**Health Analysis Quantitative Portfolio**

**1. BASS Study:** This study aims to investigate the effectiveness of a one-time educational intervention on reducing body dissatisfaction in women resulting from media images of unrealistic beauty. The study used a randomized controlled trial design, with participants being assigned to one of three groups: an educational intervention with presentation only, an educational intervention with presentation and discussion, or a control group. Body dissatisfaction was measured using the Body Areas Satisfaction Scale (BASS) and Body Mass Index (BMI) was also calculated for all participants.

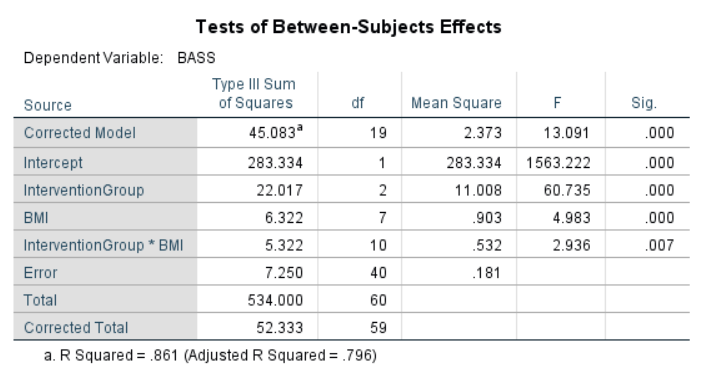
Based on the given dataset, I perform bit cleaning in the dataset like presentation + discussion change to presentation and discussion. Moreover, an ANOVA (Analysis of Variance) test would be useful in this study to determine if there is a statistically significant difference in body dissatisfaction scores (measured by BASS) between the three groups (educational intervention with presentation only, educational intervention with presentation and discussion, and control group).

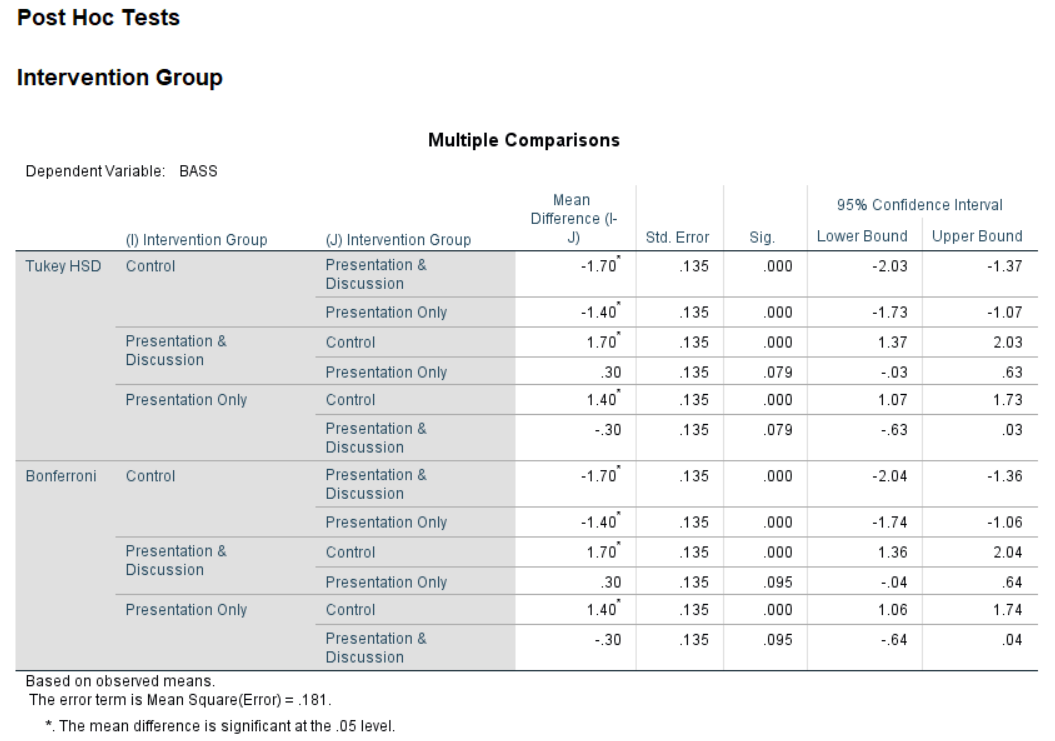
It's a statistical technique that enables to test whether the means of two or more groups are equal or not. The ANOVA test will allow to determine if the educational interventions were effective in reducing body dissatisfaction compared to the control group. If the ANOVA test yields a significant result, the researcher can then conduct post-hoc tests (such as Tukey's test or Bonferroni test) to identify which specific groups differ from one another. Also, there is interaction between the intervention and the BMI, by adding an interaction term between the two variables and run an ANOVA test again.

Setting hypothesis testing:

Null hypothesis (H0): There is no statistically significant difference in BASS scores between the three groups

Alternative hypothesis (H1): There is a statistically significant difference in BASS scores between the three groups





In this Post Hoc Tests analysis, we can see from the table above that there is some of the results but we are only interested in the differences between

1. Control and Presentation & Discussion
2. Control and Presentation only
3. Presentation & Discussion and Presentation Only

As we can see from the results, there is a statistically significant difference between Control and Presentation & Discussion (p < 0.001) and Control and Presentation only (p < 0.001) but not in between Presentation & Discussion and Presentation (p > 0.001).

Now, in between subjects test: The particular rows we are interested in are the "intervention group", "BMI" and "intervention group\*BMI" rows, and these are shown in the first table. These rows inform us whether our independent variables (the "intervention group" and "BMI" rows) and their interaction term have a statistically significant effect on the dependent variable, "BASS".

It is important to first look at the "intervention group\*BMI" interaction term results. We can see from the "Sig." column that we have a statistically significant interaction at the p = 0.007 level. On the other hand, we can see from the table above that there was a statistically significant difference in mean BASS between intervention groups (p = 0.000), inter alia, there were statistically significant differences between BASS & BMI (p < .001).

In this way, we can conclude, the effect of the educational intervention on body dissatisfaction (BASS) may be different depending on the participants' BMI. It could be that the intervention is more effective in reducing body dissatisfaction for participants with a certain BMI range.

Also found a statistically significant difference in mean BASS between the intervention groups, suggesting that the educational intervention had an effect on body dissatisfaction. The results also indicate that there were statistically significant differences between BASS and BMI, which implies that body dissatisfaction may be related to BMI.

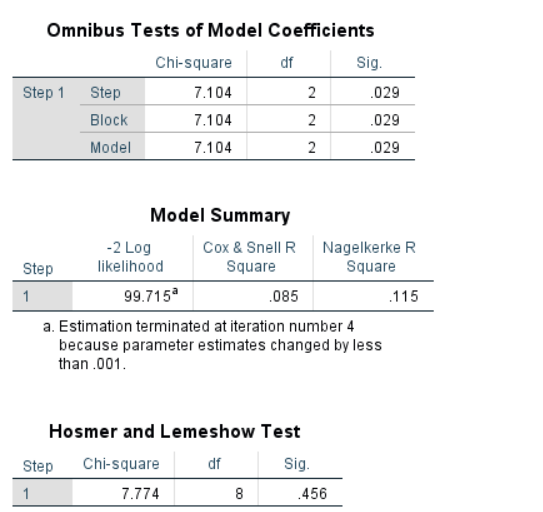
**2. CHD Symptoms:** It is clear that the dataset from a cardiac clinic study, which aimed to examine the relationship between occlusion levels, patient age, and the presence of CHD symptoms. The occlusion level is measured on a scale of 1-100, with higher numbers indicating more severe occlusion. The variable "CHD symptoms" indicates whether the patient has symptoms of CHD, with a value of 1 indicating that symptoms are present, and a value of 0 indicating that symptoms are absent. This dataset contains 80 patient records and could be used to analyze the relationship between occlusion, age, and CHD symptoms, and potentially identify risk factors for CHD.

Logistic regression is used in this dataset because the outcome variable, CHD symptoms, is binary (either present or absent) and the predictor variables, age and occlusion score, are either continuous or ordinal. Logistic regression is a statistical method that is used to model the relationship between a binary outcome variable and one or more predictor variables. It estimates the probability of the outcome variable being a certain value (in this case, 1 for symptoms present and 0 for symptoms absent) based on the values of the predictor variables.

In this study, logistic regression allows us to examine the relationship between occlusion score and age and the presence or absence of CHD symptoms. It helps us to understand how changes in occlusion score and age affect the probability of having CHD symptoms, and to identify the factors that are most strongly associated with CHD symptoms. It also allows us to estimate the odds of having CHD symptoms for different levels of occlusion and age, and to predict the probability of CHD symptoms for new patients based on their occlusion score and age.

To examine whether occlusion score and age of the patient predicts the presence or absence of CHD symptoms, statistical analysis methods such as logistic regression or decision tree could be used. Logistic regression would be appropriate if the outcome variable (CHD symptoms) is binary, and the predictor variables (occlusion score and age) are continuous or ordinal.

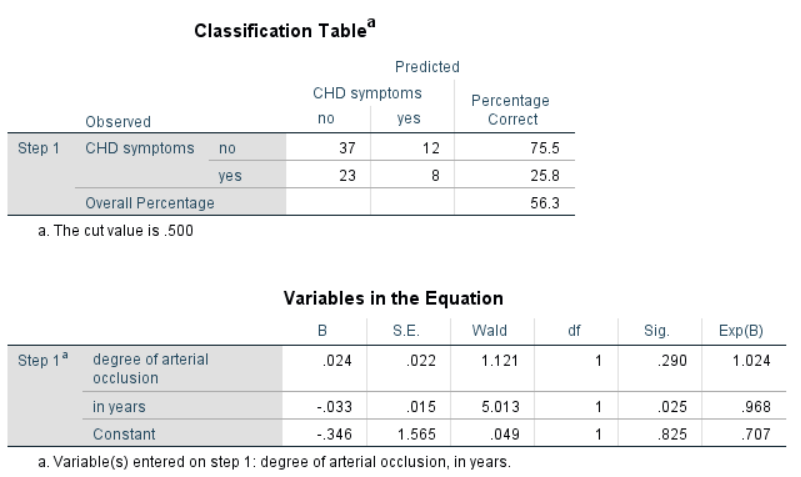
SPSS Results



The most interesting part of the output is that the overall test of the model (in the “Omnibus Tests of Model Coefficients”) and the coefficients and odds ratios (in the “Variables in the Equation” see below result). Thus, the overall model is statistically significant, χ2(4) = 7.104 and p < 0.05.

The above result contains the Cox & Snell R Square and Nagelkerke R Square values, which are both methods of calculating the explained variation. The explained variation in the dependent variable based on our model ranges from 8.5% to 11.50%, depending on whether reference the Cox & Snell R2 or Nagelkerke R2 methods, respectively.

The Hosmer-Lemeshow tests the null hypothesis that predictions made by the model fit perfectly with observed group. A chi-square statistic is computed comparing the observed frequencies with those expected under the linear model. A nonsignificant chi-square indicates that the data fit the model well.



Logistic regression estimates the probability of an event (in this case, having CHD symptoms) occurring. If the estimated probability of the event occurring is greater than or equal to 0.5 (better than even chance), SPSS Statistics classifies the event as occurring (e.g., CHD symptoms being present). If the probability is less than 0.5, SPSS Statistics classifies the event as not occurring (e.g., no CHD symptoms present). It is very common to use binomial logistic regression to predict whether cases can be correctly classified (i.e., predicted) from the independent variables. Therefore, it becomes necessary to have a method to assess the effectiveness of the predicted classification against the actual classification.

With the independent variables added, the model now correctly classifies 56.0% of cases overall i.e., Percentage accuracy in classification. 25.8% (Sensitivity) of participants who had CHD symptoms were also predicted by the model to have CHD symptoms. 84.6% (Specificity) of participants who did not have no CHD symptoms were correctly predicted by the model not to have CHD symptoms.

Now, The Wald test is used to determine statistical significance for each of the independent variables. The statistical significance of the test is found in the "Sig." column. From these results we can see that age (p = 0.025) is added significantly to the model (or prediction), however occlusion (p = 0.290) did not add significantly to the model.

To predict the probability of an event occurring based on a one-unit change in an independent variable when all other independent variables are kept constant. The coefficients indicate that the log-odds of the CHD symptoms variable for a one-unit change (increase or decrease) in the age, occlusion while controlling for the other predictors.

# **References**

Biddle, C. et al. (2019) “Gender bias in clinical decision making emerges when patients with coronary heart disease symptoms also have psychological symptoms,” Heart &amp; Lung, 48(4), pp. 331–338. Available at: <https://doi.org/10.1016/j.hrtlng.2018.11.005>.

Houston, M.C. (2022) “What are the symptoms and signs of coronary heart disease (CHD)?,” The Truth About Heart Disease, pp. 15–18. Available at: <https://doi.org/10.1201/b22808-3>.

Ogden, J. et al. (2011) “The impact of an educational intervention to protect women against the influence of media images,” Health Education, 111(5), pp. 412–424. Available at: <https://doi.org/10.1108/09654281111161239>.