D K
M D
Math200
03/07/06
Prof. Brown

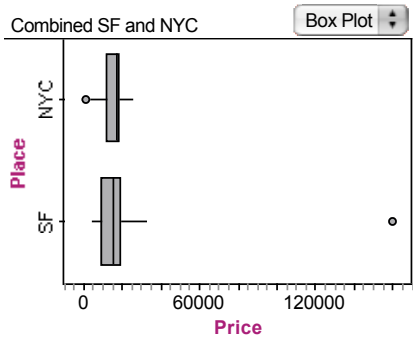
Used Car Project

In our used car project, we compared the used car, Honda Accord, between the two cities of San Francisco and New York. We had collected our data from the website www.cars.com within a seven (7) day period during the week of February 16th 2006. We obtained 106 used Honda Accords from New York and 84 used Honda Accords from San Francisco. The used cars vary in that we have cars from 1988 to present day 2006. With our data, we will discuss the relationship between the price, mileage, age and body of the used Honda Accords. We are trying to find out why the similar cars of the Honda Accords had different prices between the two cities of New York and San Francisco. We will obtain these results by looking at the used Honda Accords within the two cities, San Francisco and New York, and doing several comparisons. These would entail a Quantitative Variable Comparison between place and price, and place and age, a Categorical Variable Analysis between place and body type, and a Comparison between two variables via a regression analysis between price and mileage which will show how mileage affects price.

Combined SF and NYC

		Price
Place	NYC	106
		15417.736
		1300
		11995
		16941.5
		18995
		25995
		5405.6079
	SF	84
		16738.786
		3400
		8995
		14989.5
		19960
		159888
		17243.881
Column Summary		190
		16001.779
		1300
		9995
		15995
		18995
		159888
		12134.632

S1 = count( )
S2 = mean( )
S3 = min( )
S4 = Q1( )
S5 = median( )
S6 = Q3( )
S7 = max( )
S8 = s( )



Quantitative Variables Comparison

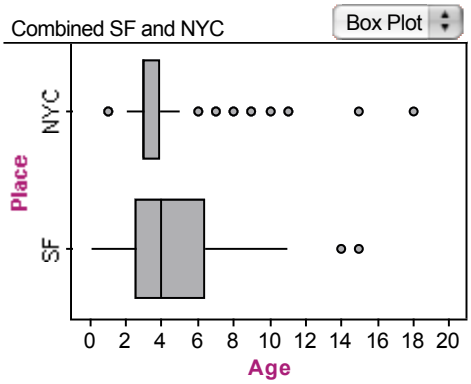
In both New York and San Francisco the mean price of an accord are pretty close in price, only about $1000 difference. There were more Accords for sale in New York but not by too much. The over all prices are clustered around the same prices ranging from about $1300 to $3500. It doesn't look like it but the two cities do have a large spread in the prices. You can see that the spread of the prices are very different, to see this, we would look at the standard deviation, which tells us how much variation is in the spread. The standard deviation for New York City is 5405.6079 and the standard deviation for San Francisco is 17243.881. We can see that the spread in San Francisco is much more than the spread in New York City. As you can see, there are not many outliers for both

cities. The formula for outliers is anything lower than Q1-IQR (1.5) and anything higher than Q3+IQR (1.5). Anything that is lower than $11,995 and higher than $18,995 is an outlier for Accords in New York City. Anything lower than $8,995 and higher than $19,960 for San Francisco is an outlier for Accords in San Francisco. For New York City the shape of the box plot seems to be more symmetrical than skewed to any one side. On the other hand San Francisco's box plot is skewed to the right because the whisker is longer on the right side than the left, also because of its outlier all the way on the right where the price is $159,888.

Combined SF and NYC

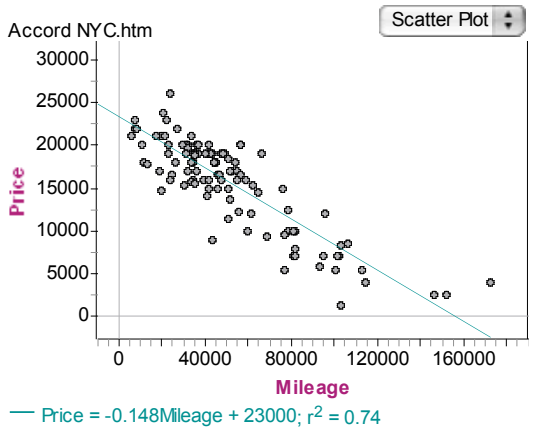
		Age
Place	NYC	106
		4.3679245
		1
		3
		3
		4
		18
		2.8328867
	SF	84
		4.7738095
		0
		2.5
		4
		6.5
		15
		3.2426049
Column Summary		190
		4.5473684
		0
		3
		4
		6
		18
		3.0194026

S1 = count( )
S2 = mean( )
S3 = min( )
S4 = Q1( )
S5 = median( )
S6 = Q3( )
S7 = max( )
S8 = s( )



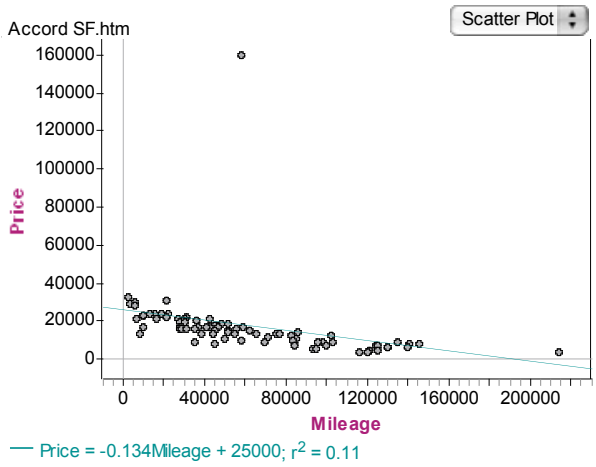
For the ages of the accords, the mean of both cities are pretty new and the median is about 4 years, but because of the outliers, the means are 4.36 for NYC, and 4.77 for

San Francisco. For New York City there are more accords that were about 3-4 years, whereas San Francisco as about the amount of accords from about 1-7 years. The spread of the age is reasonable considering how long this car has actually been around. The range in age for San Francisco is slightly higher than the range in age for New York City because the standard deviation for San Francisco is 3.24 and 2.83 for San Francisco. The shape of the age distribution is skewed to the right for both cities. On average the Honda Accords in San Francisco are being sold at older ages than in New York City because the Q1, median, and mean for San Francisco are higher than New York City's. For New York City, anything lower than 3 years old or higher than 4 years old is an outlier. For San Francisco, anything lower than 2.5 years old or higher than 6.5 years old is an outlier.

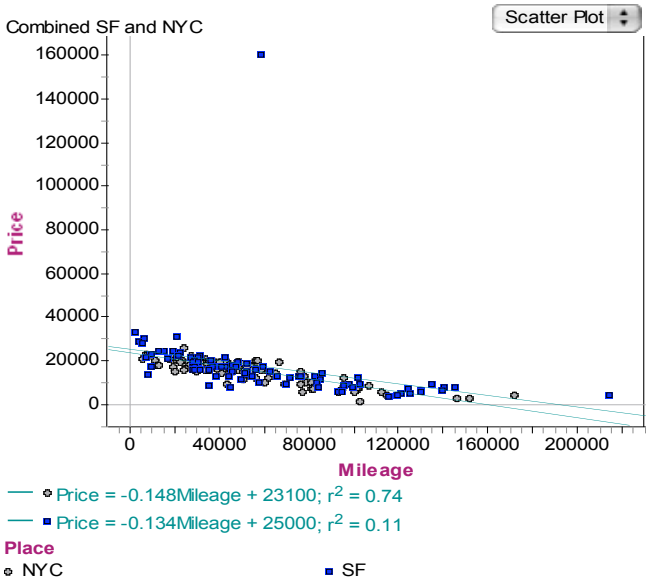


Comparison between two variables

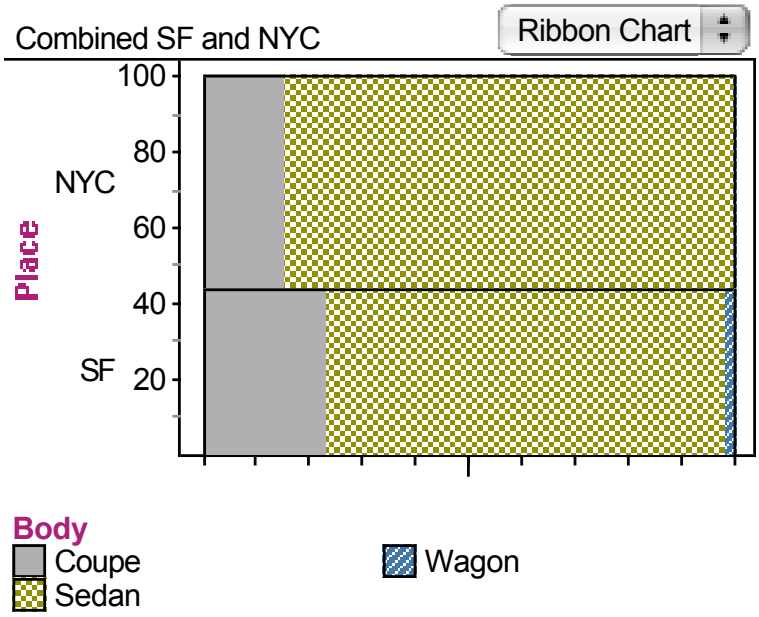
In the scatter plot above, there are variations of the Honda Accord in New York City. The equation of the least squares is Price = -0.148Mileage + 23000. The slope of -0.148 tells us that for every 10,000 miles that gets added to the mileage of the car, the price decreases by $1,480. Judging by the scatter plot, the variables have a moderately strong, negative association. This tells us that when the car has more mileage, the price decreases. There is a negative correlation for the data. For every decrease in mileage, there is an increase in price. The r-squared (r^2) tells us the proportion of variability explained. In this case r-squared (r^2) is equal to 0.74, which tells us that 74% of the variation is accounted for by the least squares regression line.



In the scatter plot above, there are variations of the Honda Accord in San Francisco. The equation of the least squares is Price = -0.134Mileage + 25000. The slope of -0.134 tells us that for every 10,000 miles that gets added onto the car, the price decreases by $1,340. Judging by the scatter plot, the variables have a strong, negative association. This tells us that when the car has more mileage, the price decreases. The correlation here is direct. For every decrease in mileage, there is a decrease in price. The r-squared (r^2) tells us the proportion of variability explained. In this case r-squared (r^2) is equal to 0.11, which tells us that 11% of the variation is accounted for by the least squares regression line. With the presence of the outlier, it would greatly alter the least squares regression line. We believe that is why, although this has a stronger negative association than New York, the r-squared (r^2) is so small compared to New York.



This scatter plot has both, New York City and San Francisco's price versus mileage data. The least squares regression line for NYC is Price = -0.148Mileage + 231000; with an r-squared (r^2) of 74%, whereas the least squares regression line for SF is Price = -0.134Mileage + 25000; with an r-squared (r^2) of 11%. The price decrease for NYC is $1,480 which is a greater decrease than San Francisco's $1,340. Also, their proportion of variation explained is very different in that NYC has an r-squared (r^2) of 74% whereas, SF has an r-squared (r^2) of 11%. This huge difference is a result of the outlier for San Francisco. If the outlier was closer to the least squares regression line for San Francisco, its proportion of variation explained would be about 68% which would make more sense.



Combined SF and NYC

		Body			Row Summary
		Coupe	Sedan	Wagon	
Place	NYC	16	87	0	103
	SF	19	60	1	80
Column Summary		35	147	1	183

S1 =

Categorical Variable Comparison

By looking at the summary table you can see that the variables place and body are not independent because if they were independent San Francisco and New York City would have the same proportions of coupes, sedans, and wagons. San Francisco has more coupes and wagons than New York City. But New York City has more sedans than San Francisco.

	NYC	SF	Both
Sedans	84.4%	75.0%	80.3%
Coupes	15.6%	23.7%	19.1%
Wagons	0.0%	1.3%	0.1%
Total	100.0%	100.0%	100.0%

Proportion of coupes in NYC = .155
Proportion of sedans in NYC = .844
Proportion of wagons in NYC = 0

Proportion of coupes in SF = .237
Proportion of sedans in SF = .75
Proportion of wagons in SF = .0125

Proportion of coupes total = .191
Proportion of sedans total = .803
Proportion of wagons total = .0054

Conclusion

After investigating the Honda Accords within the two different cities, we discovered that the Accords are more expensive in San Francisco than in New York City. Within our data, there are only 84 Accords in San Francisco and 106 Accords in New York City. It appears that both cities prefer the Sedan body type over any other one with wagon's being the least. The average price of the car, the mean, is much lower in New York City, which indicates why NYC has more Accords than SF. For every year that passes, the Accords in San Francisco decrease by $1,340 whereas in New York City, prices decrease by $1,480. Yet, when compared without the outlier, San Francisco does have a lower depreciation value than New York City. The proportion of variation explained (r^2) for San Francisco indicate that 17% of the variation in accounted for by the model, whereas in New York City, 74% of the variation is accounted for. The (r^2) for S.F. is lower than NYC because of the huge outlier in the data for S.F. If the data

point was removed, S.F would have had a higher (r^2). So, in conclusion, through our data, the prices in San Francisco are lower than in New York City.