Buying a Honda Odyssey with Linear Algebra

Andy Long

March 28, 2012

Abstract

In 2001 I decided that my wife and I were going to need a van, because we were expecting a baby in November. I wanted a used van, and wanted a reliable van, so I visited Consumer Reports and checked into "reliable used vans." At that time only two came up: the Toyota Previa, and the Honda Odyssey. It was sad that there were no American vehicles on the list, but *que sera sera*.

My older son and wife were opposed to the Toyota Previa, based on appearance (which hurt me deeply, a devoted Toyota fan). So I resigned myself to the Honda Odyssey. I'm glad that I did: ours, purchased via a little linear algebra, has worked well for 150,000 miles. This short report details the process whereby I made my choice from among the Odysseys available at that time.

1 The Data

First I went on-line, to autotrader.com, to find out what used Odysseys were available in our area. I also checked to see what characteristics were listed that I could use in a model, and settled on just year and mileage. While other information was available (e.g. color, extras), they do not appear in this model. They did, however, figure into my final choice.

year	mileage	price
1998	41296	19988
1998	64959	17995
1997	48755	17988
1997	40030	17595
1998	44249	15950
1997	42973	15695
1995	66129	13995
1997	93061	12988
1995	118006	11495
1995	126012	8995
1997	59990	18995
1998	54141	16995
1996	75730	14995
1997	85175	13995
1995	50200	12895
1997	97488	11995

2 The Model

My objective was to create a model for the price of vehicle (or rather "value" of vehicle), based on age and mileage. While one could argue about whether the model is, in fact, linear or not, I used a linear model here. That is, Price P is a function of age a and m:

$$P(a, m) = P_0 + \beta_a a + \beta_m m$$

where

1. P_0 is an intercept term (representing the price of a new car, essentially, since it corresponds to $P(0, 0)$ – a car of age 0, with 0 miles);

2. β_a is the coefficient multiplying the age of the vehicle (and this should be negative, since the older the car is, the lower the price should be); and

3. β_m is the coefficient multiplying the mileage of the vehicle. Again, this should be negative, since the more miles a car has, the lower the price should be.

In terms of linear algebra, we seek a solution to the following augmented linear system:

$constant$	$age(2002 - year)$	$Miles(K)$	$Price(\$)$
1	4	41.296	19988
1	4	64.959	17995
1	5	48.755	17988
1	5	40.03	17595
1	4	44.249	15950
1	5	42.973	15695
1	7	66.129	13995
1	5	93.061	12988
1	7	118.006	11495
1	7	126.012	8995
1	5	59.99	18995
1	4	54.141	16995
1	6	75.73	14995
1	5	85.175	13995
1	7	50.2	12895
1	5	97.488	11995

As you may well imagine, it is unlikely that this will have a solution (it is likely to be an inconsistant system). So we seek a system with minimal inconsistancy, essentially, which is achieved via linear regression. Linear regression replaces a system $A\mathbf{x} = \mathbf{b}$, with the square system $A^{trans}A\mathbf{x} = A^{trans}\mathbf{b}$ – which can generally be solved ($A^{trans}A$ is generally invertible).

To obtain the coefficients, we carry out a linear regression with the following results:

Table 1: Results of a linear regression to obtain a model for price of vehicle as a function of age (measured as 2002 minus model year) and mileage (in thousands of miles).

Coefficient	Estimate	SE	Probability
P_0	25409.1	(1912.67)	0.00000
β_a	-1060.23	(413.509)	0.02356
β_m	-66.6603	(17.1237)	0.00185
R^2:	0.773033		
$\hat{\sigma}$:	1553.92		
Number of cases:	16		
Degrees of freedom:	13		

3 Results and Analysis

All of the coefficients of our linear regression are significantly different from 0 at the .05 level. Hence we include them all, and obtain a model for the price of a vehicle of the form

$$P(a, m) = 25409 - 1060.23a - 66.66m$$

We would interpret the -1060.23 as measuring the devaluation (in dollars) due to each additional year in age; the -66.66 as the devaluation (in dollars) due to each additional 1000 miles.

Now, the actual price of vehicle i is P_i, with age a_i and mileage m_i. The estimate of the price is $P(a_i, m_i)$.

If we really believe that our model does a good job of estimating the value of a vehicle, then we compare the price the owner asks to the estimate, and we observe that if the price the owner asks is higher than the estimated value, then it is over-priced; if the price asked is lower than the estimated value, then it is under-priced. We want to buy the most under-priced vehicle.

But the level of over- or under-priced vehicle is related to the actual price. There are (at least) two ways we might evaluate that. We might do

$$100\frac{P_i - P(a_i, m_i)}{P_i},$$

which essentially gives the relative percentage deviation from the estimated value; or we might take the relative price per value:

$$\frac{P_i}{P(a_i, m_i)}$$

They both give the same result, however, in terms of the order of the vehicles. So the best car is the one with the lowest price per value number or the lowest percentage deviation.

So here are the results of our cost comparisons, ordered by relative value:

year	mileage	price	estimated value	percent deviation	relative value
1998	44249	15950	18218.55	-14.22	0.875
1995	50200	12895	14641.18	-13.54	0.881
1997	97488	11995	13609.40	-13.46	0.881
1997	42973	15695	17243.39	-9.87	0.910
1997	93061	12988	13904.51	-7.06	0.934
1995	126012	8995	9587.53	-6.59	0.938
1998	54141	16995	17559.15	-3.32	0.968
1997	85175	13995	14430.19	-3.11	0.970
1997	40030	17595	17439.57	0.88	1.009
1995	66129	13995	13579.35	2.97	1.031
1997	48755	17988	16857.96	6.28	1.067
1998	64959	17995	16838.02	6.43	1.069
1996	75730	14995	13999.57	6.64	1.071
1998	41296	19988	18415.40	7.87	1.085
1995	118006	11495	10121.21	11.95	1.136
1997	59990	18995	16109.03	15.19	1.179

4 Conclusion

In the end I purchased the 1995 with 50,200 miles at $12,895 (which rated second best in my analysis). It was an EX (a premium model), with sunroof, alloy wheels, CD, power everything, captain chairs, etc. The exterior was in great condition (the car must have been garaged). I also liked the styling of the 1995 – it was more like a station wagon. So, in the end, I went with the 1995. It's been a good 11 years!

Thanks to linear algebra for pointing me to it....

It is certainly possible to critique the model I used. For example, based on a depreciation of $1060/year, we find that at an age of about $a = 24$ years, a car with 0 miles is worth nothing ($25409 - 1060a = 0$). And if you can wait another year, someone will have to pay you $1060 to take it off their hands! Hence **functions that asymptote to zero** might be more realistic model forms. The only problem is that they're **non-linear** (and we're focused on linear algebra here!).