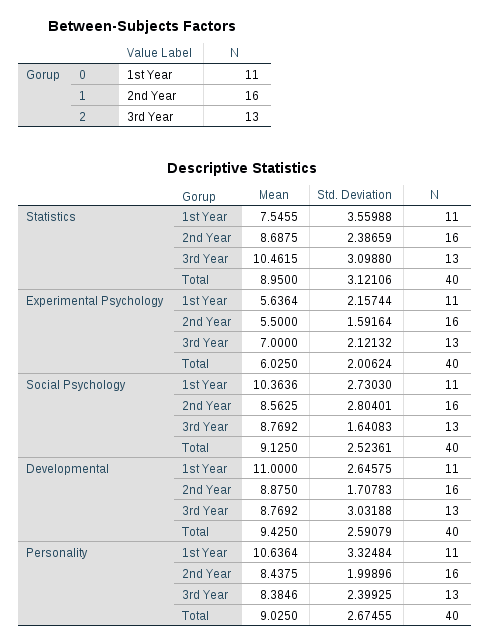
**Cogs 536 Homework 5**

**Ayse Ozdemir - e2340008**

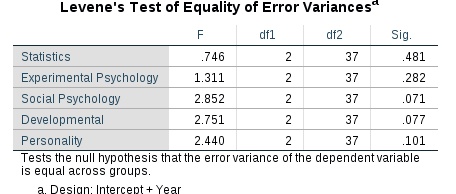
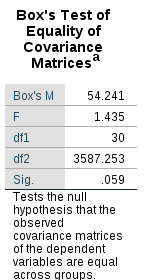
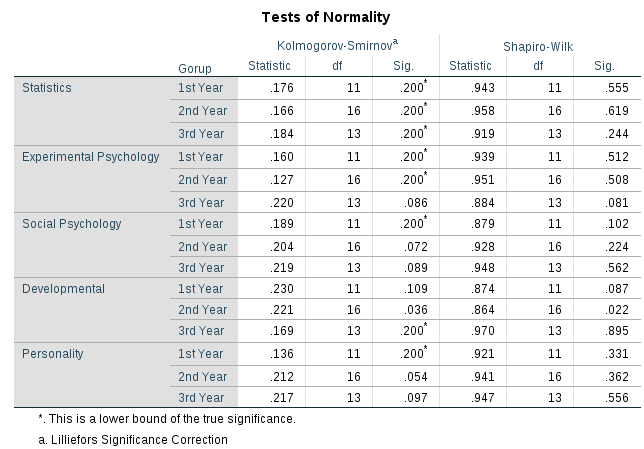
**1.**

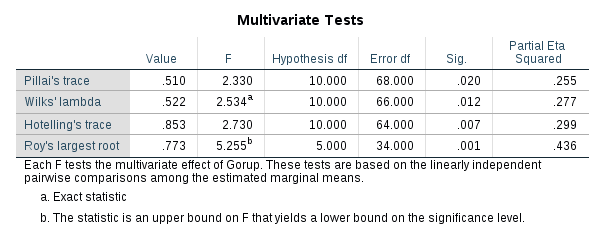
1. Dependent Variables:  
   Statistics, Experimental Psychology, Social Psychology, Developmental Psychology, Psychiatry  
      
   Independent Variables:  
   Groups   
   (Groups independent variable has 3 level(factor) which are; First year students, Second year students, and Third year students)
2. 

**Independence:** Observations should be statistically independent.

Assumption is met.  
  
**Random Sampling and Interval Measure:** Data should be randomly sampled from the population of interest, and should be measured at an interval scale. We have randomly sampled data and, have 5 dependent variables measured at interval scale.

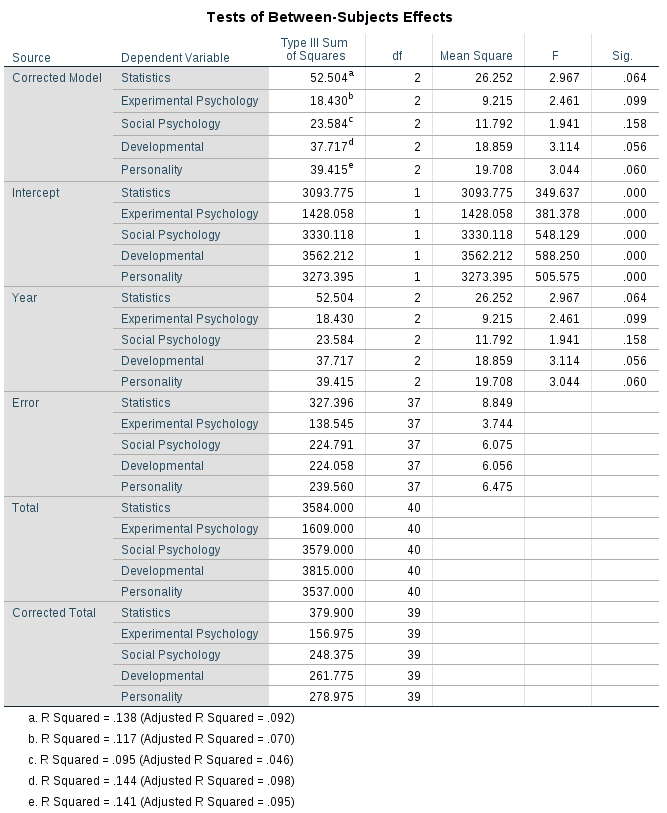
Assumption is met.

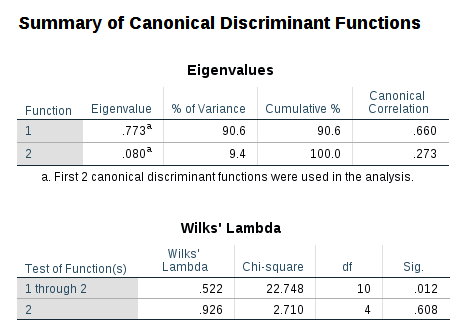
**Multivariate Normality:**According to the Shapiro-Wilk Test; as most of the Sig values are greater than 0.05, the data is could be counted as normal. Assumption is met.  
  **Homogeneity of Covariance matrices:  
  
  
  
  
  
  
  
  
  
  
  
  
  
Box’s Test**; shows the assumption of equality of covariance matrices.   
p > .05, then covariance matrices are equal.  
Assumption is met**.   
  
Levene’s Test**; shows equality of variances for each of the dependent variables.   
Assumption has been met for Statistics, Experimental Psychology, Social Psychology, Developmental, and Personality as these are not significant (p>.05).  
  
Overall, we can conclude that as Levene’s test is non-significant; the assumption of homogeneity of variance has been met.

1.   
     
     
     
     
     
     
     
     
     
     
   For these data, all test statistics are significant (p < .05).   
   We can conclude that *there is significant difference between groups*.

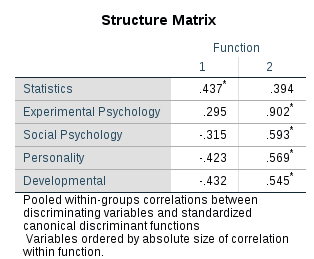
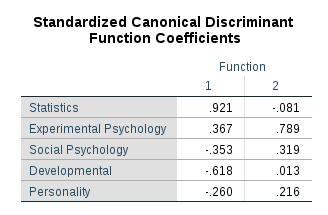
In other words, we can conclude that Multiple choice test scores of different areas of Psychology differs in first, second and third year students.   
  
Using **Roy’s largest root**, there was a significant effect on multiple choice test scored of different areas Pscycology differs in first, second and third years students.

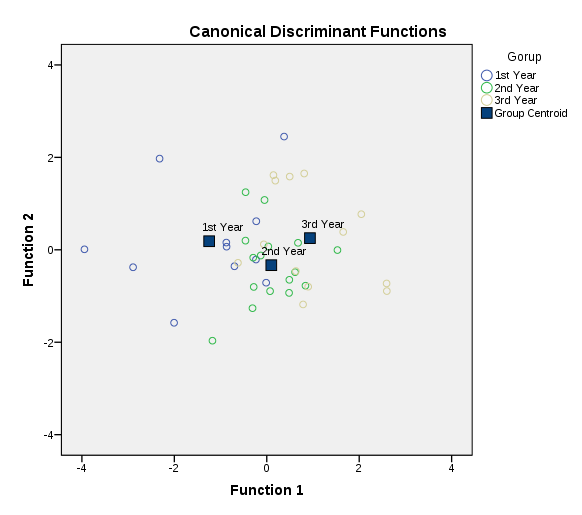
**V=.773, F(5, 34) = 5.26, p<.05**

1.   
     
     
     
   P values indicate that there was a non-significant difference between first, second and third year student groups (p > .05).   
   The *multivariate* test statistics led us to conclude that *the groups did differ*, but *univariate* test results contradicted this.  
   We don’t necessarily conduct contrasts or post-hoc tests as univariate tests results are non-significant. ConductingDiscriminant Analysis (DA) is a good way to see the differences.

1.   
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
   The larger the **E**igenvalue, the more variance the function explains in the outcome variable. Canonical correlation in Function 1 is quite high.   
     
   **W**ilk’s lambda output shows the significance of both variates, and the significance after the first variate has been removed.   
    “1 through 2” model is statistically significant (p=.012), two variates significantly discriminate the groups in combination. But second variate alone is non-significant(p=.608).

Therefore, the group differences can be explained in terms of two underlying dimensions in combination.

1.   
     
     
     
     
     
     
     
     
     
     
     
     
   The table Standardized Canonical Discriminant Function Coefficients explains relative importance of 5 predictors. Statistics is the best predictor with .921 This table is consistent with the “Structure Matrix” table, as they Statistics is the highest match in both (Function 1)

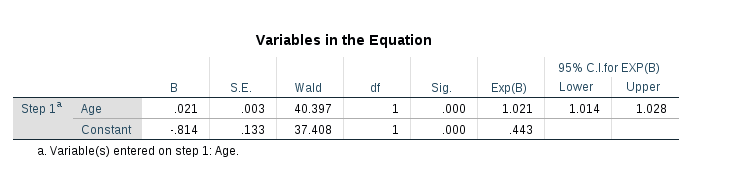
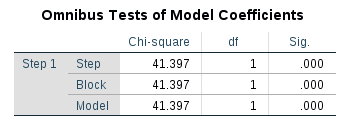
“Functions at Group Centroids” table shows the mean discriminant function scores for each group.  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
**Canonical Discriminant Functions graph shows scores of students grouped by year of student’s degree. Considering the distance between 1st year, 2nd year, and 3rd year group centroids, at horizontal axis, the graph shows that variate 1 discriminates first year students from second year and third year students.**    
Variate 2 is not as successful as variate 1 as it can’t be able to discriminate the group centroids.

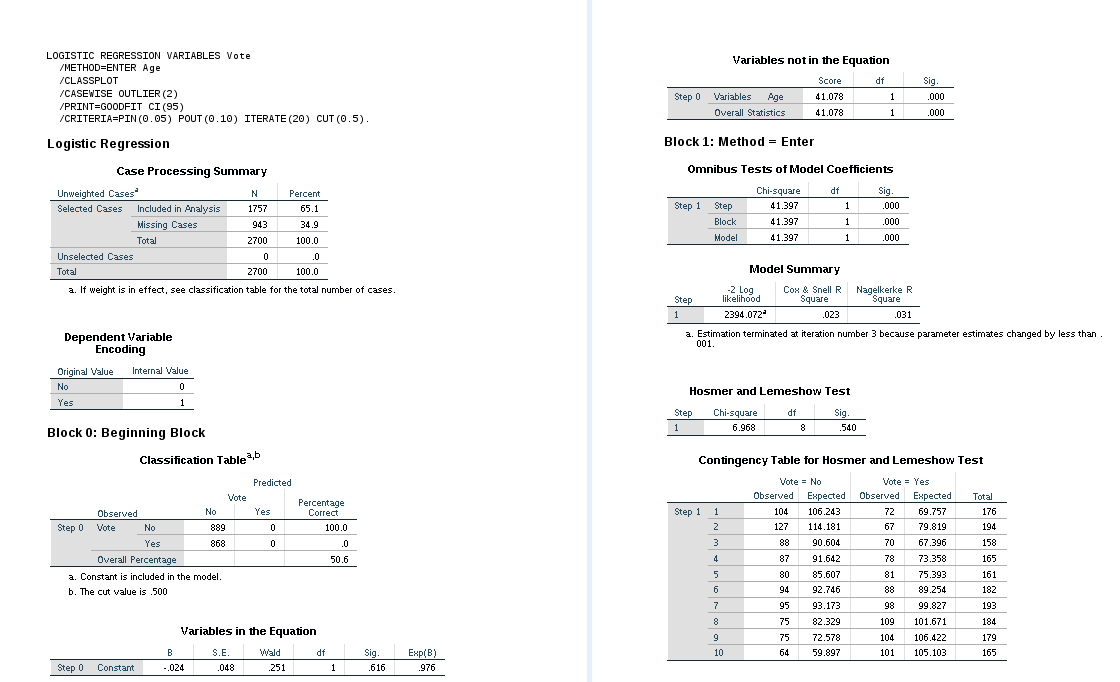
1. The *multivariate* test statistics led us to conclude that *the groups did differ*, but *univariate* test results contradicted this.  
   We don’t necessarily conduct contrasts or post-hoc tests as univariate tests results are non-significant. Conducting Discriminant Analysis (DA) was a good way. The results of DA shows that first year students discriminate from second year and third year students.   
     
   Using **Roy’s largest root**, there was a significant effect on multiple choice test scored of different areas Psychology differs in first, second and third years students.

**V=.773, F(5, 34) = 5.26, p<.05**  
However, separate univariate ANOVA’ s on the outcome variable revealed non-significant treatment effects on;  
Statistics, F(2, 37) = 2.967, p > .05  
Experimental Psychology, F(2, 37) = 2,461, p > .05

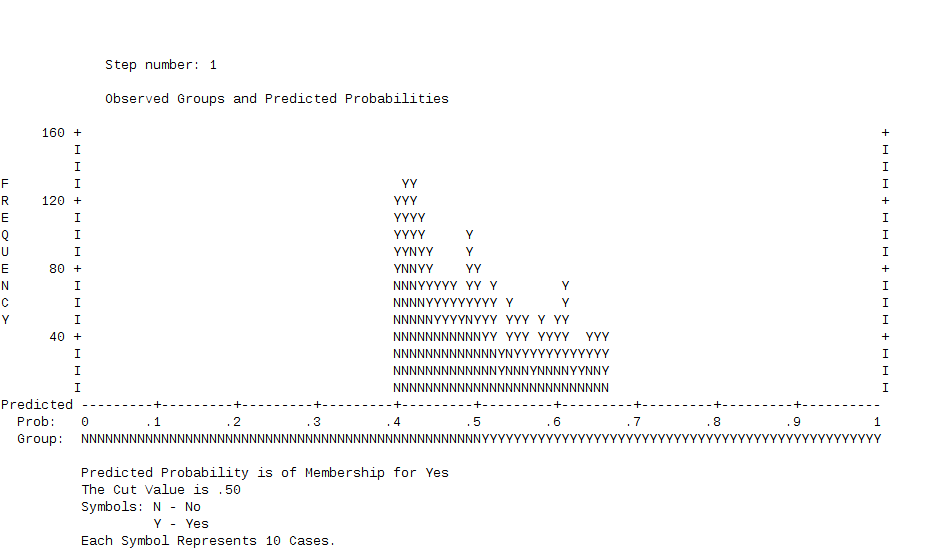
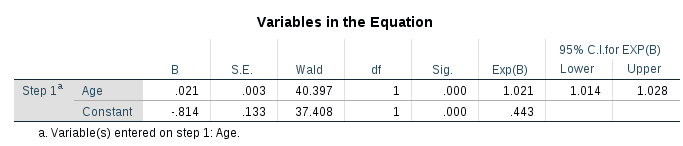
Social Psychology, F(2, 37) = 1,941 , p > .05

Developmental, F(2, 37) = 3,114 , p > .05  
 Personality, F(2, 37) = 3,044 , p > .05

**2.**

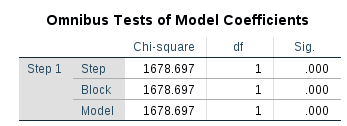
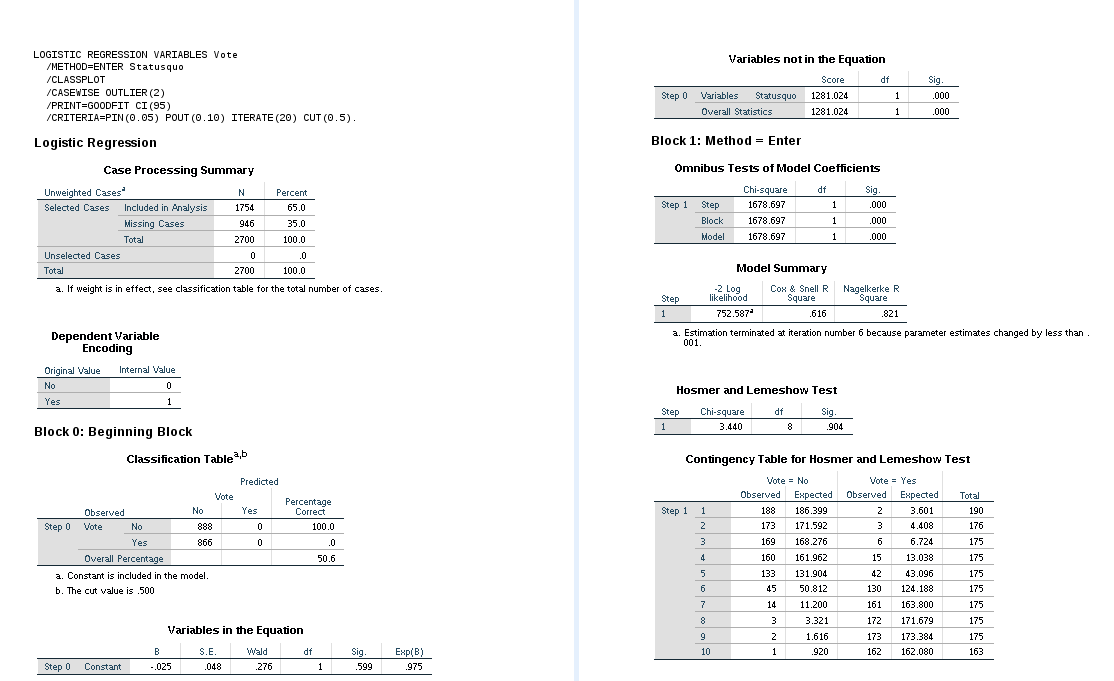
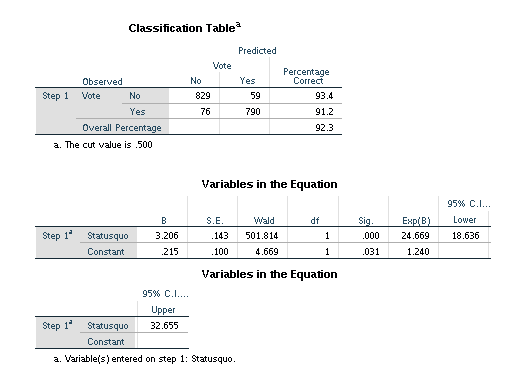
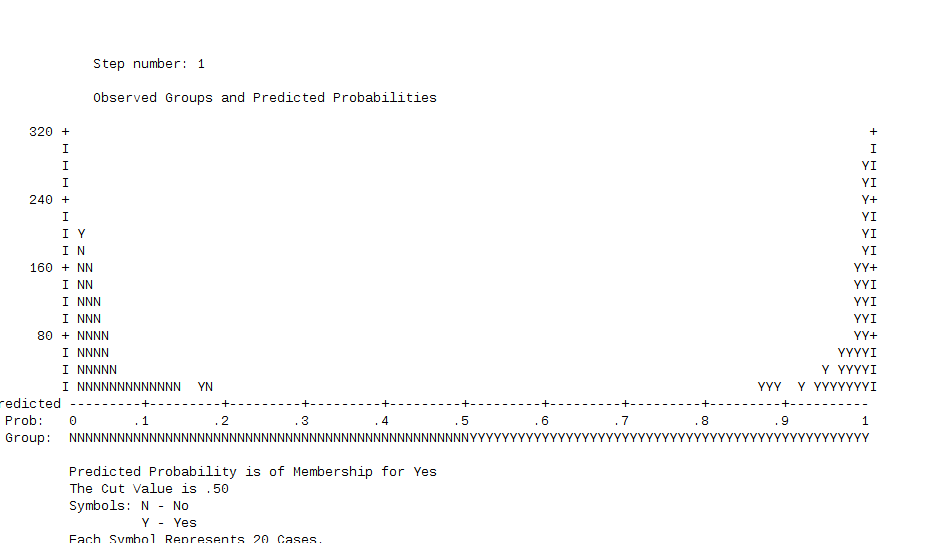
1. age is the only predictor  
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
   a-)  
     
   The value given in the Sig. column is the probability of obtaining the chi-square statistic given that the null hypothesis is true. In other words, this is the probability of obtaining this chi-square statistic is (41.397), if there is in fact no effect of the independent variables on the dependent variable.  
   **Yes**, this model is a significant fit to the data; as **p < .05.**  
     
   b-)

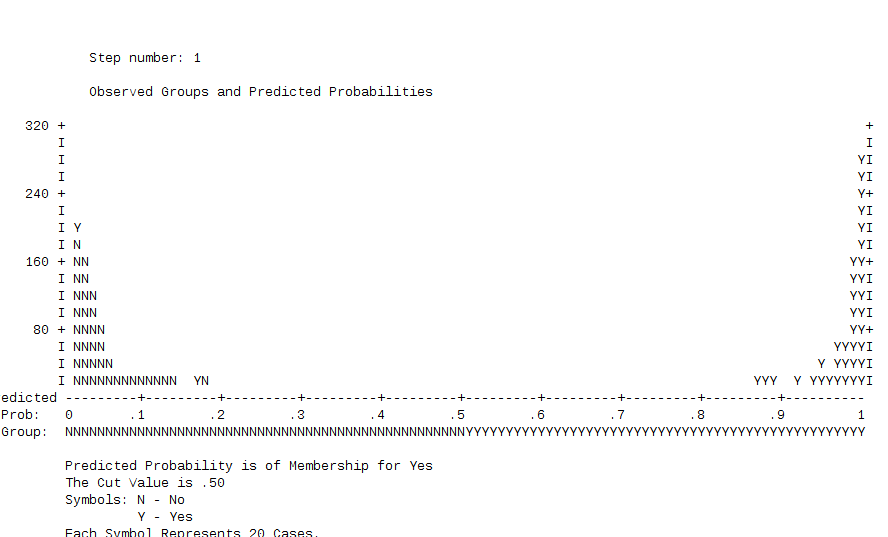
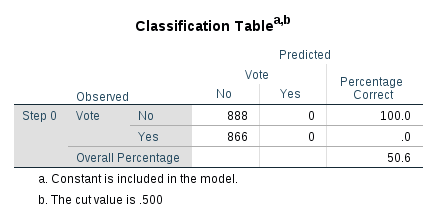
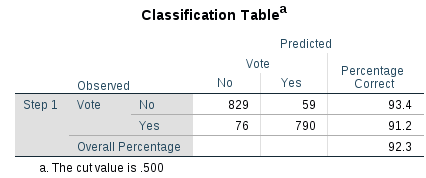
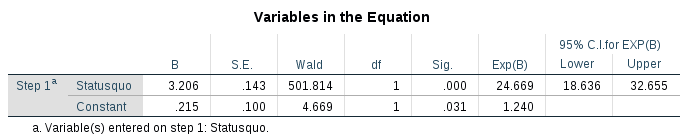
log(p/1-p) = b0 + b1 \* (Age)  
 log(p/1-p) = -.814 + .021 \* (Age)

  
c-)   
  
**Overall Percentage** in the classification tables gives the overall percent of cases that are correctly predicted by the model.   
This percentage has increased from 50.6 for the baseline model to 56.5 for the full model.  
  
d-)  
  
Yes, predictor variable making a significant contribution to the prediction, as after Age variable inserted into the model, prediction percentage has increased from 50.6 for the baseline model to 56.5 for the full model.   
  
e-)  
  
  
  
  
  
  
  
We have a statistically significant value for Age independent variable. Exp(B) an odds ratio of **1.021**.   
The definition of an odds ratio tells that for every unit increase in Age, the odds of the Voting YES increases by a factor of 1.021  
  
We conclude the result above by considering the table below also. The table below shows the predicted probability is of membership for **YES**

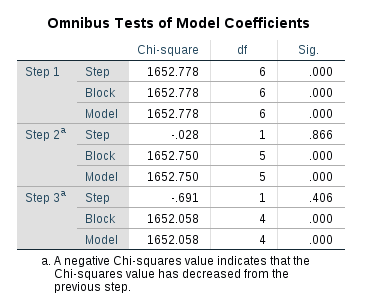
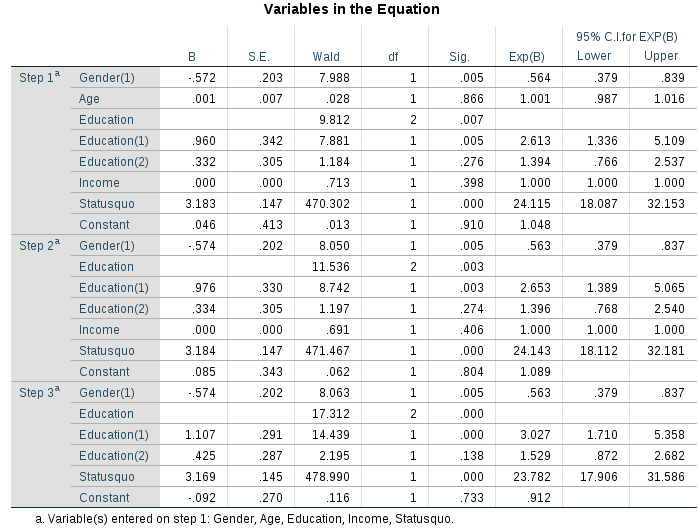
f-)  
  
Considering the coefficients, we did not want the confidence interval to include 0. If we exponentiate 0, we get 1 (exp(0) = 1). Hence, this is two ways of saying the same thing.  
  
According to the table above the 95% confidence interval **includes 1**; hence, the **odds ratio is not statistically significant**. Because the lower and upper bound of the 95% confidence interval is so close to 1, the p-value is very close to .05.

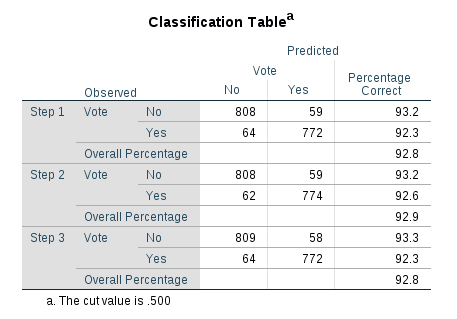
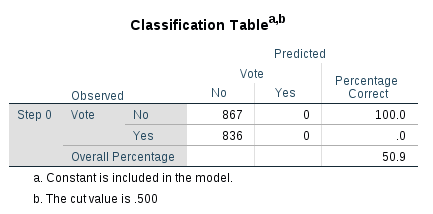
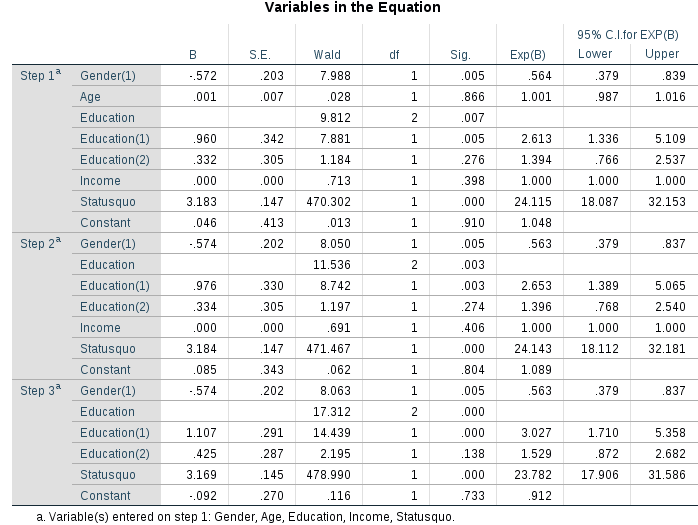
exp(b) > 1 means; while predictor is increasing. Probability of outcome occurring is also increasing.



1. government support is the only predictor  
     
     
   a-)  
     
     
     
     
     
     
     
   The value given in the Sig. Column is the probability of obtaining the chi-square statistic given that the null hypothesis is true.   
   In other words, this is the probability of obtaining this chi-square statistic (1678.697) if there is in fact no effect of the independent variables on the dependent variable.  
     
   **Yes**, this model is a significant fit to the data; as **p < .05.**  
     
   b-)  
     
     
     
     
     
    log(p/1-p) = b0 + b1 \* (Status quo)  
    log(p/1-p) = .215 + 3.206 \* (Status quo)  
     
   c-)

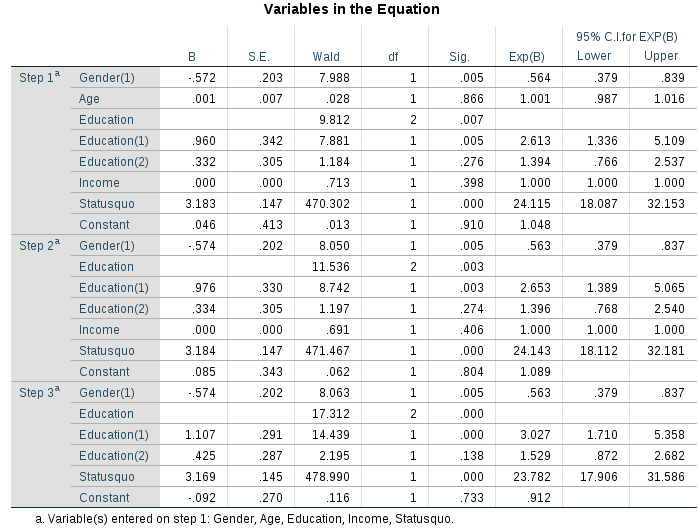
**Overall Percentage** gives the overall percent of cases that are correctly predicted by the model. This percentage has increased from 50.6 for the null model to 92.3 for the full model.  
  
d-)  
  
**Yes**, predictor variable making a significant contribution to the prediction, as after Status quo variable inserted into the model, prediction percentage has increased from 50.6 for the null model to 92.3 for the full model.   
  
e-)  
  
We have a statistically significant value for Status-quo independent variable. Exp(B) an odds ratio of **24.669**.   
  
The definition of an odds ratio tells that for every unit increase in Status-quo, the odds of the Voting YES increases by a factor of 24.669  
  
We conclude this result by considering the table below also. The table shows the predicted probability is of membership for **YES**  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
f-)  
  
According to the table above the 95% confidence interval **includes 1**; hence, the **odds ratio is not statistically significant**. Because the lower and upper bound of the 95% confidence interval is so close to 1, the p-value is very close to .05.   
  
exp(b) > 1 means; while predictor is increasing. Probability of outcome occurring is also increasing.

1. a model including all variables with Backward:Wald as the data entry method.
2.   
     
     
     
     
     
     
     
     
     
     
     
   The value given in the Sig. Column is the probability of obtaining the chi-square statistic given that the null hypothesis is true.   
   In other words, this is the probability of obtaining this chi-square statistic (1652.058) if there is in fact no effect of the independent variables on the dependent variable.  
     
   **Yes**, this model is a significant fit to the data; as **p < .05.**
3.   
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
   log(p/1-p) = -.092 + -.0574 \* Gender(1) + 1.107 \* Education(1) + .425 \* Education(2) + 3.169 \* Status-quo

1.   
     
     
     
     
     
     
     
     
     
   **Overall Percentage** gives the overall percent of cases that are correctly predicted by the model. This percentage has increased from 50.9 for the null model to 92.8 for the full model.
2. **Yes**, predictor variable making a significant contribution to the prediction, as after variables inserted into the model, prediction percentage has increased from 50.9 for the null model to 92.8 for the full model.
3.   
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
   We have a statistically significant value for Gender(1), Education(1), and Status-quo independent variables.   
   Exp(B) an odds ratio of .563 for Gender(1)  
   Exp(B) an odds ratio of 3.027 for Education(1)  
   Exp(B) an odds ratio of 23.782 for Status-quo

The definition of an odds ratio tells that;  
for every unit increase in Gender(1), the odds of the Voting YES increases by a factor of . 563  
for every unit increase in Education(1), the odds of the Voting YES increases by a factor of 3.027

for every unit increase in Status-quo, the odds of the Voting YES increases by a factor of 23.782

1.   
     
     
     
     
     
     
     
     
     
     
      
     
   According to the table above the 95% confidence interval **includes 1**; hence, the **odds ratio is not statistically significant**. Because the lower and upper bound of the 95% confidence interval is so close to 1, the p-value is very close to .05.   
     
   exp(b) > 1 means; while predictor is increasing. Probability of outcome occurring is also increasing.  
   Education(1), Education(2), and Status-quo have exp(b) values which are > 1.  
   They are the variables which make contribution to the model. As Status-quo has highest exp(b), it is the variable which makes the most contribution to the model.

for every unit increase in Status-quo, the odds of the Voting YES increases by a factor of 23.782

1. In the Backward:Wald data entry method, in each step, predictors that don’t have significant wald scores will be removed.  
   In the first step all variables are entered into the model.

In the second step; **Age** variable is removed from the model, as it’s Wald score is not significant. (p > .05, Wald= .028)

In the third step; **Income** variable is removed from the model, as it’s Wald score is not significant. (p > .05, Wald= .691)