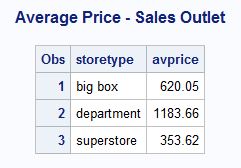
**Shruthi Sreenivasa Murthy Retail TV Purchases (Report)**

1. **Considering cases in tvfitting2017:**
   1. **How does the average price of the purchased TV depend on sales outlet?**



The Average Price of the purchased TV at a big box store is $620.05.

The Average Price of the purchased TV at a department store is $1183.66.

The Average Price of the purchased TV at a superstore store is $353.62.

* 1. **How does the payment method depend on sales outlet?**

The number of Credit card payments = 17422 [which ads up to 72.68% of the overall payments]

The number of Debit card payments = 2245 [which ads up to 9.37% of the overall payments]

The number of Premium card payments = 4303 [which ads up to 17.95% of the overall payments]

OUT OF WHICH:

Total Number of Credit Card payments at a bigbox store = 4253

Total Number of Debit Card payments at a bigbox store = 394.

Total Number of Premium Card payments at a bigbox store = 1346.

Which means the Total number of Card Payments at the Big Box Store = 5993.

Total Number of Credit Card payments at a Department store = 1440.

Total Number of Debit Card payments at a Department store = 89.

Total Number of Premium Card payments at a Department store = 864.

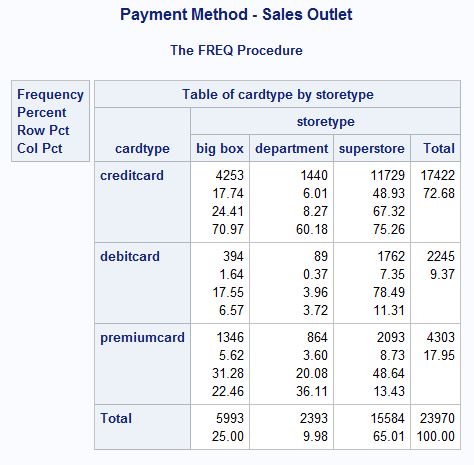
Which means the Total number of Card Payments at the Department Store = 2393.

Total Number of Credit Card payments at a superstore = 11729.

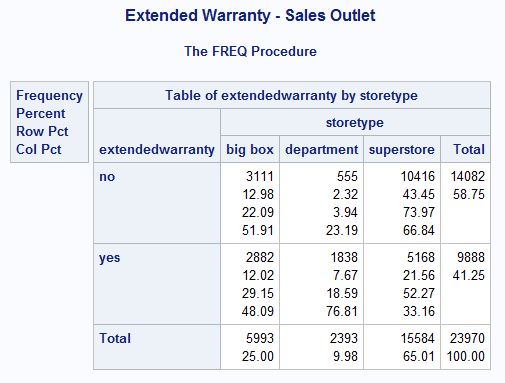
Total Number of Debit Card payments at a superstore = 1762.

Total Number of Premium Card payments at a superstore = 2093.

Which means the Total number of Card Payments at the superstore = 15584.



**c. How does the likelihood of buying the extended warranty depend on sales outlet?**



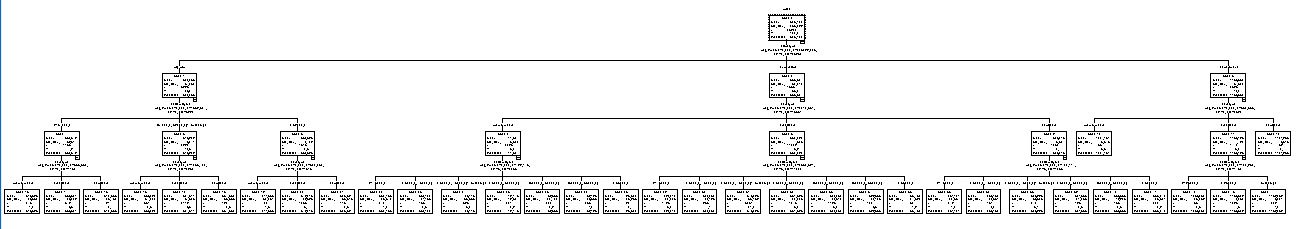
The total number of Customers who chose to buy Extended Warranty in the Bigbox store = 2882, and the total number of customers who chose **not** to buy the extended warranty in the Bigbox store = 3111.

The total number of Customers who chose to buy Extended Warranty in the Department store = 1838, and the total number of customers who chose **not** to buy the extended warranty in the Department store = 555.

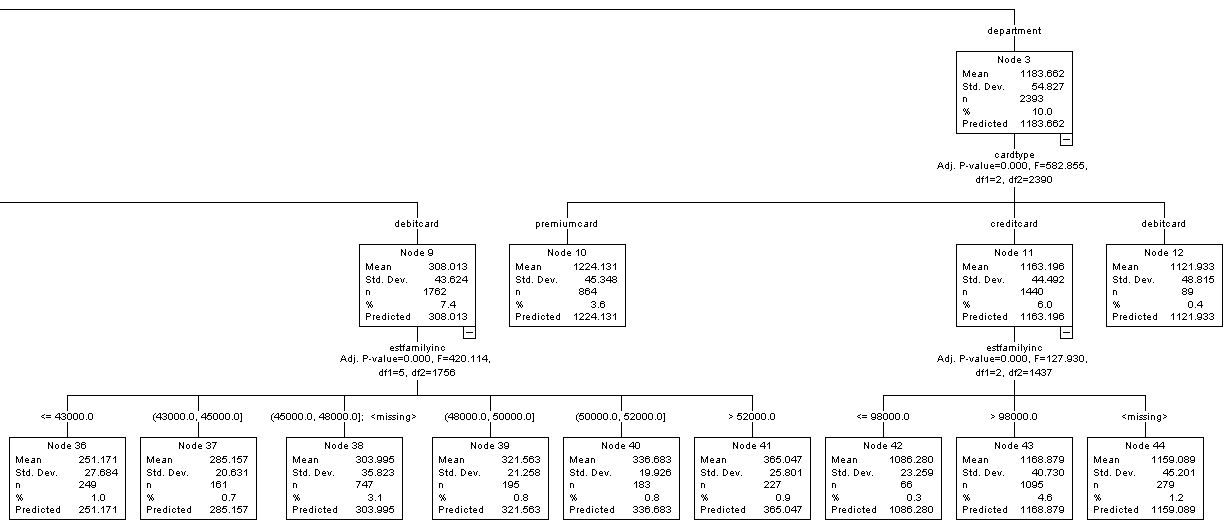
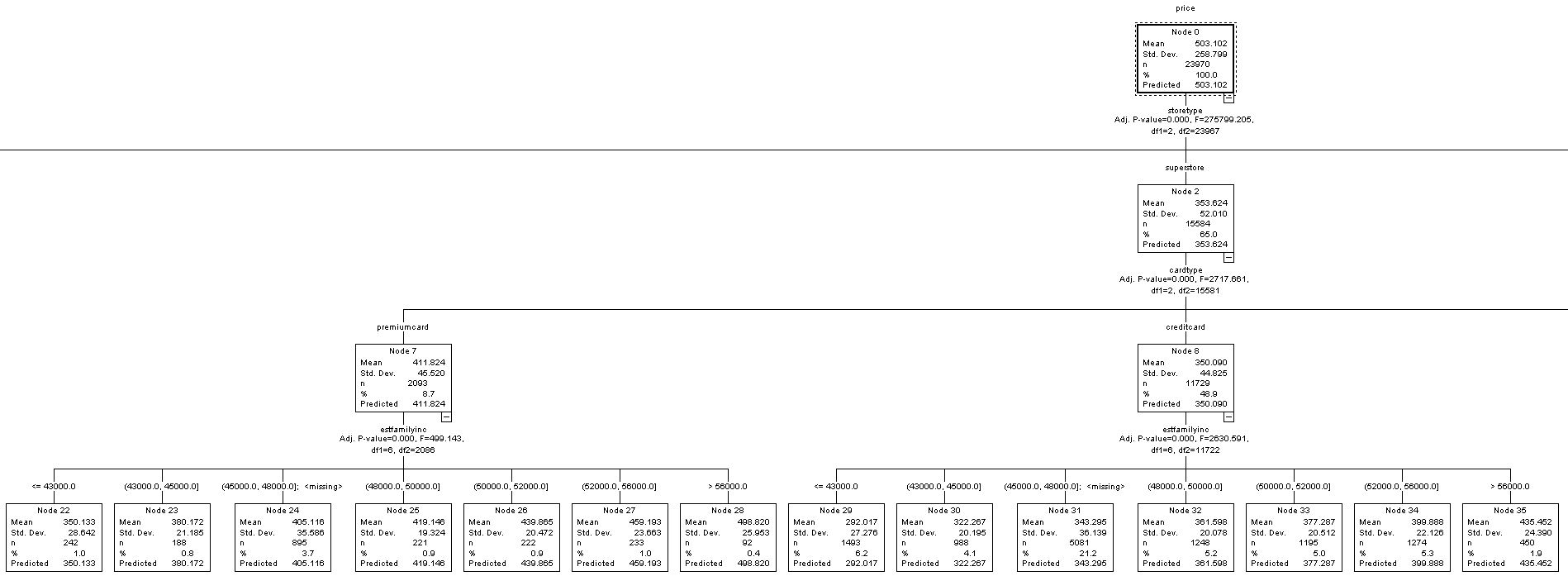
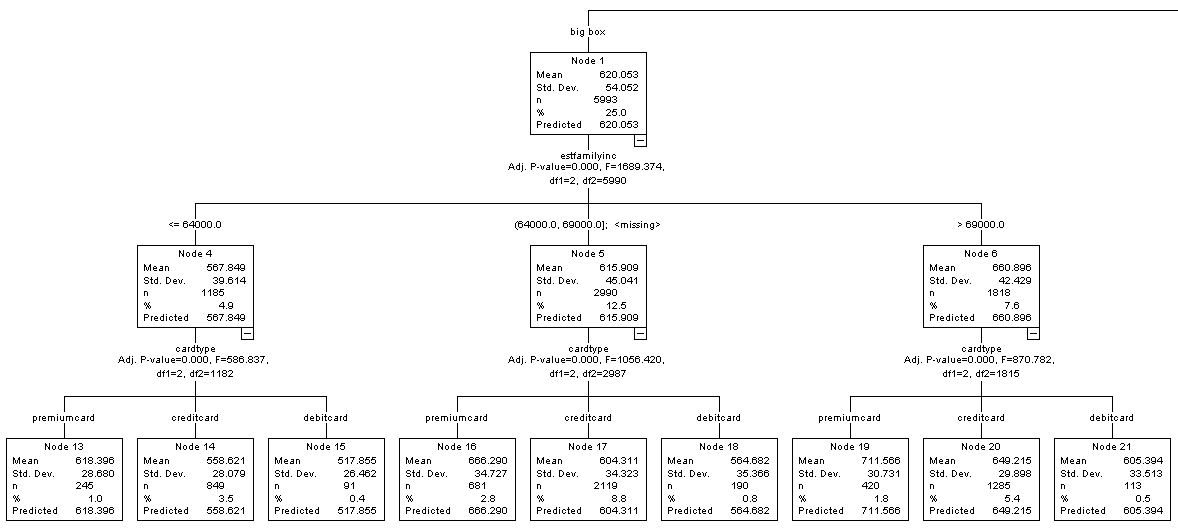
The total number of Customers who chose to buy Extended Warranty in the superstore = 5168, and the total number of customers who chose **not** to buy the extended warranty in the superstore = 10416.

So, the total number of customers who chose to buy the extended warranty = 9888, which adds up to 41.25% and the total number of customers who chose not to buy the extended warranty = 14082, which adds up to 58.75%.

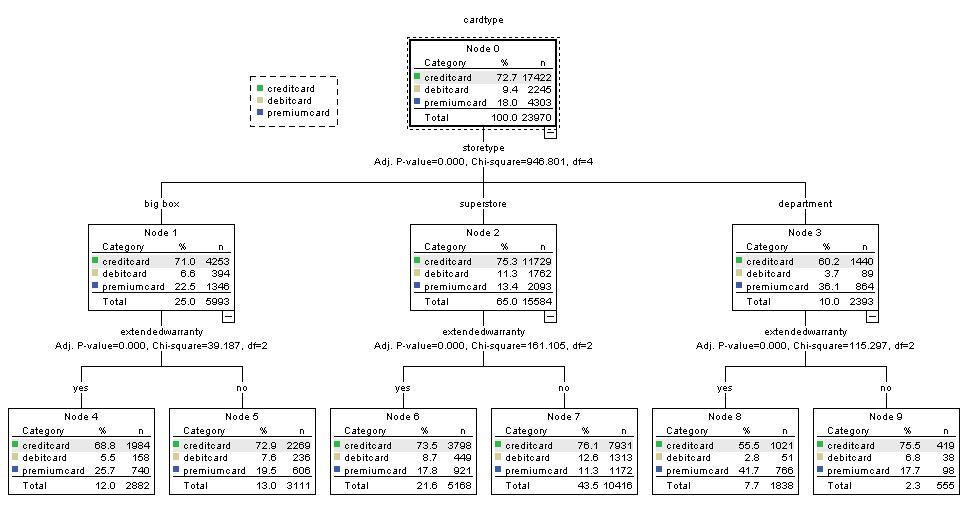
1. **Multi-level clustering trees allow the consideration of more than one explanatory factor when clustering entities. Use the SPSS CHAID decision tree module with the tvfitting2017 dataset to construct classification trees that could be used to explain or predict:**
2. **The purchase price of the TV selected, depending on type of sales outlet, type of payment used, and estimated family income**



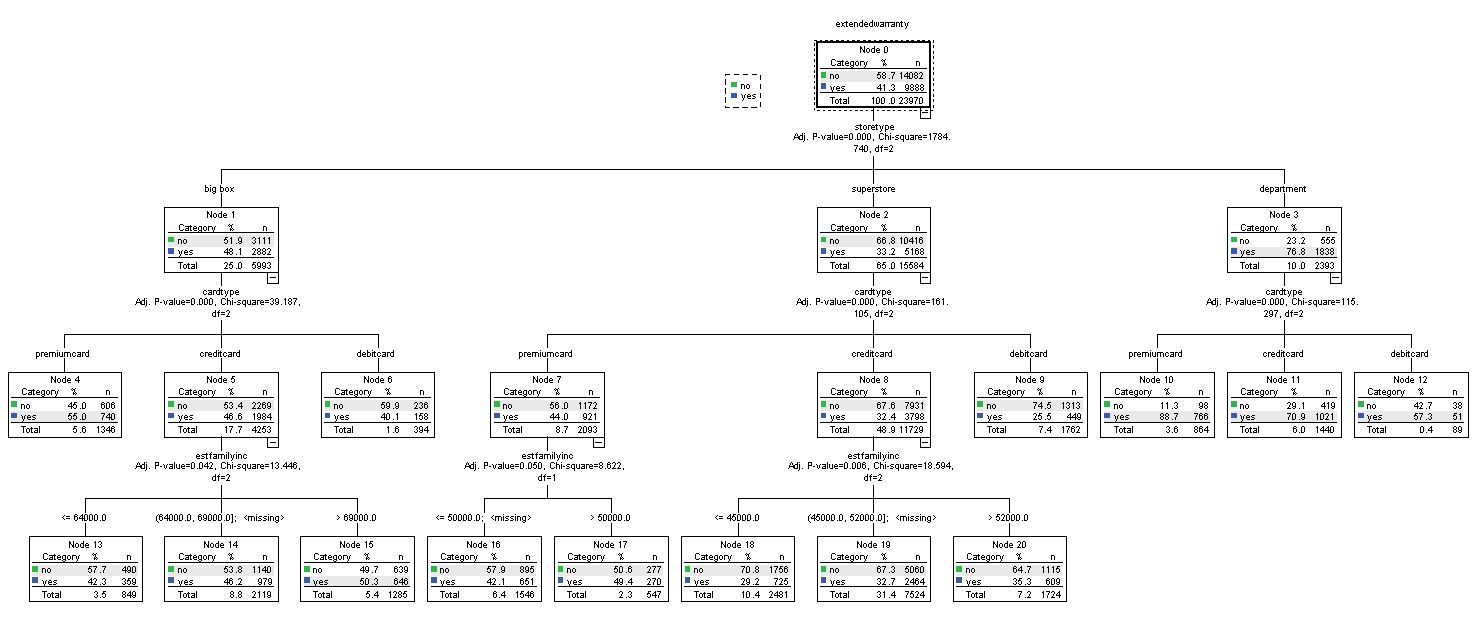
**Zoom in:**



1. **Type of payment used, depending on type of sales outlet, average income and whether an extended warrantee is purchased**



1. **Whether the person will purchase the extended warranty** depending on type of sales outlet, type of payment used, and estimated family income.



1. **For each of the three trees, describe the clusters that result. Indicate the number of clusters and range of cluster sizes. Describe the range in values of the target variables, the clusters with lowest and highest values of the target variables, the discriminating factors, and how the target variables vary systematically according to the discriminating factors.**

**a. The purchase price of the TV selected, depending on type of sales outlet, type of payment used, and estimated family income**

The Average Predicted purchase price of the TV is $503.1, where as the Predicted purchase price of the TV purchased at the bigbox store is $620.05, the Predicted purchase price of the TV purchased at the Superstore is $353.6, and the Predicted purchase price of the TV purchased at the Department store is $1183.6.

If the TV is purchased at the department store using a premium card, the Predicted purchase price of the TV is $1224.

If the TV is purchased at the department store using a credit card card, the Predicted purchase price of the TV is $1163.2.

If the TV is purchased at the department store using a debit card, the Predicted purchase price of the TV is $1122.

If the TV is purchased at the superstore using a premium card, the Predicted purchase price of the TV is $412.

If the TV is purchased at the superstore using a credit card, the Predicted purchase price of the TV is $350.

If the TV is purchased at the superstore using a debit card, the Predicted purchase price of the TV is $308.

If the TV is purchased at the bigbox by a customer whose estimated family income is less than or equal to $64000, using a premium card, the Predicted purchase price of the TV is $618.4.

If the TV is purchased at the bigbox by a customer whose estimated family income is less than or equal to $64000, using a credit card, the Predicted purchase price of the TV is $558.6.

If the TV is purchased at the bigbox by a customer whose estimated family income is less than or equal to $64000, using a debit card, the Predicted purchase price of the TV is $518.

If the TV is purchased at the bigbox by a customer whose estimated family income is less than $69000 or more than $64000, using a premium card, the Predicted purchase price of the TV is $666.3.

If the TV is purchased at the bigbox by a customer whose estimated family income is less than $69000 or more than $64000, using a credit card, the Predicted purchase price of the TV is $604.3.

If the TV is purchased at the bigbox by a customer whose estimated family income is less than $69000 or more than $64000, using a debit card, the Predicted purchase price of the TV is $564.7.

If the TV is purchased at the bigbox by a customer whose estimated family income is more than $64000, using a premium card, the Predicted purchase price of the TV is $711.6

If the TV is purchased at the bigbox by a customer whose estimated family income is more than $64000, using a credit card, the Predicted purchase price of the TV is $649.2.

If the TV is purchased at the bigbox by a customer whose estimated family income is more than $64000, using a debit card, the Predicted purchase price of the TV is $605.4.

**b. Type of payment used, depending on the type of sales outlet, average income and whether an extended warrantee is purchased :**

Among a total of 23970 Customers, 72.7% (17422) of the customers used Credit card, 9.4% (2245) of the customers used Debit card and 18% (4303) of the customers used premium card to purchase TV. Out of which,

* Among a total of 5993 customers, 4253 customers used credit card, 394 customers used Debit card, 1346 customers used Credit card, to purchase TV at the big box store, out of which 2882 customers purchased extended warranty. 1984 customers used credit card, 158 customers used debit card and 740 customers used premium card to purchase the extended warranty.
* Among a total of 15584 customers, 11729 customers used credit card, 1762 customers used debit card and 2093 customers used premium card, to purchase TV at the Superstore, out of which 5168 customers purchased extended warranty. 3798 customers used credit card, 449 customers used debit card and 921 customers used premium card to purchase the extended warranty.
* Among a total of 2393 customers, 1440 customers used credit card, 89 customers used debit card and 864 customers used premium card, to purchase TV at the department store, out of which 1838 customers purchased extended warranty. 1021 customers used credit card, 51 customers used debit card and 766 customers used premium card to purchase the extended warranty.

**c. Whether the person will purchase the extended warranty depending on the type of sales outlet, type of payment used and estimated family income.**

Among a total of 23970 Customers, 41.3% of (9888) the customers purchased the extended warranty and 58.7% (14082) of the customers did not purchase the extended warranty for the TV.

* Out of which, 2882 customers bought the extended warranty and 3111 customers did not buy the extended warranty at the big box store.

Where 740 customers used premium card, 1984 customers used credit card and 158 customers used debit card to purchase the extended warranty.

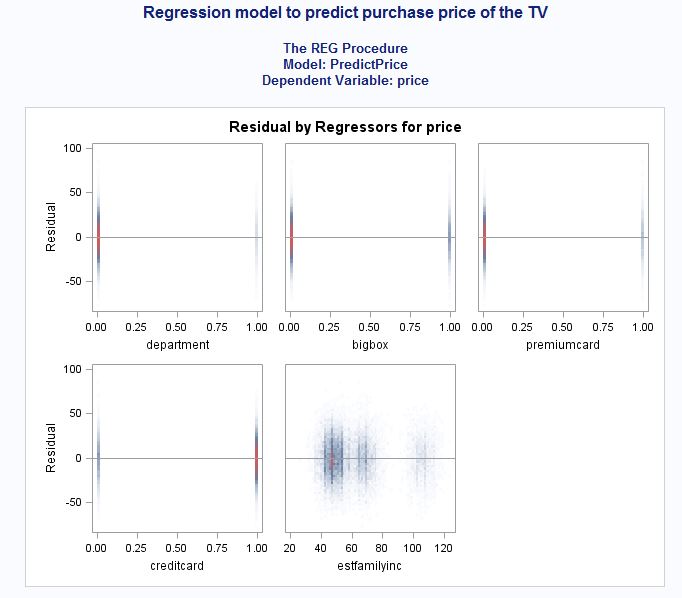
* 5168 customers bought the extended warranty and 10416 customers did not buy the extended warranty at the Superstore.

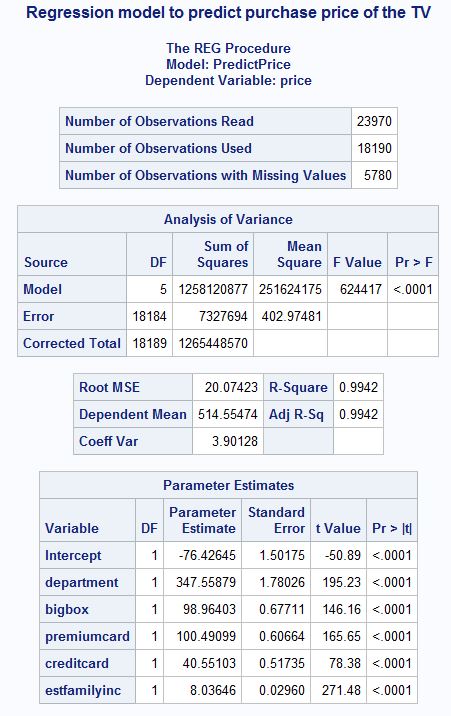
Where 921 customers used premium card, 3798 customers used credit card and 449 customers used debit card to purchase the extended warranty.

* 1838 customers bought the extended warranty and 555 customers did not buy the extended warranty at the department store.

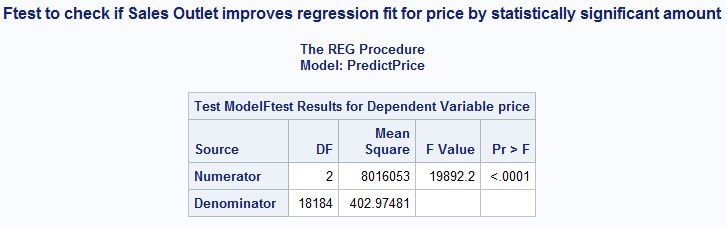
Where 766 customers used premium card, 1021 customers used credit card and 51 customers used debit card to purchase the extended warranty.

**4. Construct a regression model to predict the purchase price of a TV, depending on type of sales outlet, type of payment used, and estimated family income.**





**5. Perform a formal statistical test to gauge whether information about the type of sales outlet improves the regression fit for purchase price by a statistically significant amount after you have accounted for payment type and estimated family income (i.e., whether the better fit when sales outlet is considered could have been a random occurrence). What is the probability that the better fit could have occurred at random?**



The F-Test between the Full Model and the reduced Model shows that the **p-value is less than 0.0001**, which is less than the alpha value 0.05. This means even after considering the effects of the other variables, the variable indicators of Sales Outlets occur systematically making it statistically significant. The Sales Outlet variables must be included in the regression model.

The Probability that the better fit could have occurred at random is very low because the P value < 0.0001, which means that the occurrence is systematic.

**6. Create a logistic model for estimating the likelihood that a person will purchase the extended warranty, depending on type of sales outlet, type of payment used, product price, and estimated family income. Articulate the meaning of the “odds ratio” statistics, which are printed for the explanatory variables. Be careful to indicate the two probabilities involved. Interpret the estimated “odds-ratio” impact for the “department store” indicator variable.**

Odds Ratio:

For every $1 increase in the TV’s purchase price, the odds ratio changes by 1.002.

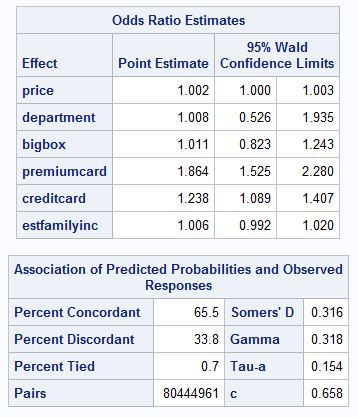
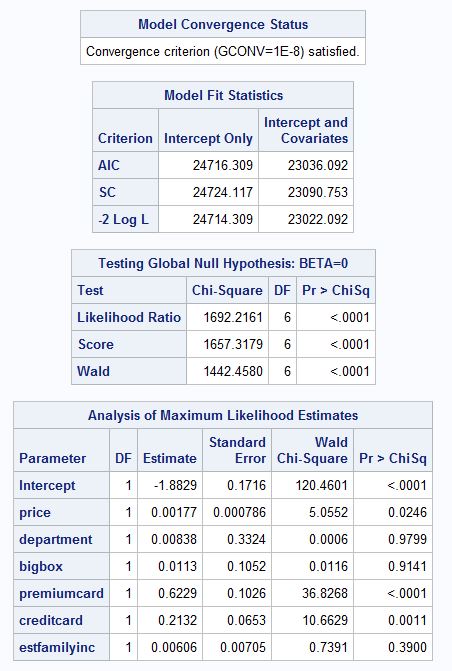
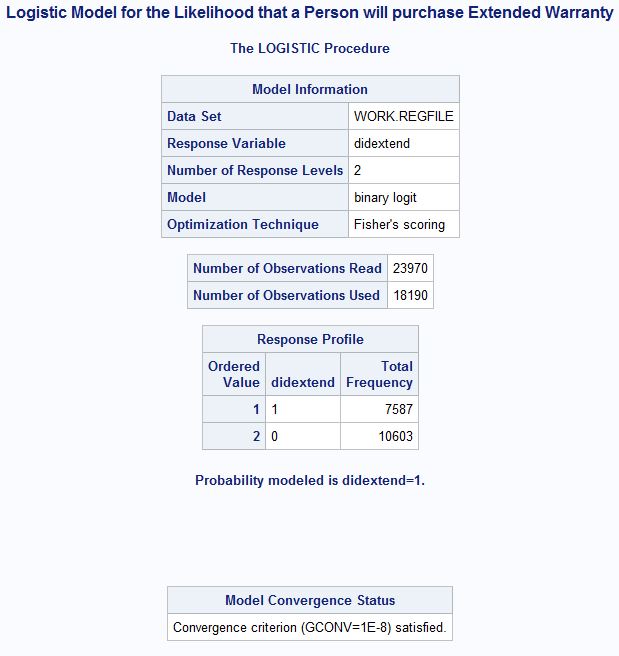
The factor by which the odds ratio changes for every TV purchase made at the Department Store is 1.008.

The factor by which the odds ratio changes for every TV purchase made at the Big Box Store is 1.011.

The factor by which the odds ratio changes for every TV purchase made using a premium card is 1.864.

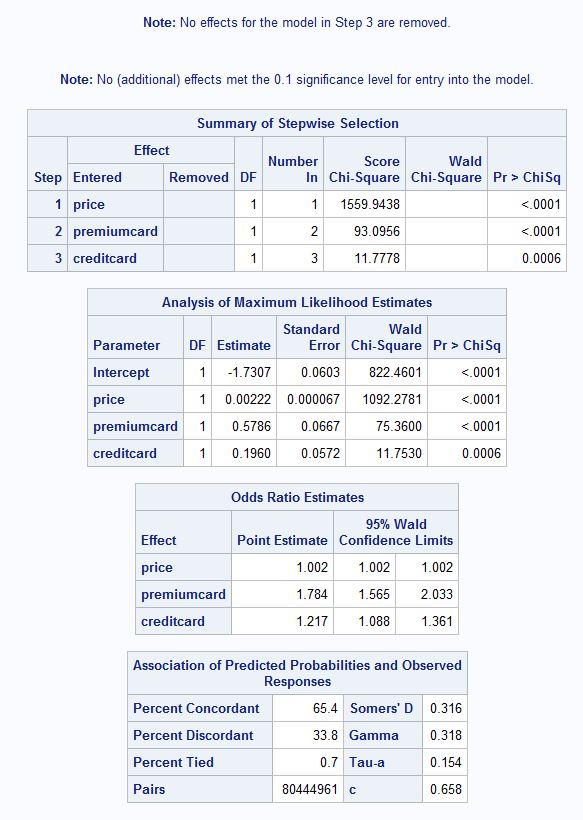
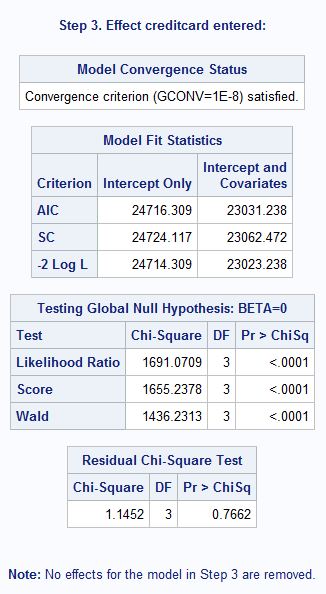
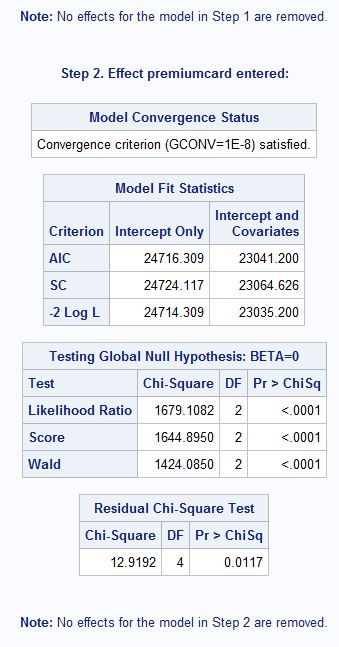
The factor by which the odds ratio changes for every TV purchase made using a credit card is 1.238.

For every $1000 increase in the estimated family income of a customer, the odds ratio changes by 1.006.

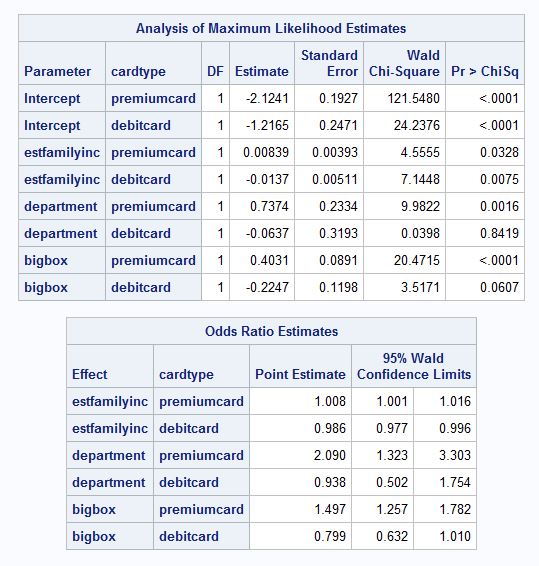
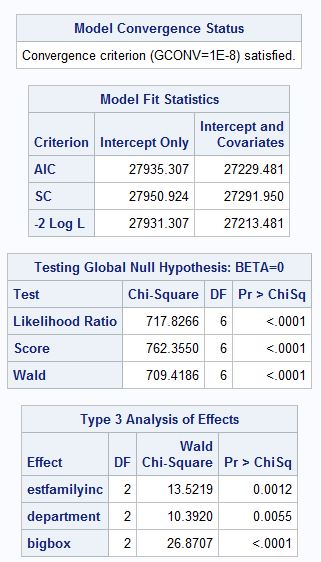
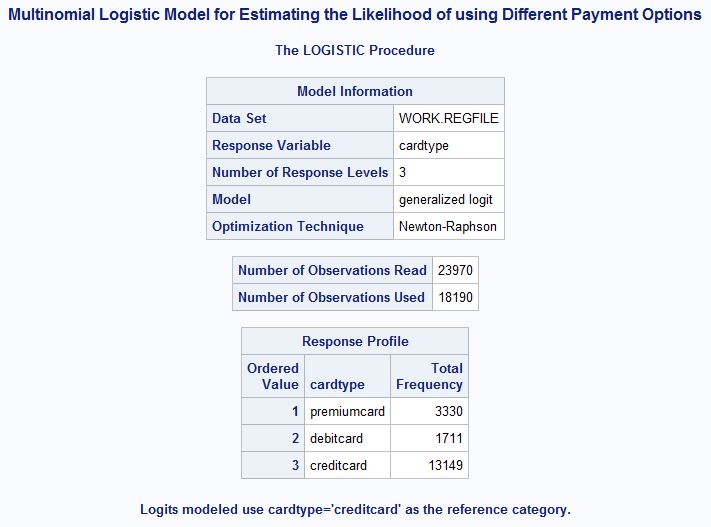


**7. Create a parsimonious logistic model for estimating the likelihood that a person will purchase the extended warranty, considering the type of sales outlet, type of payment used, product price and estimated family income, but using the stepwise option for selecting variables, with a .1 level of statistical significance required for adding or removing variables (slentry=.1 slstay=.1). Using this model, how does the likelihood of purchasing the extended warranty depend on whether the purchase is made at a department store?**

**Considering the model that is generated, the value for the purchase of extended warranty at any given store is missing which means that the likelihood of purchasing the extended warranty does not depend on whether the purchase is made at the department store or any other store.**



**8. Create a multinomial logistic model for estimating the likelihoods of using the different payment options depending on the type of sales outlet and estimated family income (in $1,000). Use the “descending” option in the proc logistic to facilitate interpretation of the logistic parameters. That will make the probability that the transaction occurred at a big box store appear as the numerator of the first odds ratio. Considering the model parameters (and associated impact on odds ratios), how do the probabilities for the alternative outcomes relate to type of retail outlet and estimated family income?**



Odds Ratio – Estimated Family Income:

If a customer uses the Premium Card for Payments, then the odds ratio changes by 1.008 times of the estimated Family Income.

If a customer uses the Debit Card for Payments, then the odds ratio changes by 0.986 times of the estimated Family Income.

Odds Ratio – Store Type:

If a customer uses the Premium Card for Payments at the Department Store, then the odds ratio changes by 2.090.

If a customer uses the Debit Card for Payments at the Department Store, then the odds ratio changes by 0.938.

If a customer uses the Premium Card for Payments at the Big Box Store, then the odds ratio changes by 1.497.

If a customer uses the Debit Card for Payments at the Department Store, then the odds ratio changes by 0.799.

**9. Apply the models developed in questions (3), (5), and (7) to Customer 000028 in the holdout sample, what are your estimates of selling price, likelihood of the customer’s choosing the extended warranty and likelihoods of the alternative payment types? How do the CHAID-tree estimates for the target variables compare with the linear regression and logistic estimates for the same case?**

**Customer 000028:**

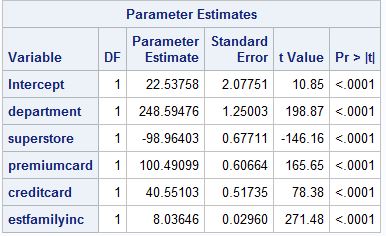
**According to the Regression Model, the estimates of selling price of the TV is :**

Y = 22.53758 + (0) \* 248.59476 – (1) \* 98.96403 + (0) \* 100.49099 + (1) \* 40.55103 + (56) \* 8.03646 = 414.16634

**Y = selling price = $414.2**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| customer | cardtype | estfamilyinc | storetype | price | ext-war | creditcard | debitcard | premiumcard | bigbox | department | superstore |
| 28 | creditcard | 56 | superstore | 411.86 | yes | 1 | 0 | 0 | 0 | 0 | 1 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| didextend | logit | probchooseew | didchooseew | didnotchooseew |
| 1 | -0.7933363 | 0.31145275 | 1 | 0 |

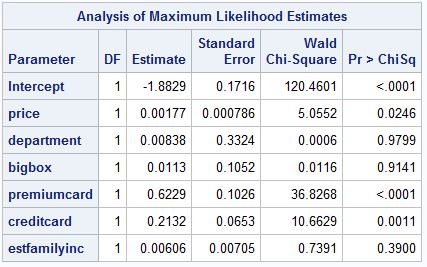


**Customer 000028:**

**According to the Logistic Model, the likelihood that the customer will purchase the Extended Warranty for the TV is : 87.8%**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter | DF | Estimate | Standard | Wald | Pr>c |
|  |  |  | Error | Chi-Sq |  |
| Intercept | 1 | -1.882923053 | 0.2218 | 71.197 | <0.0001 |
| Department | 1 | 0.008384902 | 0.1052 | 0.0116 | 0.9141 |
| SuperStore | 1 | -0.0113 | 0.2369 | 0.0002 | 0.99 |
| Premium card | 1 | 0.622892424 | 0.1026 | 36.827 | <0.0001 |
| Credit card | 1 | 0.213243584 | 0.0653 | 10.663 | 0.0011 |
| Price | 1 | 0.001767789 | 0.000786 | 5.0552 | 0.0246 |
| Estfamilyinc | 1 | 0.006060862 | 7.05E-06 | 0.7391 | 0.39 |

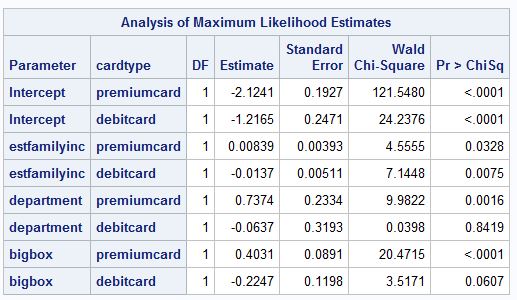
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Intercept | SuperStore | Department | Premiumcard | Creditcard | Price | Estfamilyinc | Ext Warranty | Logit | Cum Prob |
|  | -1.8716 | -0.0113 | -0.00296 | 0.6229 | 0.2132 | 0.0018 | 6.06E-06 |  | -0.6012918 | 0.354048206 |
| Cust28 | 1 | 1 | 0 | 0 | 1 | 411.86 | 56000 | yes = 1 | 0.3070216 | **0.878087704** |



**According to the Multinomial Regression Model, The likelihood of using different payment options by Customer 000028 in the Superstore is:**

**14.4% for a premium card, 10.4% for a Debit card, 75% for a credit card.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | intercept | bigbox store | Superstore | estfamilyinc | logit 1 | Probability | Card Type |
| Cust 000028 | Premium card | Logit 1 | 1 | 0 | 1 | 56000 | -1.653946067 | 0.14392371 | Premium |
|  | Debit Card | Logit 2 | 1 | 0 | 1 | 56000 | -1.98167528 | 0.10370539 | Debit card |
|  |  |  |  |  |  |  |  | 0.7523709 | Credit card |



**The CHAID-tree estimates for the target variables compare with the linear regression and logistic estimates as follows:**

According to Node 27, Node 34, and Node 41 in the Chaid Tree,

The probability that the customer000028 uses a premium card for payments in a superstore is 1%,

The probability that the customer000028 uses a credit card for payments in a superstore is 5.3%,

The probability that the customer000028 uses a debit card for payments in a superstore is 0.9%.

**10. What happens if you try to apply the respective models and the CHAID trees to Customer 000042 on the testing dataset?**

Customer 000042:

According to the Regression Model, the estimates of selling price of the TV is :

Y = 22.53758 + (1) \* 248.59476 – (0) \* 98.96403 + (0) \* 100.49099 + (1) \* 40.55103 + (63.5) \* 8.03646 = 821.99858

**Y = selling price = $822.**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| customer | cardtype | estfamilyinc | storetype | price | ext-war | creditcard | debitcard | premiumcard | bigbox | department | superstore |
| 42 | creditcard | 63.5 | department | 1112.08 | yes | 1 | 0 | 0 | 0 | 1 | 0 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| didextend | logit | probchooseew | didchooseew | didnotchooseew |
| 1 | -0.0573414 | 0.485668577 | 1 | 0 |

**According to the Multinomial Regression Model, The likelihood of using different payment options by Customer 000042 is:**

**16.35% for a premium card, 18.2% for a Debit card, 65% for a credit card.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | intercept | bigbox store | Superstore | estfamilyinc | logit 1 | Probability | Card Type |
| Cust 000042 | Premium card | Logit 1 | 1 | 0 | 0 | 0 | -1.386708289 | 0.16355679 | Premium |
|  | Debit Card | Logit 2 | 1 | 0 | 0 | 0 | -1.280163201 | 0.18194517 | Debit Card |
|  |  |  |  |  |  |  |  | 0.65449804 | Credit Card |

**11. Apply the logistic model from question (5) to each case in the holdout sample (tvtesting2017). Sort the holdout sample from highest to lowest probability of purchasing the extended warranty. Produce lift statistics and discuss the discriminating power of the model in terms of the “lift” it provides in identifying the consumers who are most likely to purchase the extended warranty.**

The LIFT statistics above was calculated with a group size of 500. The cumulative percentage (actual) is 41.365% which explains that if we randomly select a customer from the entire group, the estimated probability of purchasing the extended warranty is 41.365%.

The LIFT is the ratio of probability of the number of customers who actually purchased the extended warranty relatively to the overall number of customers in the group. The top bucket with 500 members shows that the group lift is 100% which means that, it is this group of costumers who are most likely to purchase the extended warranty.

According to this model, we predicted that 234 customers out of the top 500 customers (46.8%) would purchase the extended warranty. But 367 (73.4%) customers out of top 500 customers purchased the extended warranty.

