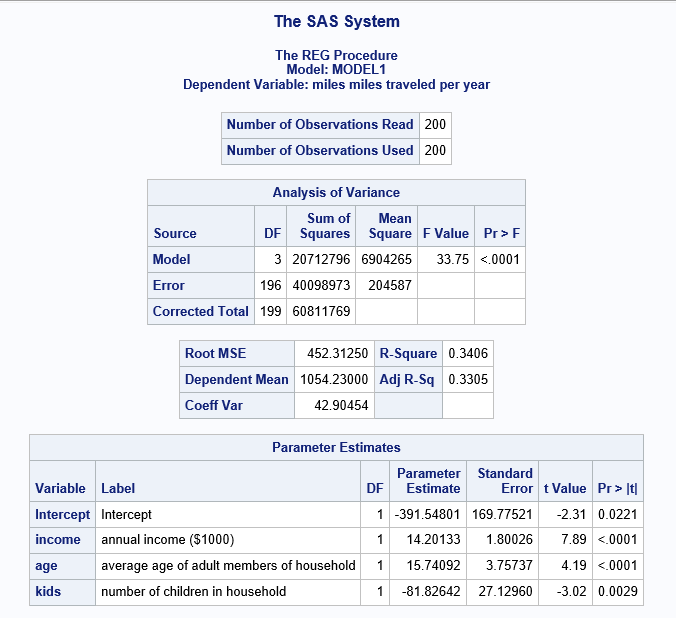
1. **Use the data in vacation.dat and run a regression with Miles (miles travelled) as the dependent variable and income, age (average age of adult members), and kids as the independent variables.**
2. Run a regression model and interpret the coefficients. Comment on the model fit.



Model fit:

The R-Square value is 0.3406, which means that the variables income, age and kids explain 34.06% of the total variation in Miles travelled. The Adjusted R-square value of 0.3305 doesn’t have a large difference form the R-Square value, which infers that the model is valid.

Interpreting the coefficients:

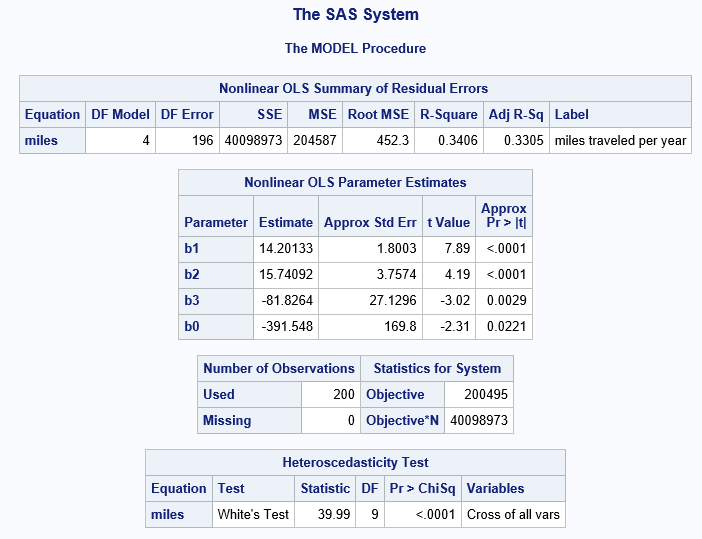
All the variables are significant at 95% confidence level as inferred from the P-value of each variable is less than 0.05.

The coefficient of the ‘income’ variable is 14.20133, which means that an increase of $1000 in annual income results in the increase of 14.2 miles travelled.

The coefficient of the ‘age’ variable is 15.74092, which means that a year increase in the average age of the adult members of household results in the increase of 15.74 miles travelled.

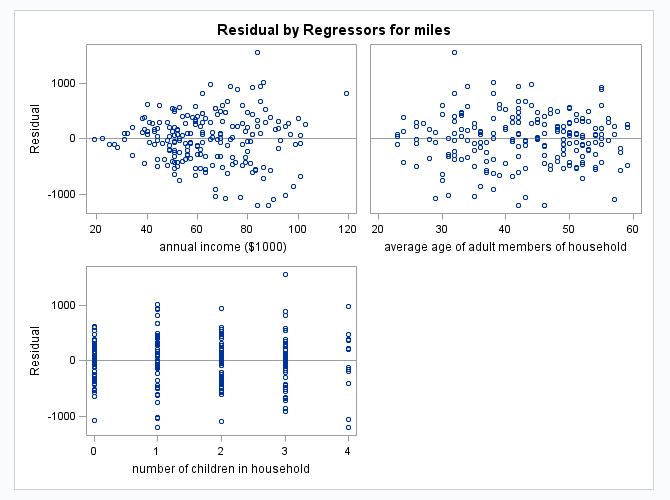
The coefficient of the ‘kids’ variable is -81.82642, which means that with an increase of 1 child per household will in the reduce the number of miles travelled by 81.82 miles.

1. Check whether there is heteroscedasticity in the model using White test.



Considering the results from the Heteroscedasticity Test table, the P-value of the White test statistic is <0.0001 implies that the null hypothesis of equality in the variances of the errors is rejected. This means that there is heteroscedasticity in the model.

1. Run a weighted Least squares (WLS) regression. Discuss your results in a paragraph. (Comment on model fit, significance of coefficients, and the effect of doing WLS.)



From the residual graphs we can observe that the variation in the residuals of the annual income variable is increasing with the increase in the annual income.

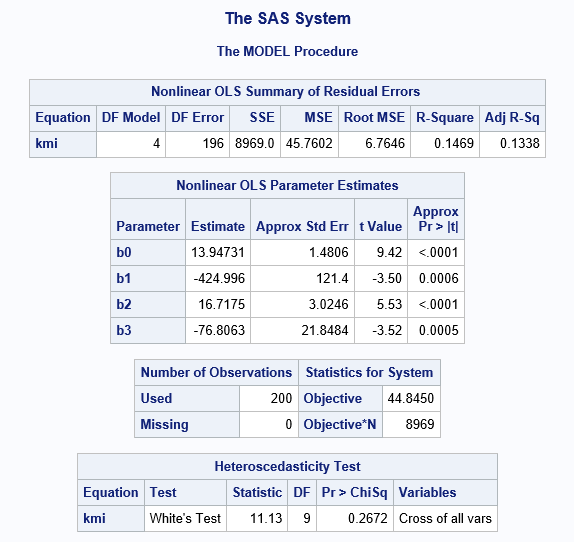
In order to reduce the heteroscedasticity of the model we ran a WLS regression by considering the following variables.

kmi=miles/income

kage=age/income

kkids=kids/income

kin=1/income



Form the Heteroscedasticity test result we can infer that the P-value of the White test statistic is 0.2672, which is more than 0.05, this implies that the null hypothesis is rejected and the variance of the errors are equal, indicating that the heteroscedasticity in the model is eliminated.

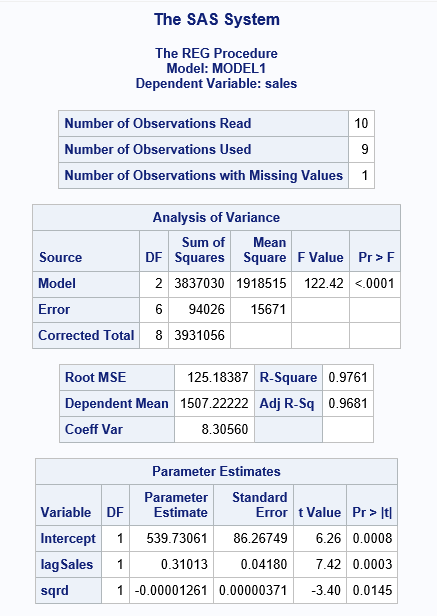
We can infer from the R-square value of 0.1469 and Adjusted R-square value of 0.1338 that the model fit is reduced after the WLS.

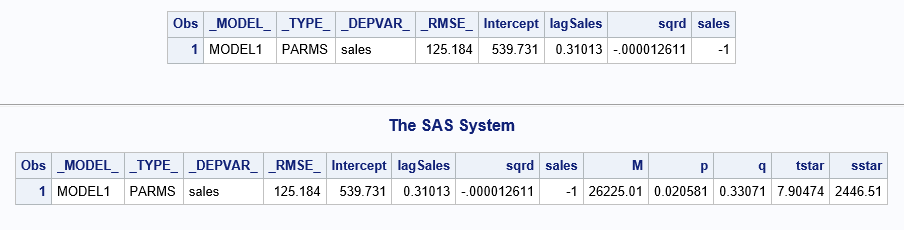
The P-value of all the variables is less than 0.05 implying that all the coefficients are significant at 95% confidence level.

By performing WLS, the effect of heteroscedasticity is reduced at the expense of model fit.

**2. Bass Model.**

1. Using SAS and regression, estimate the Bass model. Save the regression parameters using option OUTEST. Find p, q, and M and compute peak sales and the time when that peak will occur.





From the regression parameters we calculated the following:

Coefficient of innovation, p: the percent of people who buy the product on their own = 2.058%

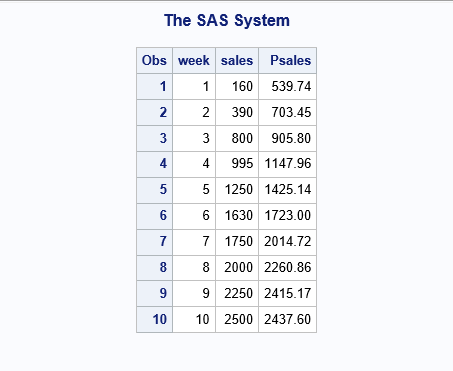
Coefficient of imitation, q: the percent of people who learn through word of mouth and buy = 33.07%

Total market potential, M: Maximum number of people that will ever buy the product = 26,225.

Peak sales, sstar = 2446 units

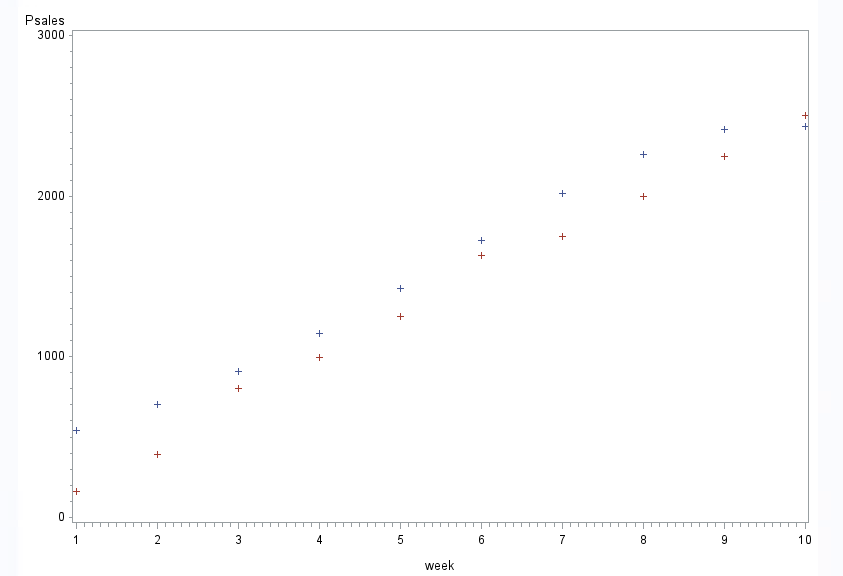
Time to reach peak sales, tstar = 7.90474 weeks = 8 weeks (approx.)

1. Predict sales in each period using only the model parameters p, q, and M and the fact that sales at time period 0=0 .



The above table shows the actual sales and predicted sales, Psales for the respective week.

1. Plot a graph of actual versus predicted sales. (SAS code given to you in the slides)

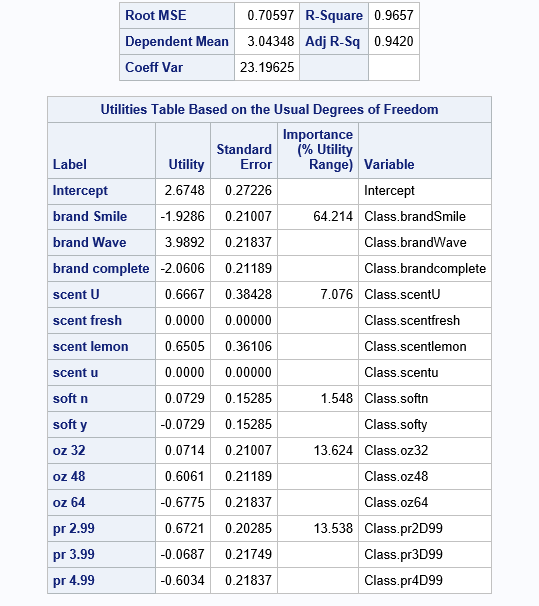


The blue points in the plot are the Predicted sales and the red points are the actual sales.

**3. A conjoint study was undertaken by a detergent manufacturer. The attributes that were considered were**

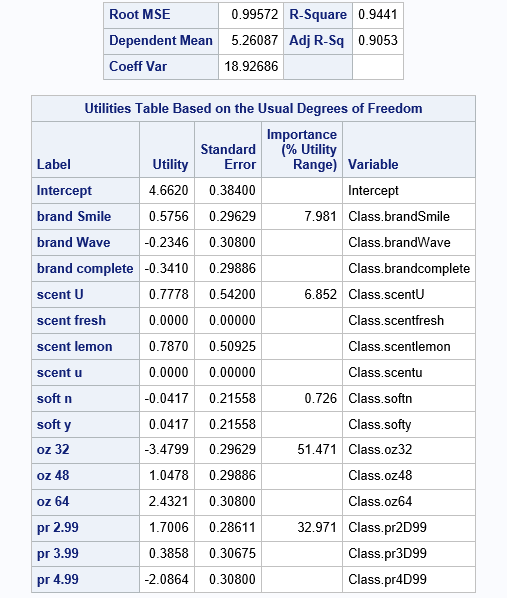
1. Find the importance weights and part-worths for each respondent using PROC TRANSREG.

The following are the importance weights and part-worths for each respondent.



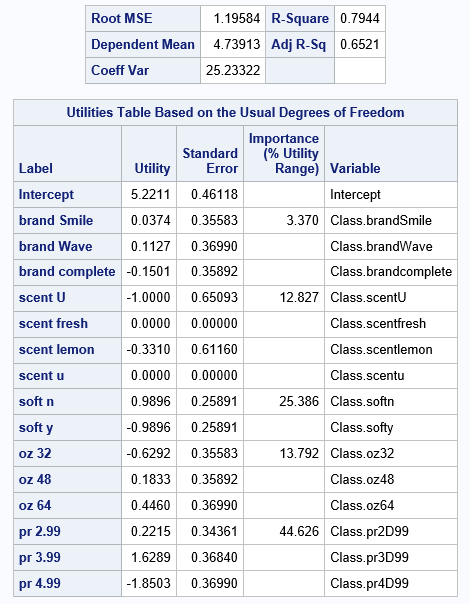
For respondent – 1:

The brand is of highest importance followed by quantity, price, scent and softness.



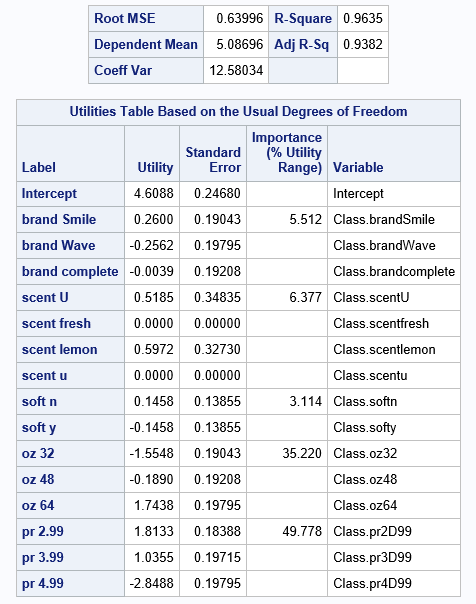
For respondent – 2:

The quantity is of highest importance followed by price, brand, scent and softness.



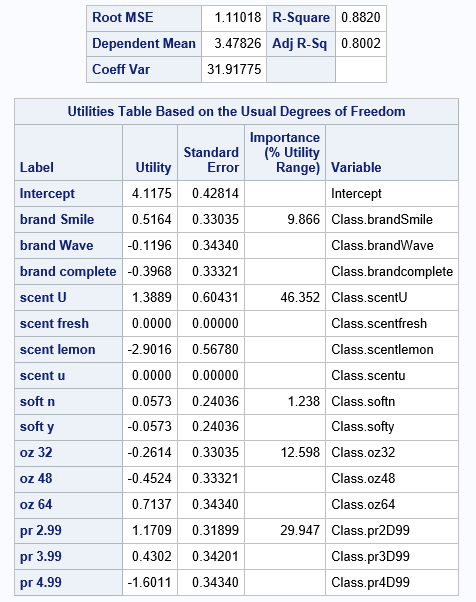
For respondent – 3:

The price is of highest importance followed by softness, quantity, scent and brand.



For respondent – 4:

The price is of highest importance followed by quantity, scent, brand and softness.



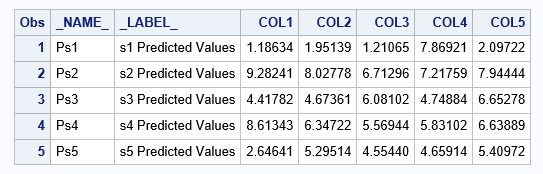
For respondent – 5:

The scent is of highest importance followed by price, quantity, brand and softness.

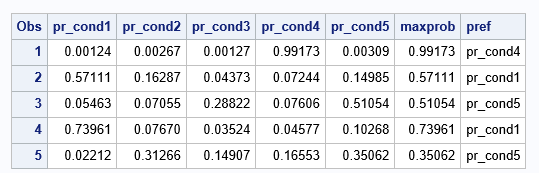
1. Predict the choice (using logit rule) for each respondent (s1-s5) for each of the following combinations using your estimates in question 1 above.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Condition -1 | complete | lemon | y | 64 | 2.99 |  |  |  |  |  |
| Condition -2 | Smile | fresh | y | 48 | 2.99 |  |  |  |  |  |
| Condition -3 | Smile | u | y | 48 | 3.99 |  |  |  |  |  |
| Condition -4 | Wave | u | y | 48 | 2.99 |  |  |  |  |  |
| Condition -5 | Smile | u | n | 48 | 2.99 |  |  |  |  |  |

Each product is named as a condition.



The combined utility values for each condition named as COL# and for each respondent mentioned as Ps#.



Applying the logit rule, the probabilities of each condition for each respondent are computed and the preferred choice of each respondent is mentioned in the ‘pref’ column.

The respondent-1 will choose the product-4.

The respondent-2 will choose the product-1.

The respondent-3 will choose the product-5.

The respondent-4 will choose the product-1.

The respondent-5 will choose the product-5.