Name: Dennis E. Miller

Course: Statistics 571 section 35870

Assignment: Course Project – Inferring Residual Value of 2003 Honda Hybrid Civic by Establishing Residual value of Toyota Prius Hybrid.

Background

In 2001, Toyota Motor Company began selling the Prius model automobile featuring what was called 'Toyota Hybrid System'. Instead of relying on the traditional gasoline fueled internal combustion engine, Prius derives its motive force through a carefully balanced harmonization between a much smaller than traditional 1.5 Liter gas engine and an electric motor/generator. The motor is driven by batteries that are alternately charged or drawn upon to suit driving conditions. This small gas engine/electric motor pair boasts both exceptional fuel mileage and emission performance.

Building on success developing similar technology in their domestic market, just this year, American Honda introduced a version of their popular Civic model incorporating Hybrid Gasoline-Electric technology to the U.S market. The Honda Civic Hybrid has caught my attention since I first saw one on the road. I have long been fascinated with the potential of this technology, but find the styling and features of the Honda Civic much more appealing than the Toyota Prius packaging.

While I have both financed purchases and leases in the past, I find leasing a more appealing option for my first hybrid technology vehicle. There are many web sites that will extol the virtues and vices of leasing as compared to purchasing, but the relative novelty of this technology in the local market gives me concern that the depreciation curve for this vehicle could be subject to greater fluctuation than for its traditionally powered counterpart. If demand for the technology falls, I do not want to be stuck with a car whose value is significantly less than what I have left to pay on it. In a closed-end lease, the future value of the car is established up front. All risk of sudden or severe depreciation is retained by the financer, not the leasor.

It is well advised that the savvy consumer do his homework prior to showing up at the new car showroom. Not doing so is an almost guaranteed formula for falling prey to the wiley schemes of the experienced dealership finance officer. The formulas for leasing are well established[1]:

Monthly Payment = Depreciation Fee + Finance Fee + Tax

Net. Cap Cost = Capitalized Cost - Cap. Cost Reduction

Depreciation Fee = (Net. Cap Cost - Residual) / Lease Term

Finance Fee = (Net. Cap Cost + Residual) x Money Factor

---

[1] http://www.pe.net/firm/dpw-designs/products/leaseit/lease_it_man_05.html

Any number of web sites (or a good business calculator) will complete the expected payment calculation given that the user enters the appropriate inputs. Most of the inputs can be reasonably estimated through sources widely available on the internet. However, the residual value, (or expected value of the vehicle at a future point in time as a percentage of the original MSRP), is not easily established for this vehicle since it has just been introduced to in this country.

Data Collection Strategy

As the subject of my course project, I gathered data on used Toyota Prius hybrid models as an approximation for the expected depreciation of the Honda Civic Hybrid. My initial sampling frame was to collect information on used Toyota Prius sales through the websites of several large Metropolitan DC dealerships that list used car inventories online. Dealers I periodically reviewed included the following:
Fitzgerald Auto Mall http://www.fitzmall.com
Jim Coleman Auto Group http://www.jimcolemanautomotive.com
Ourisman http://www.ourisman.com
Criswell http://www.criswellauto.com
King http://www.kingauto.com
Antwerpen http://www.antwerpenauto.com
Darcars http://www.darcars.com
Pohanka http://www.pohanka.com
Alexandria Toyota http://www.alexandriatoyota.com
Car Max http://www.carmax.com

It became obvious early in data collection that a significant sample of used Prius models would not be found through these local sources alone. I broadened my search to include national markets listed at Toyota's certified used care link on the Toyota website and the used cars inventory list on cars.com and autotrader.com. Finally, I solicited colleagues working in an automotive lending division of a major area bank to provide available data through consolidated list of sales at a major national used car auction. Auction data was researched through Manheim Auctions. Manheim Auctions headquartered in Atlanta, GA, claims to be the largest and highest volume wholesale automobile auction company in the world. The company operates more than 60 auction facilities in the US, moving over 70,000 vehicles a week through auction[2]. While collecting these volumes of data, generally 30 days of collected data are available through their web interface.

In addition to collecting data on individual vehicles, I researched basic information on the Toyota Prius production. I found through multiple corroborating sources that the manufacturer's suggested retail price (MSRP) on the 2001 and 2002 Toyota Prius was $19,995. The MSRP jumped to $20,480 for the 2003 model.[3] Toyota sold 15,556 units in the 2001 model year, and targeted to sell between 17,000 and 18,000 in 2002.[4] By March 2003, Knight Ridder reported that Toyota sold 44,800 Prius vehicles in the U.S. Market since its introduction.[5]

[2] www.manheim.com for Tuesday, July 22, 2003.
[3] www.toyota.com
[4] Kohn, Joe 'Toyota moves Prius into dealerships.', Automotive News, Jan 21, 2002 v76 i5967 p1
[5] Knight Ridder/Tribune Business News, 'Hybrid Cars Sell Out in Orange County, Calif., as Gasoline Hits Record Levels.' March 15, 2003. InfoTrac OneFile pITEM03074104

Collected Data

Variables targeted for collection for analytical interest included: vehicle year, listing locality, asking price, trim line, mileage, engine, transmission, color and availability of specific features (e.g. air condition, electric windows, upgraded stereo, alloy wheels, etc). In addition to variables of analytical interest, I captured certain reference and identifying information. These included: data capture date, Vehicle Identification Number (VIN), dealer name, and the web site address where the data was collected. I logged rows for data capture date where specific vehicle information was not found for administrative purposes, to keep track of sites visited. Rows logged exclusively for tracking are excluded from all analysis.

Collected Data Frequencies by Site

Web Site	Number of Visits	Number of Rows	Non-excluded Rows
http://www.manheim.com/ (auction)	1	7	7
http://www.autotrader.com/	1	34	34
http://www.alexandriatoyota.com/	2	2	0
http://www.antwerpenauto.com/	2	2	0
http://www.carmax.com/	4	4	2
http://www.cars.com/	7	67	67
http://www.criswellauto.com/	2	2	0
http://www.darcars.com/	2	2	1
http://www.fitzmall.com/	2	2	0
http://www.jimcolemanautomotive.com/	2	2	0
http://www.kingauto.com/	2	2	0
http://www.koons.com/	2	2	0
http://www.mileone.com/mileone/herbgordon.htm	2	2	0
http://www.ourisman.com/	2	2	1
http://www.pohanka.com/	2	2	0
http://www.rosenthalauto.com/	2	2	0
http://www.sheehy.com/	2	2	0
http://www.toyotacertified.com/	5	88	88
Total		226	200

The variety of sources presented many data collection challenges. Many basic data formatting and integrity issues arose. For example, the popular automatic transmission was listed on some sites as 'AUTO/OD' while others listed it by the more technical 'CVT-E' [constant velocity transmission — electronically controlled]. After researching the original vehicle information with Toyota, I found that only one automatic transmission was installed on these vehicles. I standardized the data accordingly. The same vehicle was often listed on multiple web sites and, not unexpectedly, the same listing appeared during several sit visits. In at least one case, the same vehicle was listed at multiple dealerships within a network at the same time. I used the VIN number to de-duplicate multiple listings for the same vehicle. As I collected data over time, I would find previously recorded vehicles listed at a lower price or with slightly more mileage. Since my goal is to identify the most accurate selling price for the vehicle, I eliminated the earlier citation in favor of the most current information. Finally, there was great inconsistency in how and whether specific features were installed on individual vehicles. Ultimately, I decided that optional equipment could not be reliably collected from these sources and would have to be considered 'confounding variables'.

The following original data elements were collected:

Variable	Data Type	Width	Format	Description
Capture Date	Date	10	Continuous	Data actual or attempted collection date
Dealer	Character	31	Nominal	Listing dealer name
Dealer City	Character	15	Nominal	Listing city
Dealer State	Character	3	Nominal	Listing state
Engine	Character	15	Nominal	Listed Engine size, (all the same)
External Color	Character	15	Nominal	Vehicle color as listed
Make	Character	6	Nominal	Toyota (all the same)
Miles	Numeric	16.0	Continuous	Listed miles
Model	Character	5	Nominal	Prius (all the same)
Price	Numeric	16	Continuous	Listed price
Transmission	Character	6	Nominal	Listed Transmission
VIN	Character	18	Nominal	Listed Vehicle Information Number if available
Website	Character	45	Nominal	Site where vehicle was listed
Year	Character	4	Nominal	Listed vehicle model year

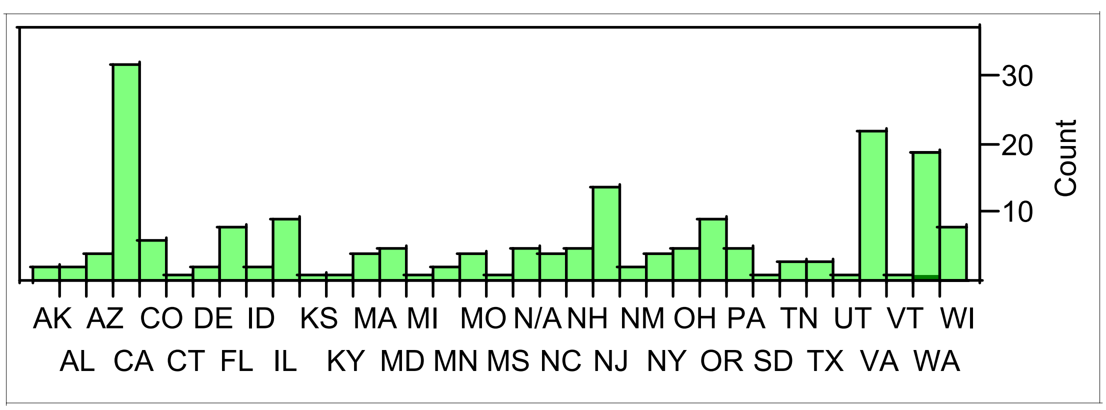
In total, data collection effort spanned the time period from June 14, 2003 to July 23, 2003.

Data Preparation

Much additional work was needed to transform the data into a form where it was useful for analysis and modeling.

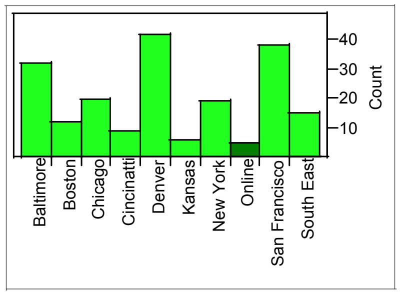
Dealer/Listing State

While it was my original intent to collect data from my local market, it soon became obvious that these sources would not support an adequate sample and I expanded my sampling frame. Ultimately, data was collected from 34 states with 5 'for sale by owner' listings where the state could not be determined:

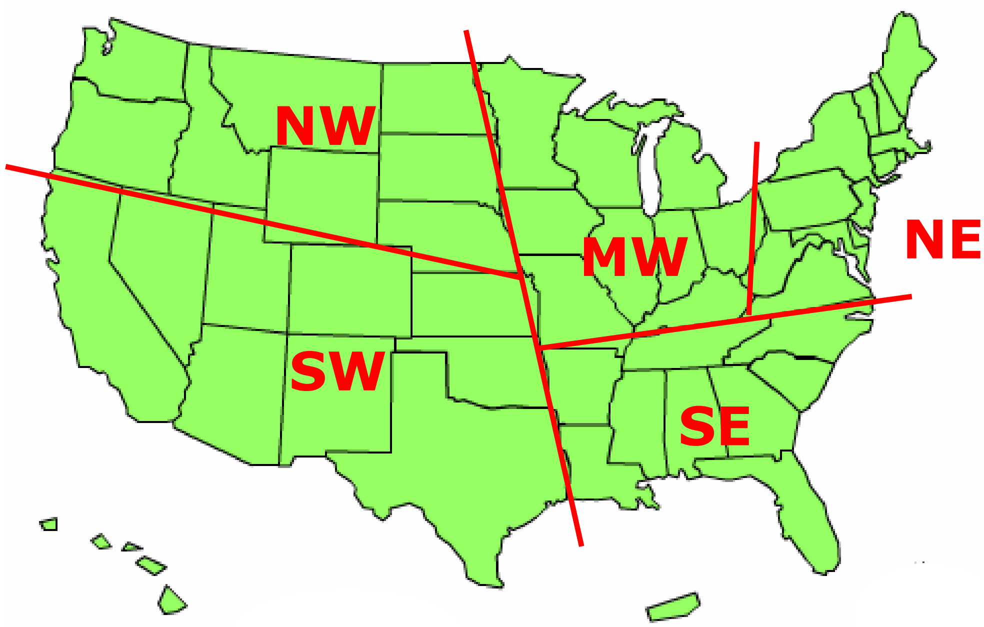


In order to account for price variability across markets, the listing state data needed to be encoded in such a way that a reasonable number of levels are defined. Two data coding schemes were considered. Additional research determined that Toyota, under the management of Toyota Motor Sales, USA Inc., divides the country into nine sales regions, each with a regional sales headquarters. While the demarcation lines between sales regions were not forthcoming, I used my best judgment to assign each observation to a specific sales region, with the 'for sale by owner' classified as 'online'.

Sales Region	Count	Distribution
Baltimore	32	16.16%
Boston	12	6.06%
Chicago	20	10.10%
Cincinnati	9	4.54%
Denver	42	21.21%
Kansas	6	3.03%
New York	19	9.59%
Online	5	2.52%
San Francisco	38	19.19%
South East	15	7.57%



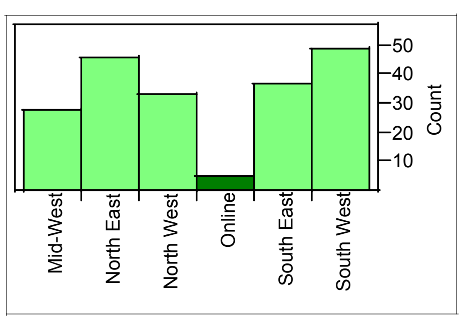
While the sales region scheme seemed defendable, some of the regions are still likely to be too small to be meaningful.
I devised a second scheme that classified each observation grossly into north-east, mid-west, south-east, south-west and north-west according to the following map.



The resulting distribution is as follows:

Distribution by Region

Region	Count	Distribution
Mid-West	28	14.14%
North East	46	23.23%
North West	33	16.67%
Online	5	2.53%
South East	37	18.69%
South West	49	24.75%



Vehicle External Color

It was truly amusing to see the variety of color names listed since Toyota produced the vehicle in only a few colors. Like state, in order to account for price variability across vehicle colors, color needed to be encoded in such a way that a reasonable number of levels are defined. I started with the basic assumption that owners did not flock to their local body shop to have their vehicles repainted, and the variety of listed colors was due purely to marketing. Recoding colors took less imagination than it took the dealers to come up with their color descriptions.

Contingency Table
Listed Color by Coded Color

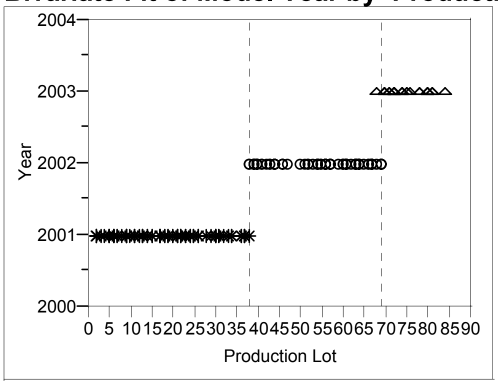
Count	Aqua	Blue	Green	Silver	White	Total
Aqua	18	0	0	0	0	18
Blue	0	19	0	0	0	19
Blue Green	0	1	0	0	0	1
Blue Moon Pearl	0	1	0	0	0	1
Champagne	0	0	0	0	1	1
Dark Blue	0	5	0	0	0	5
Dark Green	0	0	1	0	0	1
Green	0	0	27	0	0	27
Grey	0	1	0	0	0	1
Light Blue	0	3	0	0	0	3
Light Green	0	0	1	0	0	1
Mint	0	0	1	0	0	1
N/A	0	0	0	3	0	3
Pearl	0	0	0	0	1	1
Silver	0	0	0	75	0	75
Super White	0	0	0	0	2	2
Turquoise	0	1	0	0	0	1
White	0	0	0	0	37	37
	18	31	30	78	41	198

VIN Encoded 'Production Lot' as a proxy for Vehicle Age

The age of the vehicle was not available at any greater granularity than production year. However, a significant amount of production information is coded in the Vehicle Identification Number (VIN)[6]. By law each vehicle is required to bear permanently affixed VIN information that conforms to government standards. The VIN is an 18 digit code. The first three digit positions correspond to country (J=Japan), Manufacturer (T=Toyota), and Type of vehicle (2=passenger car). Digit positions four through eight capture model and body specific information. A complicated, and for the purposes of this analysis irrelevant, check digit scheme occupies position 9. The production year is encoded in position ten. The assembly plant code can be found in position eleven. Positions twelve through eighteen are taken by a manufacturer specific sequential production number. Examination of this data shows the sequence is continuous across model years. In fact, when ordered by production sequence, the data exhibit perfect separation across model years.

[6] VIN Library at www.autoinsurancetips.com/toyota_vin.htm

Bivariate Fit of Model Year by 'Production Lot'



Given the relation between VIN encoded sequential production number and production year, I have tentatively dubbed the 10^4 through 10^6 digits of the production sequence number as 'production lot'.

Treatment of Missing Predictor Variables

While most of the data elements were complete for all observations, a small number of predictor attributes were not populated. In total, 22 observations were missing one or more data elements. In order to retain these observations in the analysis, the following substitutions were made:

Attribute Name	Substitution Strategy	VIN (model year) for observations with Missing Data	Substitution Value
Vehicle Color	Most Probable Value	JT2BK18U720054665 (2002) JT2BK18UX20054787 (2002) JT2BK18U020061621 (2002) JT2BK18U430078682 (2003)	Silver Silver Silver Silver
Production Lot	Mean for Model Year	7 Auction vehicles, VIN N/A (2001) 2 Carmax vehicles: stock# 1894805 (both 2002) stock# 1906865	19 56 56
Mileage	Mean for Production Lot	JT2BK12U610011765 (2001) JT2BK18U120039501 (2002) JT2BK18U820054139 (2002) JT2BK18U720054665 (2002) JT2BK18UX20054787 (2002) JT2BK12U320062089 (2002)	40,179 24,824 10,464 10,464 10,464 8,939
Price	Average for vehicles w/I 1000 mile bracket	JT2BK18U820054139 (2002) JT2BK12U020057125 (2002) JT2BK18U020061621 (2002)	$19,683 $21,200 $19,636

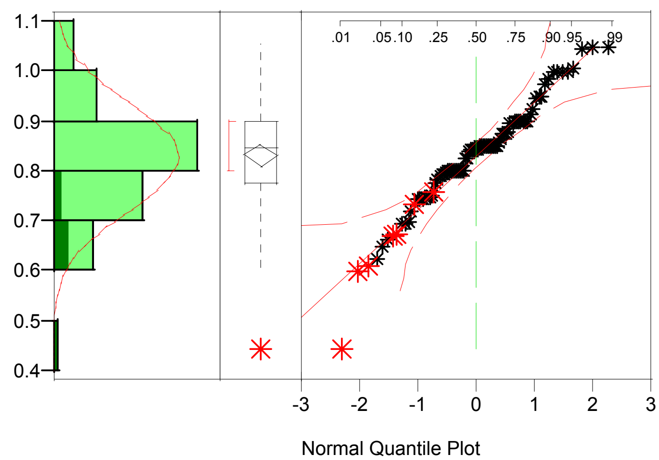
Examination of Data Structure

With data collected and preliminarily encoded for analysis, the response variable will be examined for normality and outliers and the explanatory variables will be examined for linearity with the response variable.

Response Variable: MSRP Residual

The response variable for this analysis and modeling exercise is MSRP residual. In lease terms, the MSRP residual is the percentage of the original Manufacturers Suggested Retail Price that the vehicle cost at the end of the lease. The MSRP residual is calculated by dividing the listing price by the original MSRP. This amount is expected to be bounded between 0.0 and 1.0, but if the listing price of the used vehicle exceeds the original MSRP, while irrational from an economic prospective, the used vehicle is listing for more than the original price and MSRP residual will exceed 1.0. Since I intend to lease the vehicle for 36 months, the data in the analysis has been restricted to 2001 model year vehicles.

Distributions MSRP Residual for Model Year 2001



Quantiles

100.0%	maximum	1.0502
90.0%		0.9879
75.0%	quartile	0.8987
50.0%	median	0.8450
25.0%	quartile	0.7749
10.0%		0.6907
0.0%	minimum	0.4451

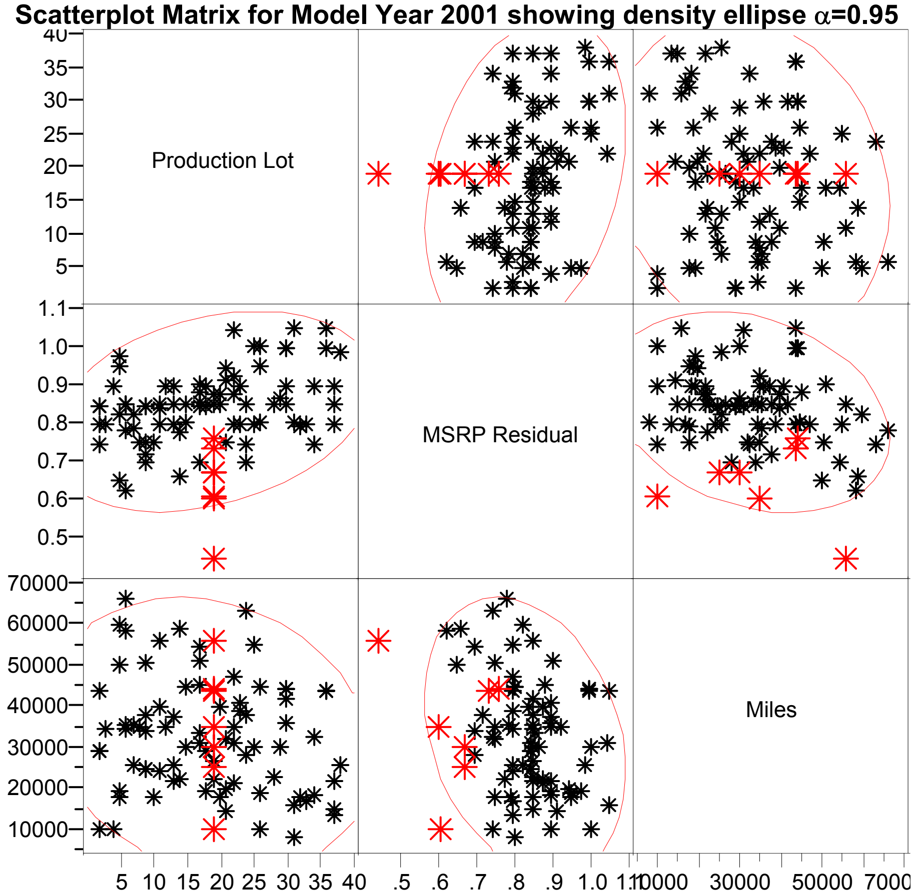
The above plot shows a histogram distribution, outlier box plot, and normal plot for the model year 2001 vehicles in the sample. Reviewing the normal plot, MSRP residual seems to be normally distributed with minimum value 0.4451 and maximum value 1.0502. The outlier box plot shows the minimum value of 0.4451 to be an outlier. I went back and reviewed this source of data for this observation. The outlier observation was one of the seven observations collected from the auction. All seven auction observations are shown in red on the normal plot. I reverified the data for this specific observation and found it to be consistent with what the auction listed. Since I can't determine any source of irregularity in the data collection, I will leave this point in the sample.

Continuous Numeric Explanatory Variable: Production Lot and Mileage

Multivariate Correlations for Model Year 2001

	Production Lot	MSRP Residual	Miles
Production Lot	1.0000	0.3027	-0.1733
MSRP Residual	0.3027	1.0000	-0.2713
Miles	-0.1733	-0.2713	1.0000

Scatterplot Matrix for Model Year 2001 showing density ellipse α=0.95



The two continuous explanatory variables in the dataset are Miles and Production lot. Both are shown in a scatter plot matrix against the dependent outcome MSRP Residual. Miles and Production Lot are shown to have a low level of multicollinearity. In fact, if we model MSRP Residual using only these two variables, we see the VIF statistic to be less than 10 for both production lot and miles.

Parameter Estimates

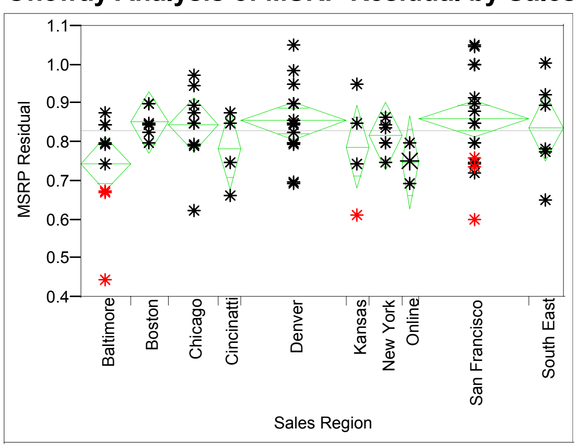
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.8634862	0.029284	29.49	<.0001	.
Production Lot	0.0019933	0.000419	4.75	<.0001	2.2451434
Miles	-0.000002	6.28e-7	-3.09	0.0023	2.2451434

Since neither miles or production lot demonstrate strong linear patterns against the response variable, I investigated several transformations of x to see if one could be found to improve linear fit.

Predictor Attribute	Transform	R^2	F	Prob >F
Miles	None – Baseline	0.0637	5.5745	0.0206
	Centered Polynomial of X	0.0951	4.2576	0.0175
	Natural Log of X	0.0384	3.2764	0.0739
	Square Root of X	0.0516	4.4626	0.0877
	X squared	0.0814	7.2644	0.0085
	Inverse of X	0.0149	1.2405	0.2686
Production Lot	None – Baseline	0.0895	8.0647	0.0057
	Centered Polynomial of X	0.1175	5.3897	0.0063
	Natural Log of X	0.0470	4.0531	0.0474
	Square Root of X	0.0699	6.1622	0.0151
	X squared	0.1104	10.1760	0.0020
	Inverse of X	0.0141	1.1752	0.2815

Nominal Categorical Explanatory Variable: Sales Region, Region, and Color

Oneway Analysis of MSRP Residual by Sales Region



Oneway Anova Summary of Fit

Rsquare	0.142595
Adj Rsquare	0.042378
Root Mean Square Error	0.105585
Mean of Response	0.830136
Observations (or Sum Wgts)	87

Analysis of Variance

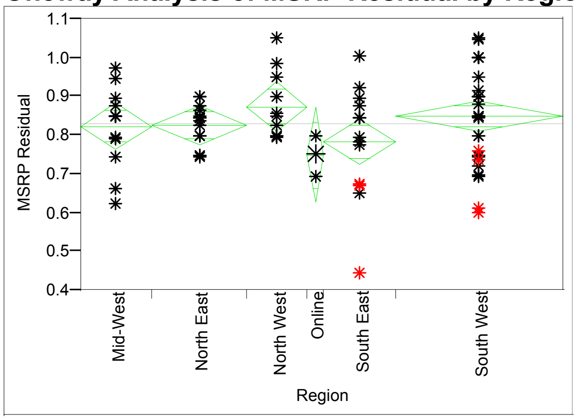
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Sales Region	9	0.142	0.0159	1.4229	0.1932
Error	77	0.858	0.0111		
C. Total	86	1.001			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Baltimore	9	0.743436	0.03519	0.67335	0.81352
Boston	7	0.852106	0.03991	0.77264	0.93157
Chicago	9	0.844572	0.03519	0.77449	0.91465
Cincinnati	4	0.783796	0.05279	0.67867	0.88892
Denver	19	0.855272	0.02422	0.80704	0.90351
Kansas	4	0.788822	0.05279	0.68370	0.89395
New York	6	0.816346	0.04310	0.73051	0.90218
Online	3	0.748437	0.06096	0.62705	0.86982
San Francisco	20	0.862396	0.02361	0.81538	0.90941
South East	6	0.838843	0.04310	0.75301	0.92468

Std Error uses a pooled estimate of error variance

Oneway Analysis of MSRP Residual by Region



Oneway Anova Summary of Fit

Rsquare	0.080479
Adj Rsquare	0.023718
Root Mean Square Error	0.106609
Mean of Response	0.830136
Observations (or Sum Wgts)	87

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Region	5	0.081	0.016	1.4179	0.2266
Error	81	0.921	0.011		
C. Total	86	1.001			

Means for Oneway Anova

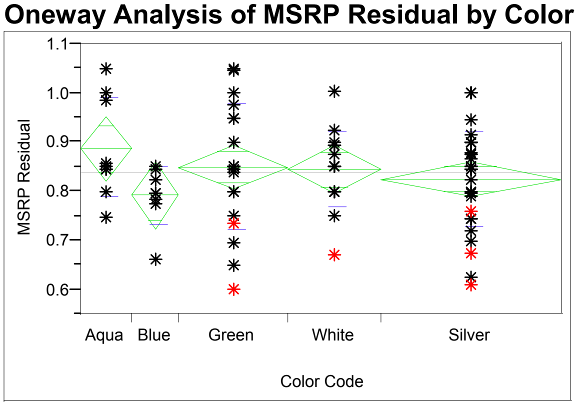
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Mid-West	13	0.823737	0.02957	0.76491	0.88257
North East	17	0.825345	0.02586	0.77390	0.87679
North West	11	0.873714	0.03214	0.80976	0.93767
Online	3	0.748437	0.06155	0.62597	0.87090
South East	13	0.782988	0.02957	0.72416	0.84182
South West	30	0.848245	0.01946	0.80952	0.88697

Std Error uses a pooled estimate of error variance

The previous page shows analysis of variance on the two geographic grouping schemes. The plots are meant to compare the viability of either scheme as a likely predictor attribute in a linear model. Note that all the groups in both schemes have a small number of observations. Also note that both schemes do not show significant evidence for rejecting the null hypothesis that $H_0 : \mu_1 - ... - \mu_n = \delta_0$ at even the 90% level and we can assume that the means are not equal across levels using either scheme. Only one scheme will be allowed to stay in the model. Attribute statistics will decide which scheme is retained.

Likewise, a similar analysis can be shown for coded color:

Oneway Analysis of MSRP Residual by Color Code Model Year 2001



Oneway Anova Summary of Fit

Rsquare	0.056853
Adj Rsquare	0.008486
Root Mean Square Error	0.099734
Mean of Response	0.837728
Observations (or Sum Wgts)	83

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Color Code	4	0.0468	0.0117	1.1755	0.3282
Error	78	0.7759	0.0099		
C. Total	82	0.8226			

Analysis of Variance Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Aqua	9	0.886466	0.03324	0.82028	0.95265
Blue	8	0.791442	0.03526	0.72124	0.86164
Green	19	0.849175	0.02288	0.80362	0.89473
White	16	0.844521	0.02493	0.79488	0.89416
Silver	31	0.825000	0.01791	0.78934	0.86066

Level Number Mean Std Error Lower 95% Upper 95%
Std Error uses a pooled estimate of error variance

The above analysis shows a one way analysis of variance of color against MSRP residual. Like the analysis for region schemes, the color scheme does not show significant evidence for rejecting the null hypothesis that $H_0 : \mu_1 - \ldots - \mu_n = \delta_0$ at even the 90% level and we can assume that the means are not equal across levels.

Linear Model Development

Preliminary Models

To develop and validate the model, I have selected a strategy of randomly divide the data into development and validation samples. A new 'Sample Usage' column was created in the dataset to randomly (Randomuniform) assigned half of the 2001 observations for development ('Dev') and the remaining observations for validation ('Val'). Several models were developed on the 'Dev' sample data to model response 'MSRP Residual' by combinations of Production Lot, Miles, Sales Region or Region, and color code. The first series of models set the most probable Sales Region (South West), Region (San Francisco) and Color (Silver) as reference levels. Since the dataset was small, I used the "All Possible Models" option on the Fit Model Platform to exhaustively examine all possible attribute combinations. The best model for each number of predictor terms was as follows:

Best Models Including Sales Region by Number of Attributes (16,383 models total)

Model Designation	Model Attributes	Number of Attributes	R^2	sqrt(MS)	Cp
SRM1	Sales Region {Baltimore & South East & Cincinnati &Chicago - Boston & New York & Denver & San Francisco & Kansas}	1	0.145962	0.250266	0.162834
SRM2	Miles ,Sales Region{Baltimore & South East & Cincinnati & Chicago - Boston & New York & Denver & San Francisco & Kansas}	2	0.282364	0.246135	-4.57121
SRM3	Production Lot, Miles ,Sales Region{Baltimore & South East & Cincinnati & Chicago - Boston & New York & Denver & San Francisco & Kansas}	3	0.301018	0.2075	-3.49211
SRM4	Production Lot, Miles ,Sales Region{Baltimore & South East & Cincinnati & Chicago - Boston & New York & Denver & San Francisco & Kansas},Color Code{Blue & White & Silver - Green & Aqua}	4	0.324396	0.186548	-2.64624
SRM5	Production Lot, Miles ,Sales Region{Baltimore & South East & Cincinnati & Chicago - Boston & New York & Denver & San Francisco & Kansas},Sales Region{Baltimore & South East - Cincinnati},Color Code{Blue & White & Silver - Green & Aqua}	5	0.343454	0.171685	-1.58711

Best Models Including Sales Region by Number of Attributes (16,383 models total)

Beyond five attributes, the process began to include multiple groupings of attributes that were already selected. The Cp statistic became positive and the R^2 ceased to show meaningful improvement. Therefore models with more than five attributes were not entertained.

Best Models Including Region by Number of Attributes (1023 models total)

Model Designation	Model Attributes	Number of Attributes	R^2	sqrt(MS)	Cp
RM1	Region{South East & Mid-West & North East - North West & South West}	1	0.132767	0.238687	3.466612
RM2	Miles ,Region{South East & Mid-West & North East - North West & South West}	2	0.265952	0.238874	-1.5159
RM3	Miles ,Region{South East & Mid-West & North East - North West & South West},Region{South East & Mid-West - North East}	3	0.289689	0.203558	-0.76034
RM4	Miles ,Region{South East & Mid-West & North East - North West & South West},Region{South East & Mid-West - North East},Color Code{Blue & White & Silver - Green & Aqua}	4	0.303941	0.180571	0.492455

Like the models containing sales region, the models containing region began to include multiple groupings of already selected attributes with the third selection. This is demonstrated in Models RM3 and RM4 where multiple instances of Region were included. The Cp statistic became positive and the R^2 showed marginal improvement beyond three attributes. Therefore models with more than three attributes were not entertained.

The preliminary modeling exercise suggests that the Sales region scheme is more predictive than the simpler region scheme. With R^2 values topping out in the low 30s, it also suggests that the data do not strongly support prediction of this outcome. Even so, models SRM2, SRM3, SRM4 and RM2 are selected for further evaluation. The development sample model parameter estimates and effect test are as follows:

SRM2: $R^2 = 0.282364$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	0.12116469	0.060582	8.4595
Error	43	0.30794273	0.007161	Prob > F
C. Total	45	0.42910742		0.0008

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	0.9313527	0.035996	25.87	<.0001	0.8587597	1.0039456
Miles	-0.000003	0.000001	-2.86	0.0065	-0.000005	-8.564e-7
Sales Region{Baltimore & South East & Cincinnati & Chicago - Boston & New York & Denver & San Francisco & Kansas}	-0.041046	0.012732	-3.22	0.0024	-0.066722	-0.015369

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Miles	1	1	0.05853147	8.1731	0.0065
Sales Region	1	1	0.07442978	10.3931	0.0024

For Model SRM2, the Effect Tests show both parameters are significant at the 5% level.

SRM3: $R^2 = 0.301018$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	0.8961127	0.04899	18.29	<.0001	0.7972464	0.994979
Production Lot	0.0013596	0.001284	1.06	0.2958	-0.001232	0.0039514
Miles	-0.000003	0.000001	-2.48	0.0172	-0.000005	-4.877e-7
Sales Region{Baltimore & South East & Cincinnati & Chicago - Boston & New York & Denver & San Francisco & Kansas}	-0.037459	0.013158	-2.85	0.0068	-0.064013	-0.010906

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Production Lot	1	1	0.00800441	1.1208	0.2958
Miles	1	1	0.04396851	6.1569	0.0172
Sales Region	1	1	0.05788206	8.1052	0.0068

For Model SRM3, the Effect Tests show Production Lot is not significant at the 5% level.

SRM4: $R^2 = 0.303941$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	0.8869014	0.049357	17.97	<.0001	0.7872219	0.9865808
Production Lot	0.0016744	0.001305	1.28	0.2067	-0.000961	0.0043097
Miles	-0.000002	0.000001	-2.16	0.0370	-0.000004	-1.466e-7
Sales Region{Baltimore & South East & Cincinnati & Chicago - Boston & New York & Denver & San Francisco & Kansas}	-0.033063	0.013603	-2.43	0.0195	-0.060535	-0.005591
Code{Blue & White & Silver - Green & Aqua}	-0.016712	0.01403	-1.19	0.2405	-0.045046	0.0116234

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Production Lot	1	1	0.01164148	1.6464	0.2067
Miles	1	1	0.03286047	4.6473	0.0370
Sales Region	1	1	0.04177233	5.9076	0.0195
Coded Color	1	1	0.01003151	1.4187	0.2405

For Model SRM4, the Effect Tests show Production Lot and Coded Color are not significant at the 5% level.

RM2: $R^2 = 0.265952$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	0.11412203	0.057061	7.7896
Error	43	0.31498539	0.007325	Prob > F
C. Total	45	0.42910742		0.0013

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	0.942337	0.036704	25.67	<.0001	0.8683161	1.0163578
Miles	-0.000003	0.000001	-2.79	0.0078	-0.000005	-7.981e-7
Region{South East & Mid-West & North East-North West &South West}	-0.038767	0.012782	-3.03	0.0041	-0.064544	-0.012991

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Miles	1	1	0.05715062	7.8019	0.0078
Region	1	1	0.06738711	9.1993	0.0041

For Model RM2, the Effect Tests show both parameters are significant at the 5% level.

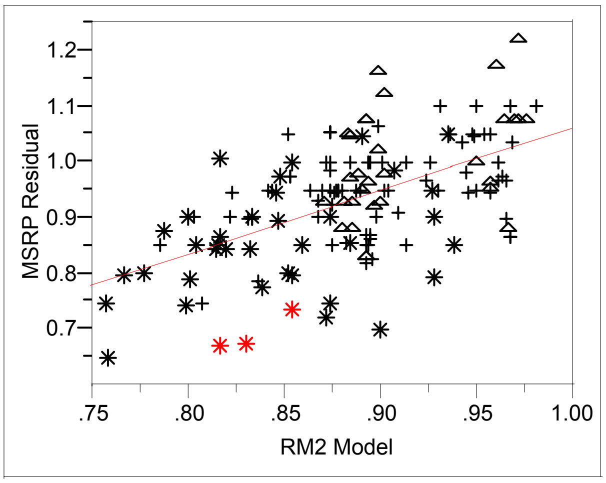
Evaluation of all four preliminary models shows the two-attribute models are the most robust on the development data. Leverage and residual plots for models SRM2 and RM2 (not shown) were examined and do not display any obvious patterns. The Press statistics are Model SRM2=0.35066, Model RM2=0.36011.

Model Validation

Approximately half of the 2001 model year observations and all 2002 and 2003 model year observations were withheld from the development exercise for use in model validation. The below charts show fit curves of the known outcome (MSRP residual) against each individual model. A Summary of Fit and Analysis of Variance was included on the linear fit between each model and the expected outcome. While neither model did not fit well to just the remaining 2001 model year vehicles alone (∗ symbol), the fit across all remaining data was excellent (2002 shown with + and 2003 shown with a △ symbol).

Fit Y by X Group

Bivariate Fit of MSRP Residual by RM2 Model



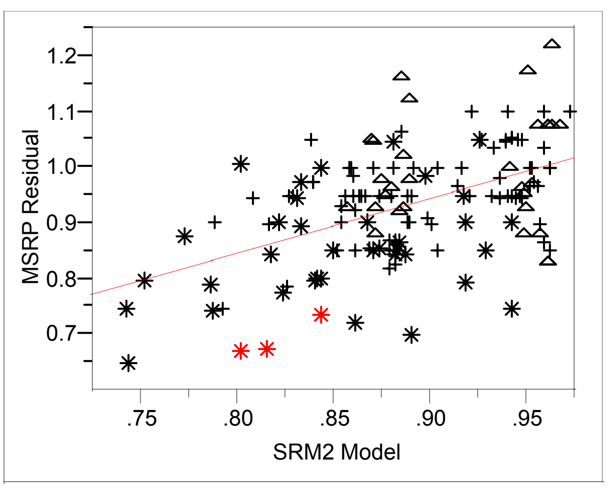
Summary of Fit

RSquare	0.316929
RSquare Adj	0.312282
Root Mean Square Error	0.086633
Mean of Response	0.934186
Observations (or Sum Wgts)	149

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.5118879	0.511888	68.2045
Error	147	1.1032635	0.007505	Prob > F
C. Total	148	1.6151514		<.0001

Bivariate Fit of MSRP Residual by SRM2 Model

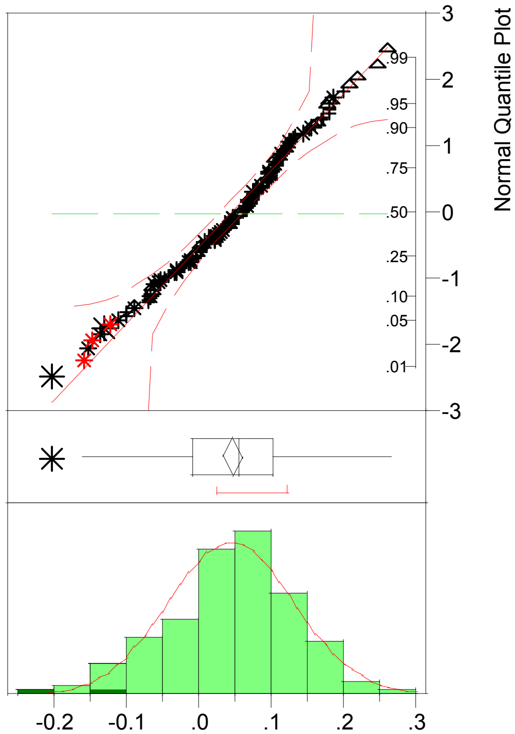


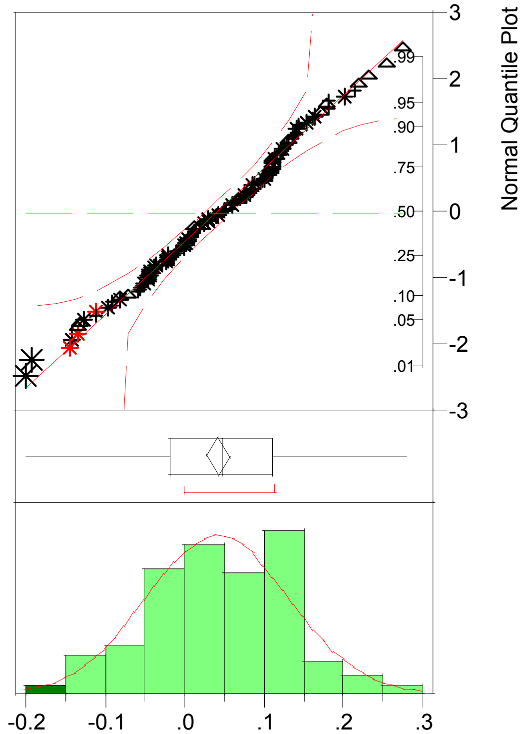
Summary of Fit

RSquare	0.25201
RSquare Adj	0.246921
Root Mean Square Error	0.090656
Mean of Response	0.934186
Observations (or Sum Wgts)	149

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.4070335	0.407034	49.5266
Error	147	1.2081178	0.008218	Prob > F
C. Total	148	1.6151514		<.0001





Above are shown distributions of the RM2 and SRM2 model residuals respectively for the validation data. While the RM2 model shows an outlier observation, both plots show that the residuals are normally distributed.

Predicted MSRP Residual:

Individual prediction intervals of the MSRP residual for both models for a vehicle with 36,000 miles in the Metropolitan DC area are as follows:

Model	Predicted MSRP Residual	Lower 95% Prediction Interval	Upper 95% Prediction Interval
RM2	0.800	0.624	0.976
SRM2	0.786	0.610	0.961

For all practical purposes, the two outcomes are the same.

Therefore, if the Honda Civic Hybrid follows a similar depreciation curve in its first three years of Toyota Prius production, a Honda Civic leased today should be worth between 61% to 97% of its current selling value. With this figure in hand, the expected lease payments for 36 month lease on a 2004 Honda Civic Hybrid assuming the current MSRP for the Civic is $19,550, $1,000 down (cap cost reduction) and a current market interest rate of 6.0% would be calculated as follows:

Net. Cap Cost = Capitalized Cost - Cap. Cost Reduction = $19,550 - $1,000 = $18,550

Depreciation Fee = (Net. Cap Cost - Residual) / Lease Term=

($18,550 – [$18,550 x 0.61])/ 36 = $200.96

Money factor = Interest Rate / 2400 = 6.0 / 2400 = 0.0025

Finance Fee = (Net. Cap Cost + Residual) x Money Factor=

(($18,550 + [$18,550 x 0.61]) x 0.0025 = $74.67

Maryland Tax = 5% x Capitalized Cost = 0.05 x $19,550 =

$977.50/36 = $27.15 per payment.

Monthly Payment = Depreciation Fee + Finance Fee + Tax =

$200.96 + $74.67 +$27.15 = **$302.78**

Since the majority of the lease cost is generated by the depreciation fee, higher depreciation values from the model prediction or upper prediction interval would result in a lower payments.