
modelyr_mvd	0.7805	0.0519	15.03	0.0000	***
origin_mvd2	2.6340	0.5665	4.65	0.0000	***
origin_mvd3	2.8557	0.5528	5.17	0.0000	***
hp_c	-0.0174	0.0137	-1.27	0.2056	
displ_c	0.0241	0.0077	3.14	0.0018	**
cylnum_mvd	-0.5123	0.3222	-1.59	0.1126	
acc_c	0.0845	0.0984	0.86	0.3913	

Table 1: R Summary of original full model (relating mpg to the rest)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significance
wgt_c	1	16470.05	16470.05	1505.92	0.0000	***
modelyr_mvd	1	2756.94	2756.94	252.08	0.0000	***
origin_mvd	2	261.23	130.61	11.94	0.0000	***
hp_c	1	8.96	8.96	0.82	0.3659	
displ_c	1	77.03	77.03	7.04	0.0083	**
cylnum_mvd	1	29.10	29.10	2.66	0.1037	
acc_c	1	8.06	8.06	0.74	0.3913	
Residuals	382	4177.89	10.94			

Table 2: R ANOVA of original full model (relating mpg to the rest)

	GVIF	Df	GVIF^(1/(2*Df))
wgt_c	11.07	1.00	3.33
modelyr_mvd	1.30	1.00	1.14
origin_mvd	2.09	2.00	1.20
hp_c	9.98	1.00	3.16
displ_c	22.87	1.00	4.78
cylnum_mvd	10.74	1.00	3.28
acc_c	2.62	1.00	1.62

Table 3: VIF of each regressor in the Full Original Model

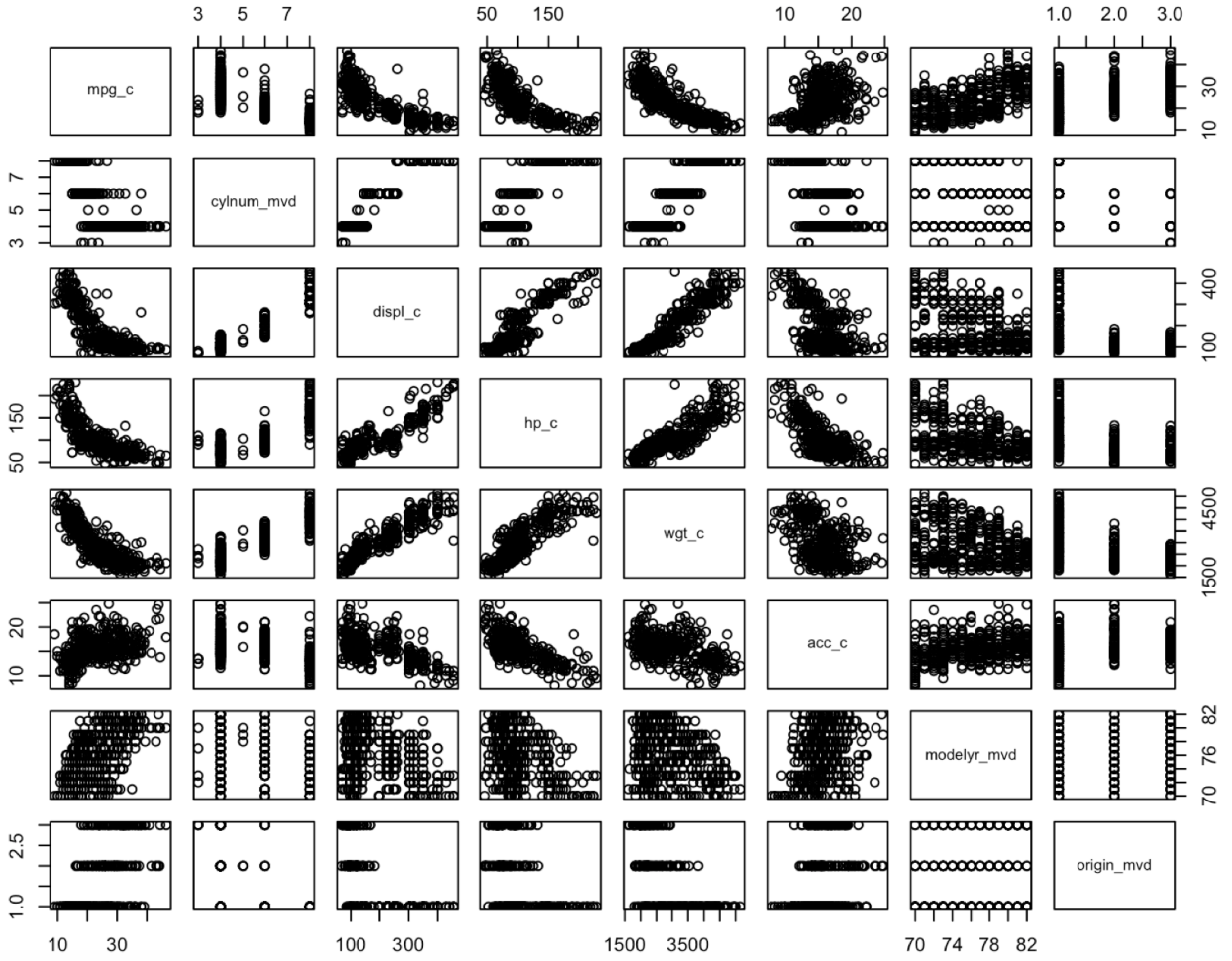

Figure 1: A scatterplot matrix between all the variables in the automobile data set

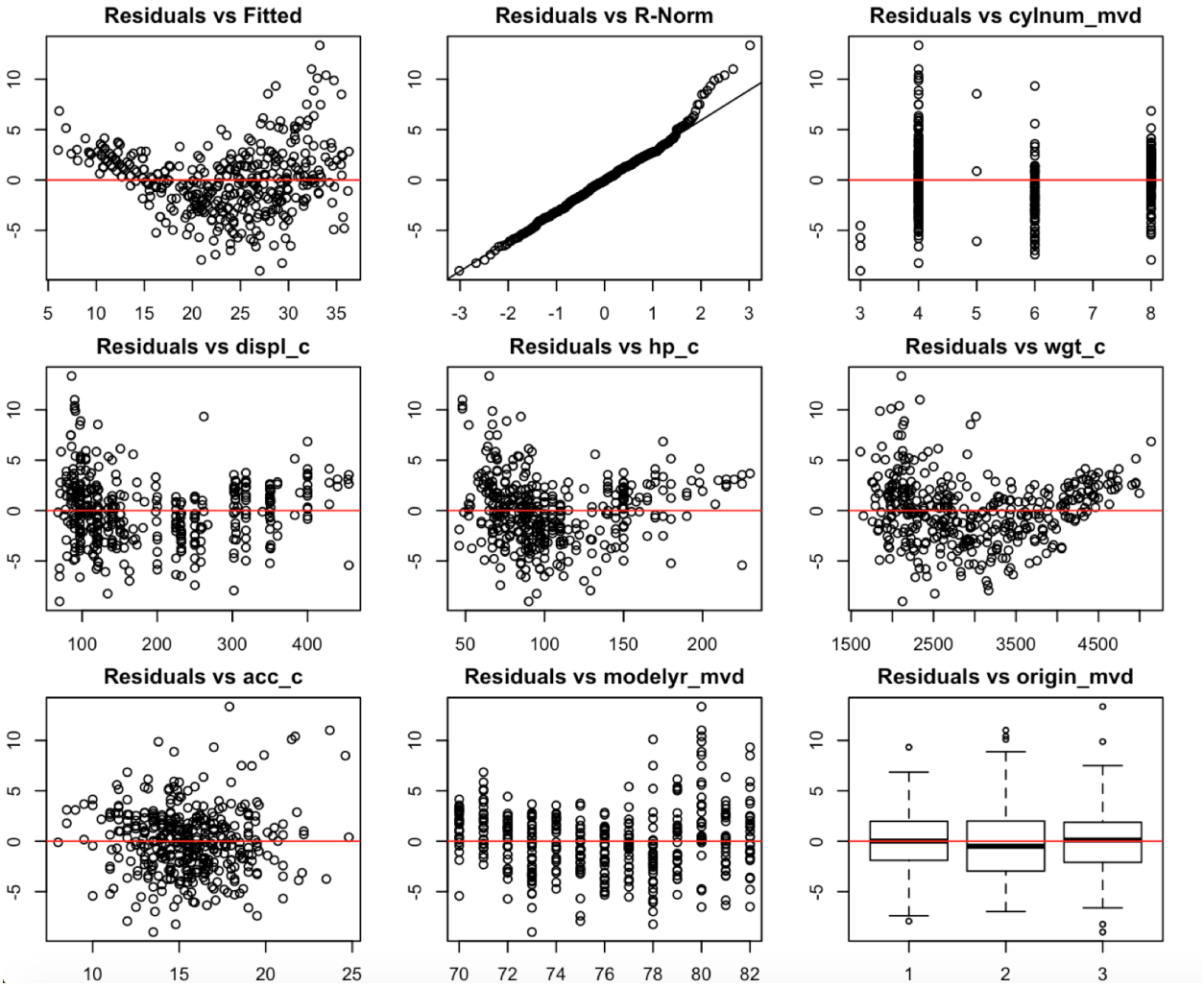

Figure 2: A Residual vs Fitted, a Residual vs R-Norm, and Residual vs Regressors plots of the Original Full Model

	Regressor	F_Statistic	P_Value	Significance
1	Displacement	9.89	0.00	**
2	Weight	104.63	0.00	***
3	HP	1.61	0.21	none
4	Cylinder Num	2.53	0.11	none

Table 4: A partial F test on each of the regressors with high VIF scores

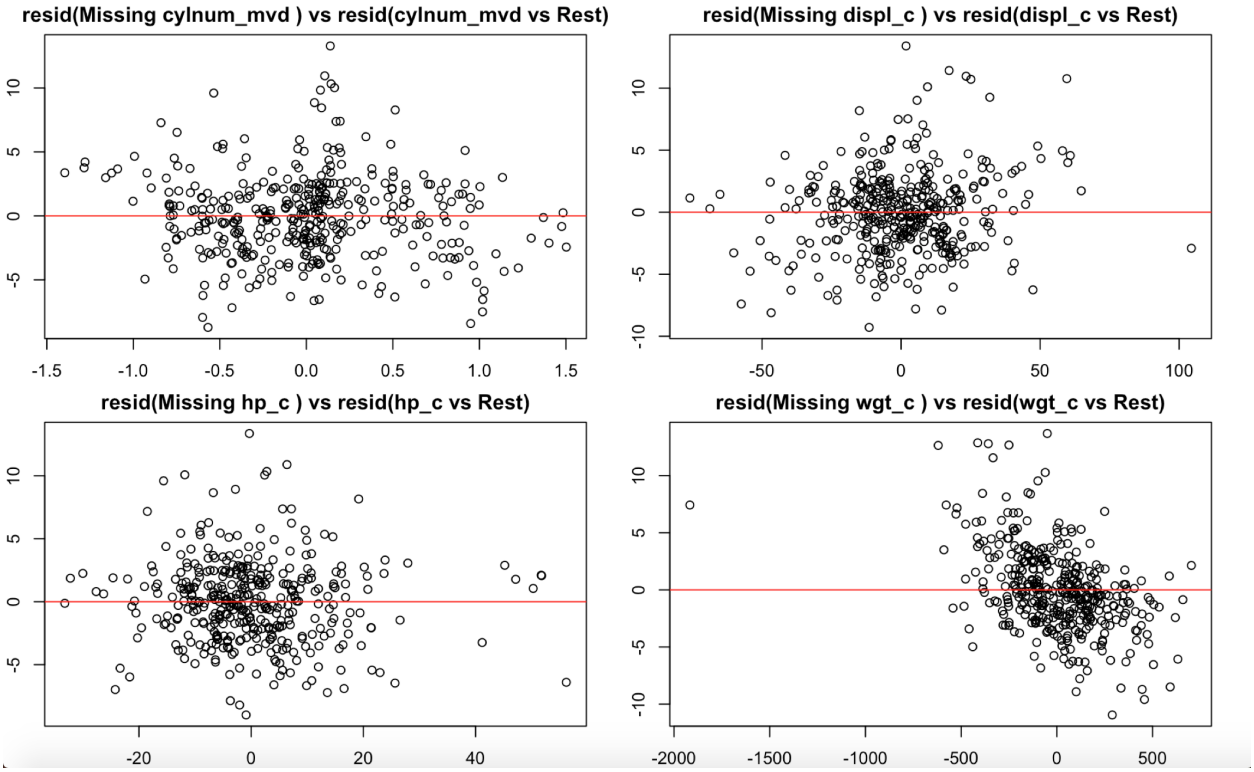

Figure 3: Partial regression plots on each of the regressors with high VIF scores

POST TRANSFORMATION

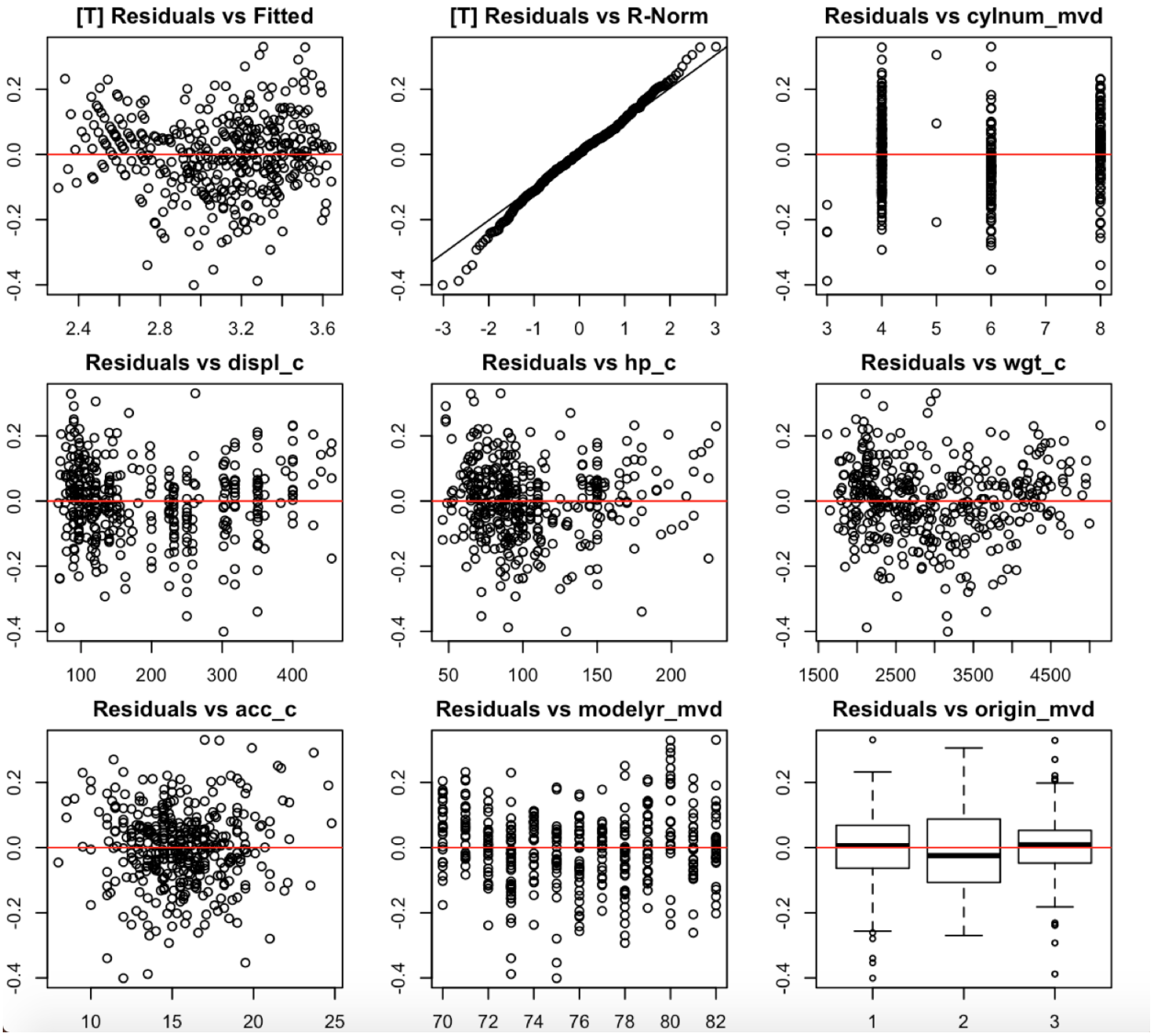

Figure 4: A Residual vs Fitted, a Residual vs R-Norm, and Residual vs Regressors plots of the Transformed Full Model

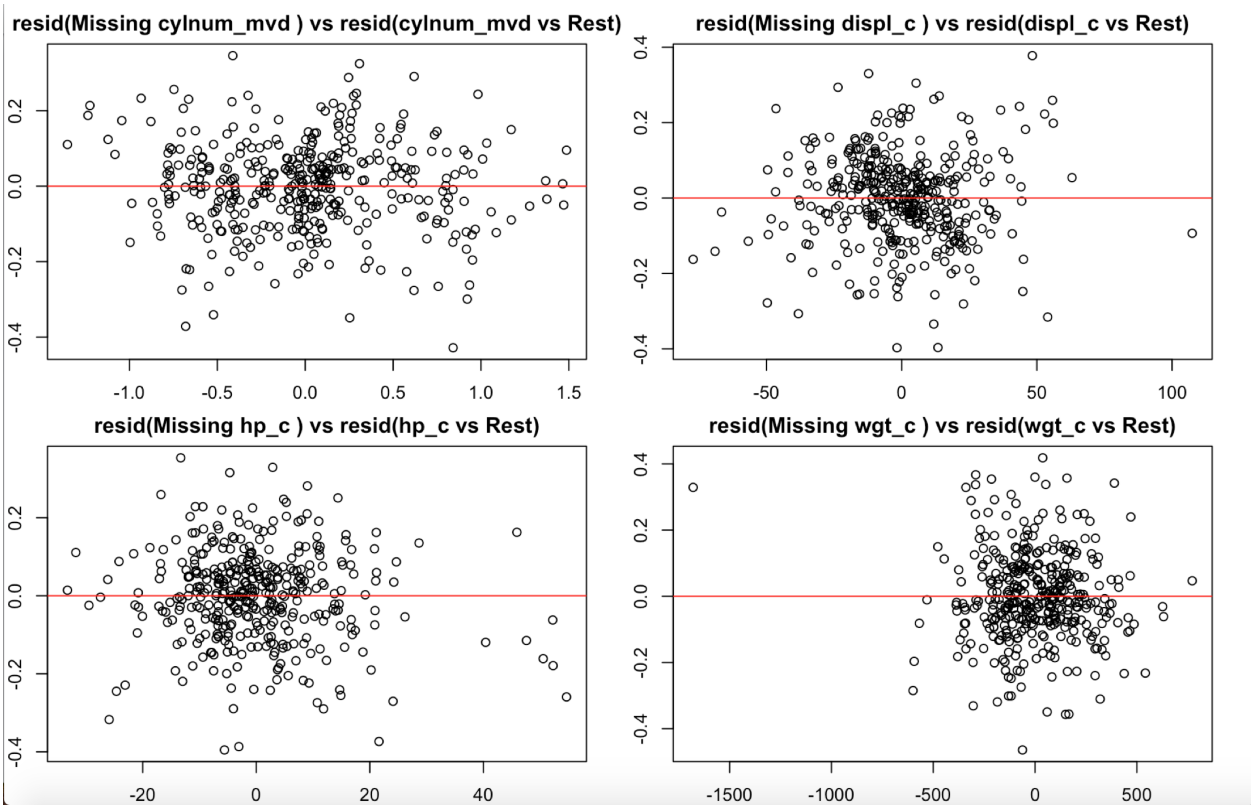

Figure 5: Partial regression plots on each of the regressors with high VIF scores on the Transformed Full Model

	Regressor	F_Statistic	P_Value	Significance
1	Displacement	8.39	0.00	**
2	Weight	126.68	0.00	***
3	HP	9.10	0.00	**
4	Cylinder Num	6.35	0.01	*

Table 5: A partial F test on each of the regressors with high VIF scores on the Transformed Full Model

	Model	R_Sq	AR_Sq	MS_res
1	Interaction	0.88	0.87	7.90
2	Transformed + Interaction	0.90	0.90	0.01

Table 6: A chart comparing the Untransformed Full Model with all combinations of interaction terms for the high VIF regressors against the same model with a log transformation on the response variable (mpg)

	Selection_Method	Num_Regressors	R_Sq	Adj_R_Sq	MS_res
1	Forward	6.00	0.89	0.89	0.01
2	Backward	16.00	0.90	0.90	0.01
3	Stepwise	6.00	0.89	0.89	0.01

Table 7: Statistics about the models outputted from Forward, Backward, and Stepwise Selection algorithms in R (note that the model selected by Forward and Stepwise selection is identical, so just the Forward model will be considered in further sections)

	GVIF	Df	GVIF^(1/(2*Df))
wgt_c	13.83	1.00	3.72
modelyr_mvd	1.27	1.00	1.13
origin_mvd	1.74	2.00	1.15
hp_c	37.47	1.00	6.12
acc_c	2.61	1.00	1.62
wgt_c:hp_c	58.06	1.00	7.62

Table 8: VIF of each regressor in the Forward Model

	GVIF	Df	GVIF^(1/(2*Df))
wgt_c	3110.02	1.00	55.77
modelyr_mvd	1.44	1.00	1.20
origin_mvd	3.01	2.00	1.32
hp_c	2806.06	1.00	52.97
displ_c	10568.29	1.00	102.80
cylnum_mvd	975.92	1.00	31.24
acc_c	3.64	1.00	1.91
wgt_c:hp_c	27058.36	1.00	164.49
hp_c:displ_c	39680.75	1.00	199.20
wgt_c:displ_c	11724.72	1.00	108.28
wgt_c:cylnum_mvd	9069.37	1.00	95.23
hp_c:cylnum_mvd	9125.25	1.00	95.53
displ_c:cylnum_mvd	19861.83	1.00	140.93
wgt_c:hp_c:cylnum_mvd	41867.56	1.00	204.62
wgt_c:displ_c:cylnum_mvd	15077.88	1.00	122.79
hp_c:displ_c:cylnum_mvd	44842.29	1.00	211.76

Table 9: VIF of each regressor in the Backward Model

	Model	Num_Infl_Pnts	Percent_Infl_Pnts	Common_Infl_Pnts
1	Forward	20.00	5.12%	14.00
2	Backward	36.00	9.21%	14.00

Table 10: Influential point comparison of Forward Model vs Backward Model

	Model	R_Sq	AR_Sq	MS_res
1	Forward w/o Infl	0.91	0.91	0.01
2	Backward w/o Infl	0.90	0.90	0.01

Table 11: Forward Model with no influential points vs Backward Model with no influential points

FINAL MODEL: mpg (c) ~ modelyr (mvd) + origin (mvd) + hp (c) + acc (c) + wgt (c) * hp (c)

	Estimate	Std. Error	t value	Pr(>|t|)	Significance
(Intercept)	2.1373	0.1735	12.32	0.00001	***
wgt_c	-0.0004	0.0000	-14.76	0.00001	***
modelyr_mvd	0.0309	0.0018	17.59	0.00001	***
origin_mvd2	0.0558	0.0177	3.14	0.00180	**
origin_mvd3	0.0455	0.0180	2.52	0.01210	*
hp_c	-0.0064	0.0009	-7.06	0.00001	***
acc_c	-0.0053	0.0034	-1.59	0.11180	
wgt_c:hp_c	0.0000013	0.0000002	6.71	0.00001	***

Table 12: R Summary of the final model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significance
wgt_c	1	34.62	34.62	2714.64	0.0000	***
modelyr_mvd	1	4.72	4.72	369.94	0.0000	***
origin_mvd	2	0.25	0.12	9.78	0.0001	***
hp_c	1	0.11	0.11	8.87	0.0031	**
acc_c	1	0.001	0.00	0.26	0.6078	
wgt_c:hp_c	1	0.57	0.57	45.04	0.0000	***
Residuals	383	4.88	0.01			

Table 13: R ANOVA of the final model