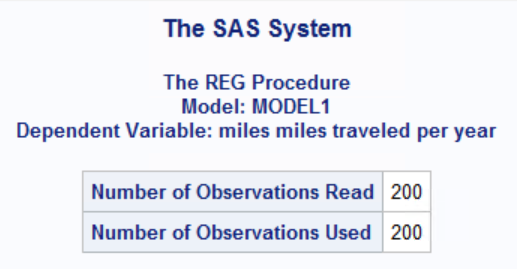
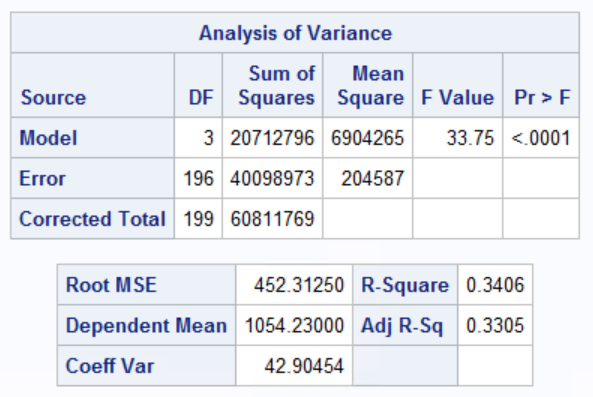
**SAS Homework 3**

**Group assignment**

**Group -10**

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.

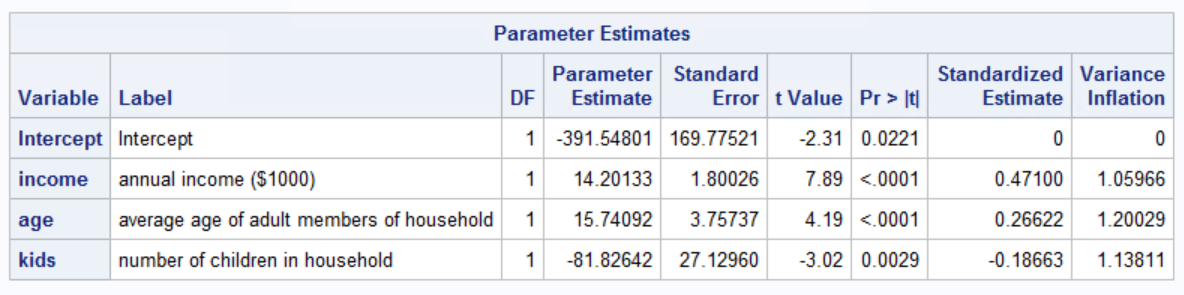
 

Above regression is done with **Null hypothesis**: Coefficient of income, age and kids is zero; and **Alternate hyposthesis**: Atleast one of the coefficient is not zero.

As the p-value (0.001) is less than alpha (0.05), so reject the null, that is the coefficient of independent variables is not zero. Also the p-value of all the coefficient less than 0.05%, hence all are significant.

For the above resultset, the R-squared is 0.34 and ajusted R-squared is 0.33, that means the given 3 independent variables are able to explain only 33% of variation in dependent variable (miles).

**Interpretation of coffecient:**



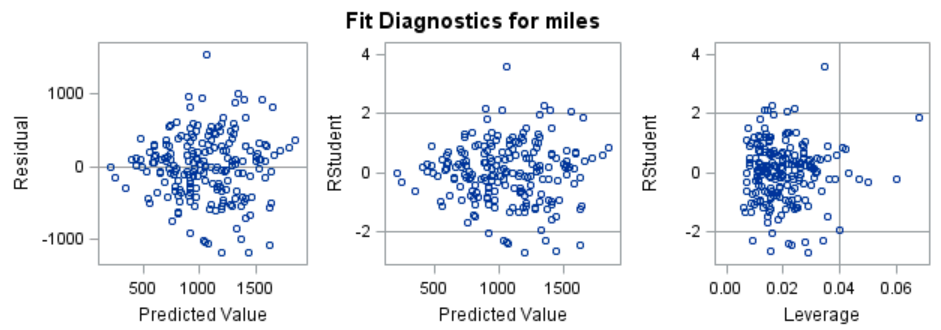
Miles = -391.548 + 12.20 income + 15.74 age - 81.826 kids

**Income**: For a unit increase ($1000) in income, miles travelled will increase by 14.20 units. So people will travel more with increase in income.

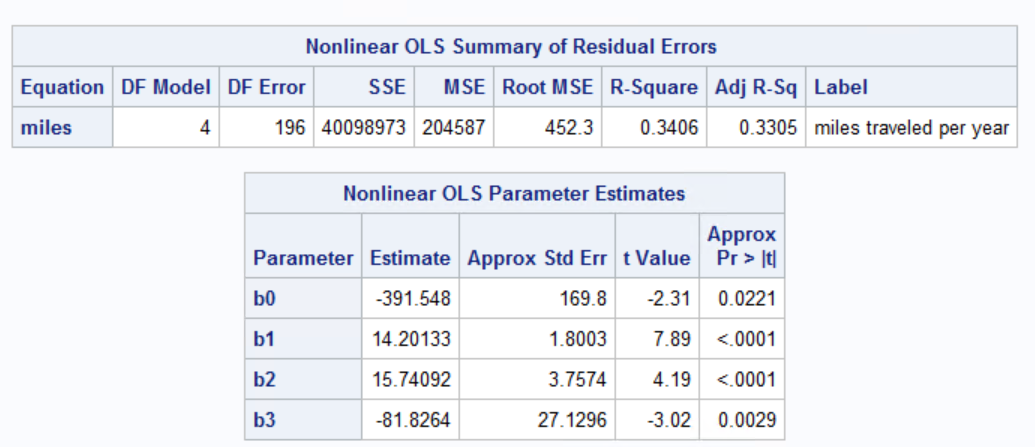
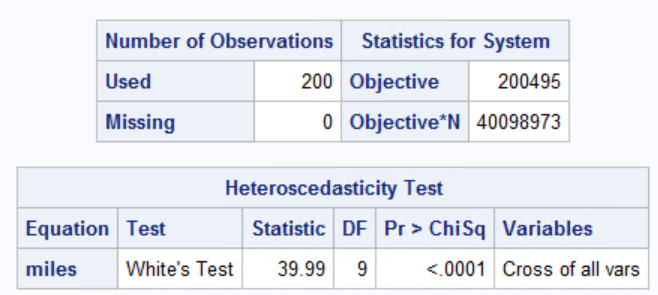
**Age**: For a unit increase (average age of adult members) in age, miles travelled will increase by 15.74 units. So people as they grow old will travel more.

**Kids** : With a unit increase in number of kids, the miles travelled will decrease by 81.82 units. More children tend to reduce the number of miles travelled in a family.

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



From the above plot of residual versus predicted value, we can see that the variance in error is not constant. To test further:

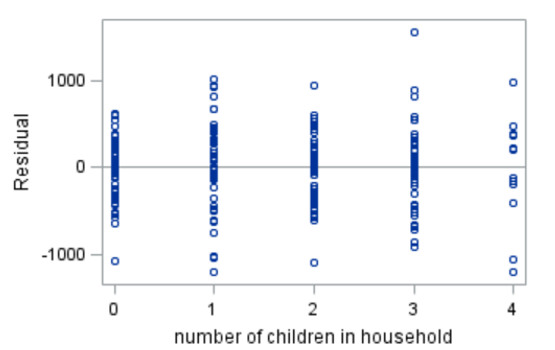
White test is used to test whether the errors are homoscedastic or not, in other words whether the error has constant variance.

**Null Hypothesis:** Error are homoscedastic

**Alternate Hypothesis**: Errors are heteroskedastic.

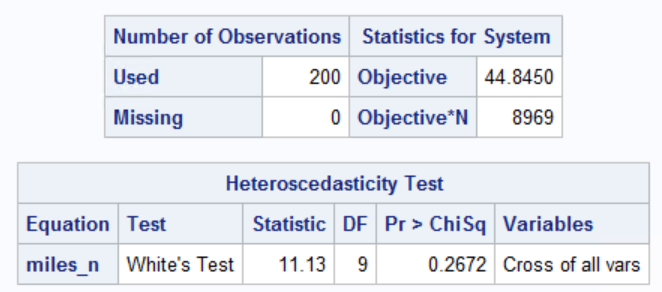
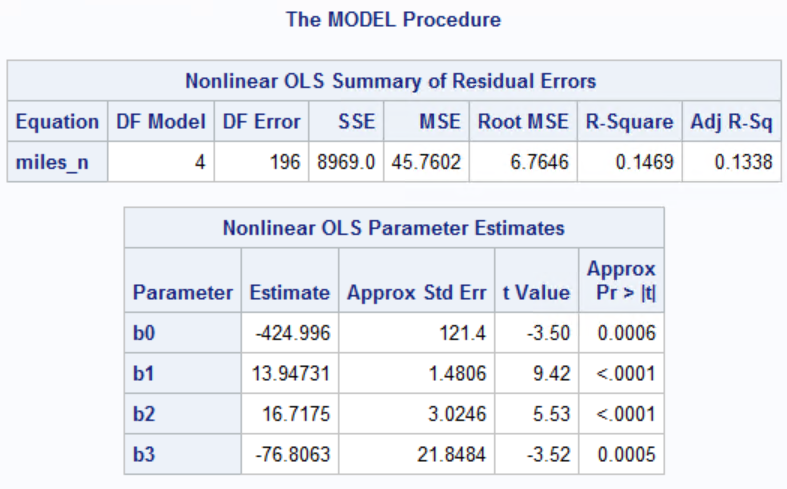
As the p value (0.001) in white test is less than alpha (0.05), we reject the null that is errors are not homoscedastic. There is heteroskedasticity in our model.

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.)



As per the distribution of error with independent variables, we can see that there is heteroskedasticity in the distribution of income variable as the error is showing an increasing pattern with the increase in income.

On running the regression equation with inverse of income variable as weight in Proc Model:



White test with weighted least square is used to test whether the errors are homoscedastic or not, in other words whether the error has constant variance after weighting the independent variables.

**Null Hypothesis:** Error are homoscedastic

**Alternate Hypothesis**: Errors are heteroskedastic.

As the p value (0.2672) in white test is greater than alpha (0.05), we do not reject the null, hence we can statistically prove that is errors are homoscedastic. There is no heteroskedasticity in our model.

**Model fit**: As we can see from above result sets, the R-squared and adjusted R-square is decreased in WLS model from 33% to 13.38%.

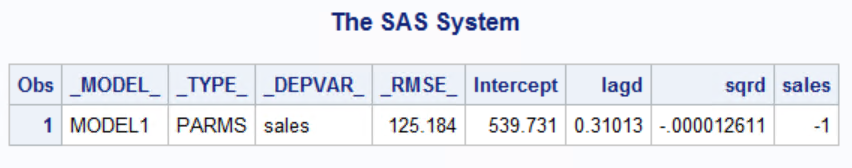
**Significance of coefficients**: As the p-value of the coefficients is less than alpha (0.05), we can statistically prove that all the three independent variables are significant.

**Effect of doing WLS**: WLS is used to improve the heteroskedasticity of the model, but it has negatively affected the model fit.

1. I have provided the Sales of a durable good.

|  |  |
| --- | --- |
| Week | Sales |
| 1 | 160 |
| 2 | 390 |
| 3 | 800 |
| 4 | 995 |
| 5 | 1250 |
| 6 | 1630 |
| 7 | 1750 |
| 8 | 2000 |
| 9 | 2250 |
| 10 | 2500 |

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.





p (**coefficient of innovation):** From the above model we can see that 2 percent of people will buy the product on their own.

q (**coefficient of imitation):** 33% percent of people who hear about a product from others will buy it.

M (**total market potential)**: The maximum number of people that will ever buy this product is 26225.01

Time to reach peak sales is 7.90 units.

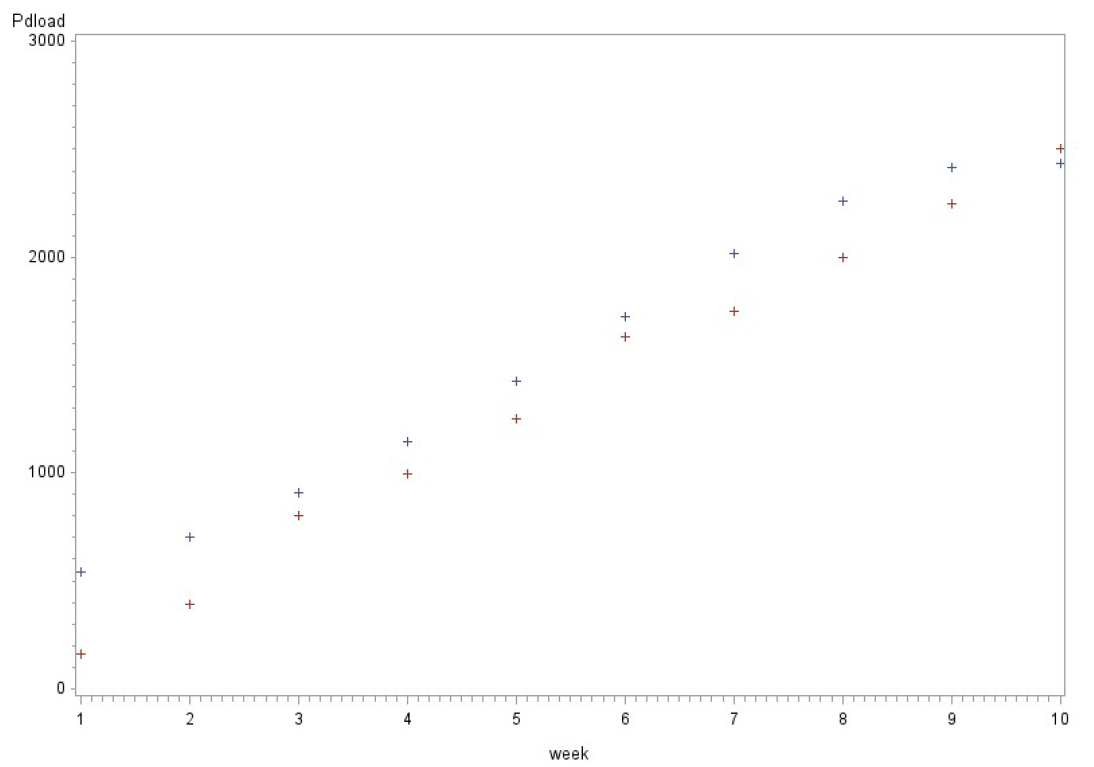
Peak sales is 2446.51 units.

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 predicted value of sales is:



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



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

Brand (Complete, Smile, Wave)

Scent (fresh, lemon, Unscented)

Whether there was a softener or not (Y, N)

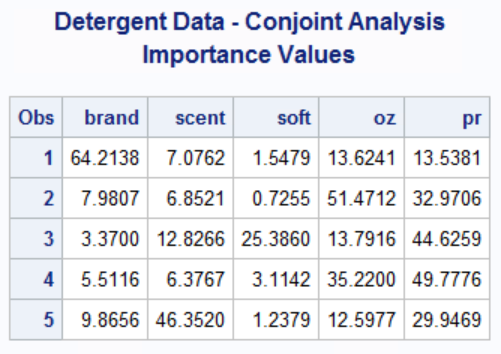
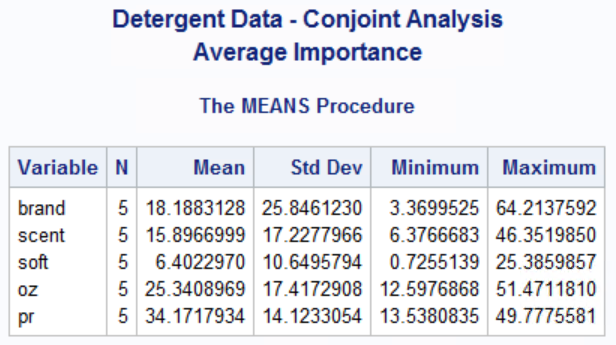
Size of packet (32, 48, 64)

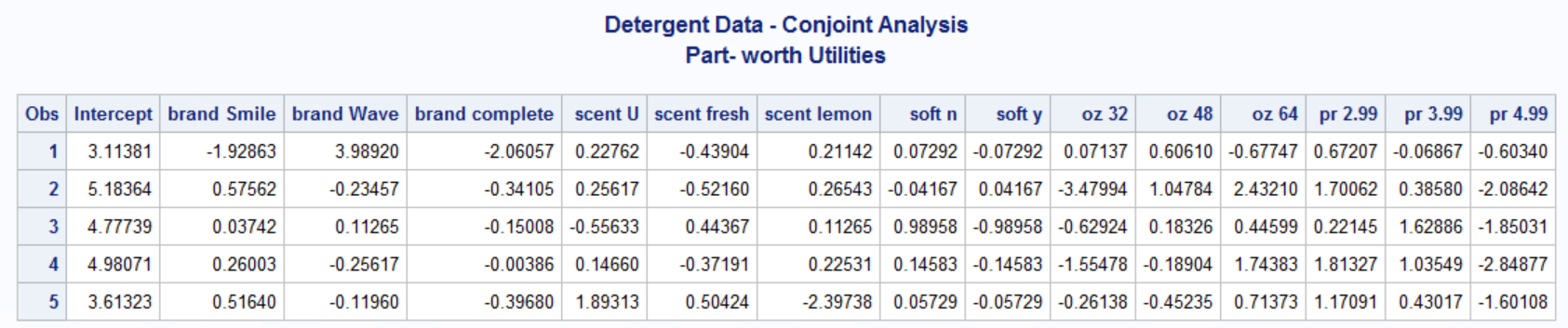
Price (2.99, 3.99, 4.99)

The preferences of five respondents s1, s2, s3, s4, s5 were obtained for some combination of attributes on a 1-9 point scale with 9 indicating a higher preference.

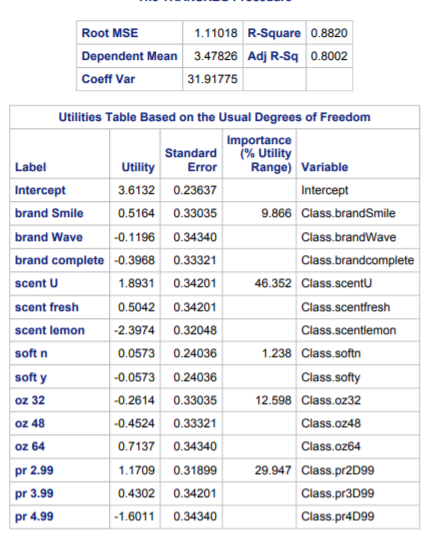
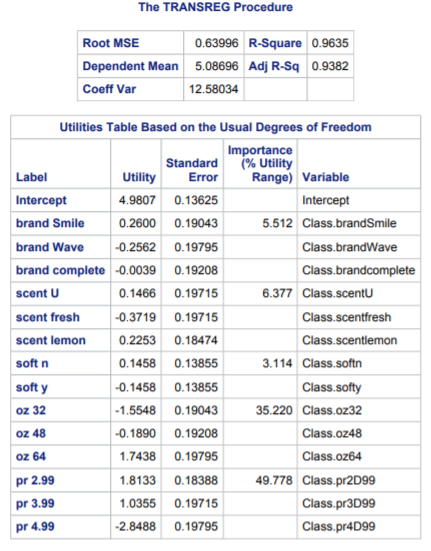
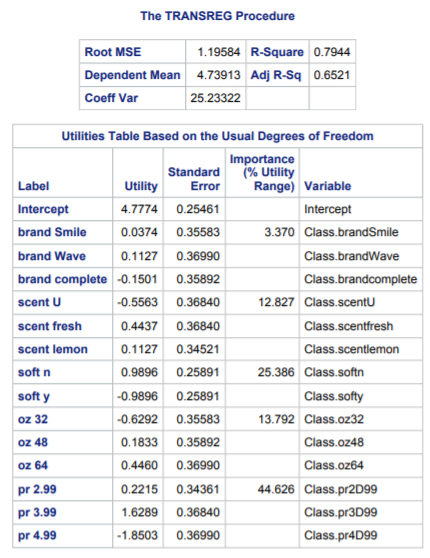
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| brand | scent | soft | oz | pr | s1 | s2 | s3 | s4 | s5 |
|  |  |  |  |  |  |  |  |  |  |
| complete | fresh | n | 48 | 4.99 | 1 | 3 | 3 | 2 | 2 |
| complete | fresh | y | 32 | 2.99 | 1 | 3 | 3 | 5 | 5 |
| complete | lemon | n | 32 | 2.99 | 1 | 2 | 7 | 5 | 1 |
| complete | lemon | y | 64 | 3.99 | 1 | 9 | 5 | 8 | 1 |
| complete | U | n | 64 | 3.99 | 1 | 9 | 7 | 8 | 7 |
| complete | U | y | 48 | 4.99 | 1 | 3 | 3 | 2 | 3 |
| Smile | fresh | n | 64 | 2.99 | 1 | 9 | 9 | 9 | 6 |
| Smile | fresh | y | 48 | 3.99 | 1 | 7 | 7 | 6 | 5 |
| Smile | lemon | n | 48 | 3.99 | 1 | 7 | 7 | 6 | 1 |
| Smile | lemon | y | 32 | 4.99 | 1 | 1 | 1 | 1 | 1 |
| Smile | U | n | 32 | 4.99 | 1 | 1 | 3 | 1 | 2 |
| Smile | U | y | 64 | 2.99 | 1 | 9 | 3 | 9 | 9 |
| Wave | fresh | n | 32 | 3.99 | 7 | 1 | 7 | 4 | 5 |
| Wave | fresh | y | 64 | 4.99 | 5 | 5 | 3 | 3 | 2 |
| Wave | lemon | n | 64 | 4.99 | 5 | 5 | 5 | 3 | 1 |
| Wave | lemon | y | 48 | 2.99 | 9 | 9 | 5 | 7 | 1 |
| Wave | U | n | 48 | 2.99 | 9 | 9 | 5 | 7 | 7 |
| Wave | U | y | 32 | 3.99 | 7 | 1 | 5 | 4 | 5 |
| Wave | lemon | n | 64 | 2.99 | 8 | 9 | 6 | 9 | 3 |
| Smile | lemon | n | 32 | 4.99 | 2 | 1 | 3 | 2 | 1 |
| Smile | fresh | y | 48 | 2.99 | 2 | 8 | 4 | 5 | 5 |
| complete | U | y | 32 | 2.99 | 2 | 4 | 2 | 5 | 6 |
| complete | lemon | y | 48 | 3.99 | 2 | 6 | 6 | 6 | 1 |

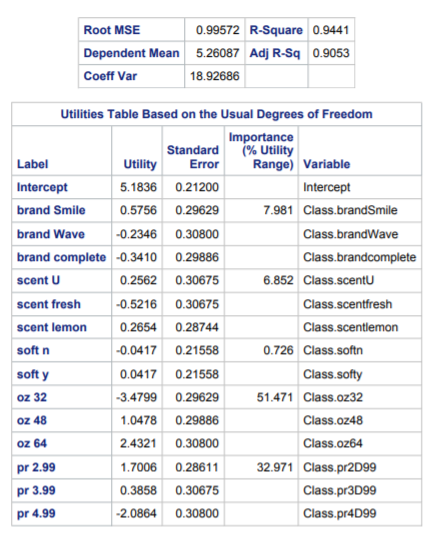
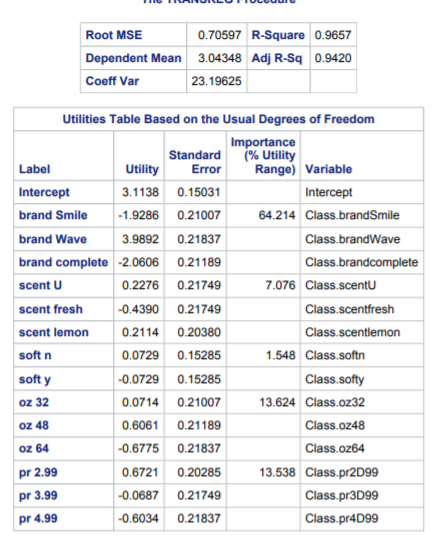
Find the importance weights and part-worths for each respondent using PROC TRANSREG.



Based on the importance values, price is the most important attribute for some of the respondents, but brand and size of packet is most important for others. On the average, price is most important followed closely by size of packet. Soft and scent are less important. Both Wave and Complete are the most preferred brands by some of the respondents.

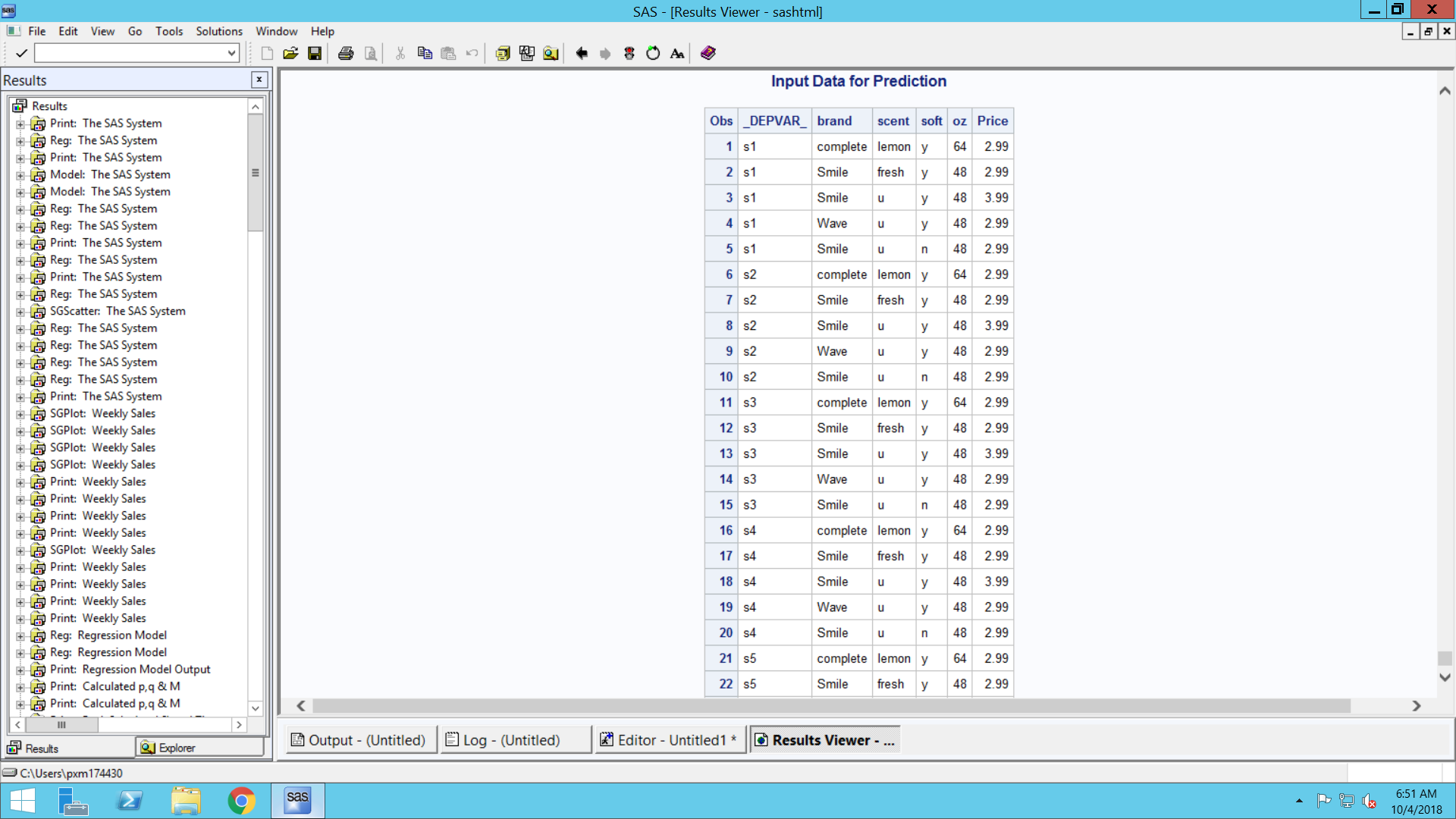
  

Predict the choice (using logit rule) for each respondent (s1-s5) for each of the following combinations using your estimates in question 1 above.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| complete | lemon | y | 64 | 2.99 |  |  |  |  |  |
| Smile | fresh | y | 48 | 2.99 |  |  |  |  |  |
| Smile | u | y | 48 | 3.99 |  |  |  |  |  |
| Wave | u | y | 48 | 2.99 |  |  |  |  |  |
| Smile | u | n | 48 | 2.99 |  |  |  |  |  |

Input Data:

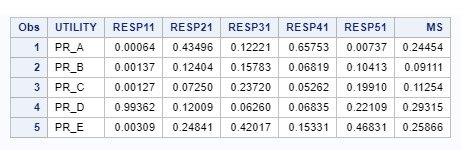




Where PR is the Probability and Utili is the predicted probability and predicted utility values respectively for each respondent for respective combination.



Based on the above dataset, market share is:



**Deliverable:**

1. Electronic copy of SAS code
2. Word document with answers to the above questions.