

**Categorical Data Analysis**

**Faith, Ethics, and Beliefs in Science's Benefits**

2nd Report

Prof. Rania Mamdouh

Nada Hamouda Mohamed

5210898

**Introduction**

Science and technology play a significant role in shaping the way we live today, offering countless improvements in health, convenience, and overall quality of life. However, not everyone views these advancements the same way—opinions about their impact can vary widely depending on personal beliefs and demographics.In this project, we set out to explore these diverse perspectives by analyzing what influences people's agreement with the statement: “Science and technology are making our lives healthier, easier, and more comfortable.” Initially, we worked with a three-category response (Completely Agree, Neutral, Completely Disagree), but to simplify the analysis and focus on clear trends, we transformed the responses into two categories: Completely Agree (coded as 1) and Neutral or Completely Disagree (coded as 0)This change allowed us to use binary logistic regression to better understand the factors driving agreement with this statement. By examining beliefs, attitudes, and demographics such as age, gender, and education, this study sheds light on what shapes people’s views on science and technology—and ultimately, how these innovations are perceived in our everyday lives.

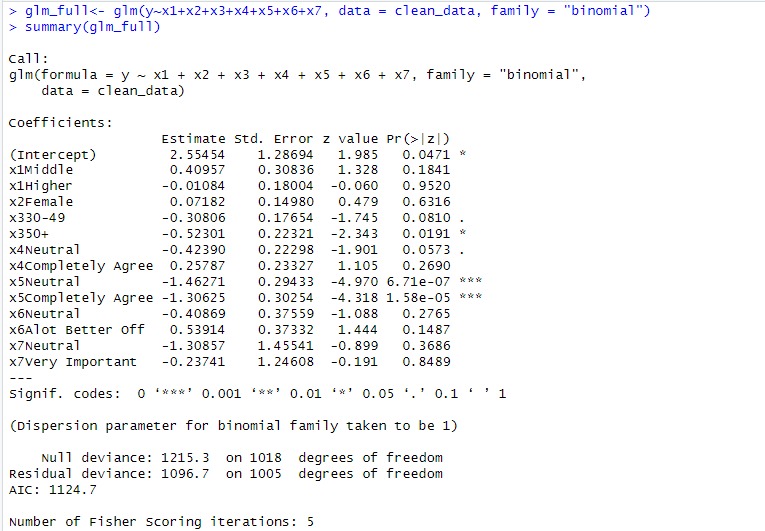
**Fitting Binary Logistic Model: Fitted model:** 𝐋𝐧 ( ) = 𝛂 + 𝛃𝟏D𝟏 + 𝛃𝟐D𝟐 + 𝛃𝟑D𝟑 + 𝛃𝟒D𝟒 + 𝛃𝟓D𝟓 + 𝛃𝟔D𝟔 + 𝛃𝟕D𝟕+ 𝛃𝟕D8+ 𝛃𝟕D9+ 𝛃𝟕D10+ 𝛃𝟕D11 +𝛃𝟕D12+ 𝛃𝟕D13.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| X1 | X2 | X3 | X4 | X5 | X6 | X7 | y |
| Q275R: Highest educational level: Respondent (recoded into 3 groups)  D1,D2 | Q260: Sex  D3 | X003R2: Age recoded (3 intervals)  D4,D5 | Q160: We depend too much on science and not enough on faith.  D6,D7 | Q161: One of the bad effects of science is that it breaks down people’s ideas of right and wrong  D8,D9 | Q163: The world is better off, or worse off, because of science and technology  D10,D11 | Q164: Importance of God  D12,D13 | Q158: Science and technology are making our lives healthier, easier, and more comfortable |

**Estimated model:** 𝐋𝐧() = 2.55454+0.40957D1-0.01084D2+0.07182D3-0.30806D4-0.52301D5-0.42390D6+0.25787D7-1.46271D8-1.30625D9-0.40869D10+0.53914D11-1.38057D12-0.23741D13.

**Checking significance and interpreting parameters**

**H0: βJ= 0 v J= 1, 2, ... ,7. H1: βJ ≠ 0 v J= 1,2, .... ,7**

****

it takes five iterations weighted least squares to obtain maximum likelihood estimates of the parameters

* Don’t reject H0 for every **(βk v k=1,2,3,4,6,7,10,11,12,13)** at level of significance 0.05 as p value for each parameter greater than 0.05 so these parameters have no significance effect.
* Reject H0 for every **(βN v N=5,8,9)** at level of significance 0.05 as p value for each parameter less than 0.05 so these parameters have significance effect.
* According to p-value of chi-square that less than 0.05 that model is better than model with intercept only
* by comparing between the predicted values from the fitted model to those from the null model we found pseudo-R squared equal 0.097
* **Interpreting the significant parameters:**

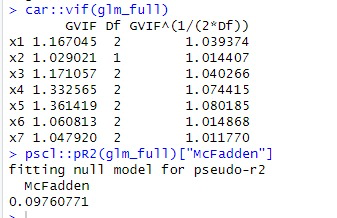
**1-  𝒆𝜷5=-0.523** **=0.593:** Respondents aged 50+ have 0.59 times the odds of agreeing with Q158 compared to the baseline category (ages 16–29). This suggests that older respondents are less likely to agree with the statement.Holding other factors constant.

**2-  𝒆𝜷8=-1.463** **=0.2315:** Respondents who answered Neutral to that science breaks down the ideas of right and wrong have higher odds (0.2943) of agreeing with Q158 compared to those who completely disagreed with Q161.This indicates that respondents in the Neutral category are 34% more likely to agree with Q158 than those in the reference category.Holding other factors constant.

**3-  𝒆𝜷9=-1.31** **=0.269:** Respondents who completely agree that science breaks down the ideas of right and wrong have higher odds (0.269) of agreeing with Q158 compared to those who completely disagreed with Q161.This indicates that respondents in the Completely Agree category are 35% more likely to agree with Q158 than those in the reference category. Holding other factors constant.

* ➢ **Goodness of fit and multicollinearity**

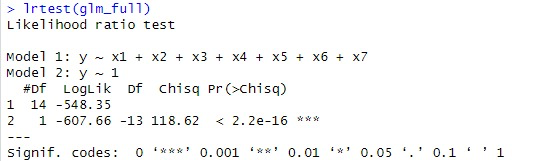
**1-By checking multicollinearity of the model, we found that:**



All variables have VIF less than 10 so there isn’t multicollinearity between the variables in the model

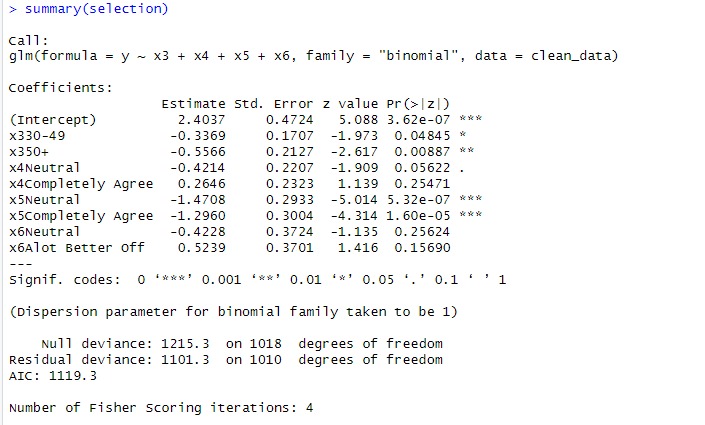
**2- goodness of fit of the model**

1. To determine if the model fits the data well

**Ho: model fits the data well. H1: model doesn’t fit the data well** 

Reject **Ho** at level of significance 0.05 as p-value < 0.05 so the model doesn’t fit the data well.

So we are going to apply the stepwise methods to find the best combination of variables

****

* Don’t reject H0 for every **(βk v k=1,4,5,8,9)** at level of significance 0.05 as p value for each parameter greater than 0.05 so these parameters have no significance effect.
* Reject H0 for every **(βN v N=2,3,6,7)** at level of significance 0.05 as p value for each parameter less than 0.05 so these parameters have significance effect.
* According to p-value of chi-square that less than 0.05 that model is better than model with intercept only
* by comparing between the predicted values from the fitted model to those from the null model we found pseudo-R squared equal 0.0938

**1--  𝒆𝜷5=-0.3369** **=0.713 :** Respondents aged (30-49) have 0.713 times the odds of agreeing with Q158 compared to the baseline category (ages 16–29). This suggests that older respondents are less likely to agree with the statement.Holding other factors constant.

**2-  𝒆𝜷5=-0.5566** **=0.57:** Respondents aged 50+ have 0.57 times the odds of agreeing with Q158 compared to the baseline category (ages 16–29). This suggests that older respondents are less likely to agree with the statement.Holding other factors constant.

**3-  𝒆𝜷8=-1.47=0.229:** Respondents who answered Neutral to that science breaks down the ideas of right and wrong have higher odds (0.229) of agreeing with Q158 compared to those who completely disagreed with Q161.This indicates that respondents in the Neutral category are 34% more likely to agree with Q158 than those in the reference category.Holding other factors constant.

**4-  𝒆𝜷9=-1.29** **=0.275:** Respondents who completely agree that science breaks down the ideas of right and wrong have higher odds (0.275) of agreeing with Q158 compared to those who completely disagreed with Q161.This indicates that respondents in the Completely Agree category are 35% more likely to agree with Q158 than those in the reference category. Holding other factors constant.

* ➢ **Goodness of fit and multicollinearity**

**1-By checking multicollinearity of the model, we found that:**

**A screenshot of a computer

Description automatically generated**

All variables have VIF less than 10 so there isn’t multicollinearity between the variables in the model

**2- goodness of fit of the model**

1. To determine if the model fits the data well

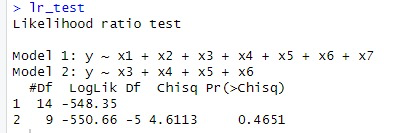
**Ho: model fits the data well. H1: model doesn’t fit the data well**

**A close-up of a computer code

Description automatically generated**

Reject **Ho** at level of significance 0.05 as p-value < 0.05 so the model doesn’t fit the data well.

We perform a likelihood ratio test to compare between the full model and the reduced model to know which fits the data better



**Ho: model 1 is better than model 2 H1: model 1 isn’t better than model 2**

Base on the test, we reject **Ho** at signicance 0.05,means that model 2 is better.

➢ **For summarizing the predictive power of the model**

we will construct classification table to determine

1-  sensitivity which is the probability of correctly predicting a success

2-  specificity which is the probability of correctly predicting a failure

3-  determine misclassification error

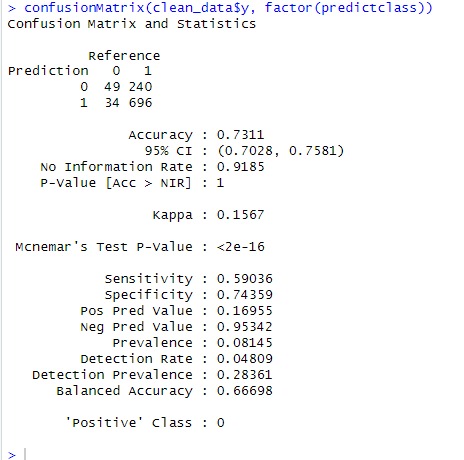
4-  Determine the accuracy of correct classification

But first of all, we need to determine optimal cutoff point as classification table very sensitive to the cutoff point

**A graph showing the growth of a number of cutoff

Description automatically generated with medium confidence**➢ **Determining the optimal cutoff point**We plot the curve of the accuracy of each cutoff point against all possible cutoff points and from the plot:

From this graph we found that at point 0.48 have the maximum accuracy

****➢ **The classification table**

**A close-up of a number

Description automatically generated**

From the classification table we got this

1-  The model classifies 73.1% of the observation correctly which greater than 60% which is good

2-  the probability of correctly predicting the agreement on that science and technology are making our lives healthier, easier, and more comfortable is 0.59 which is greater than 0.6

3-  the probability of correctly predicting predicting the disagreement on that science and technology are making our lives healthier, easier, and more comfortable is 0.74 which is greater than 0.6

4-  the model classifies 26.9% of the observation wrong.

➢ **Roc curve and area under the curve**

**A graph of a curve

Description automatically generated**we will plot Roc curve as it more powerful than classification table as it plots sensitivity as a function in (1-specificity) for the all-possible cutoff points. Thus, it is more informative than a classification table and determine area under the curve which called concordance index

* From this curve we found that it above 45-degree line which the model is better from model with intercept only and the curve close but not too much from point (0,1) that model has good predictive power
* The area under the curve (concordance index) is 0.7 which is higher than 0.5 which the model is better from model with intercept only as it is the probability that the predictions and outcomes are concordant.

**A graph showing the difference between a number of fractions

Description automatically generated**

The ROC curve evaluates the performance of a binary classification model by showing the trade-off between the True Positive Rate and False Positive Rate. The curve rises steeply in the early stages, indicating good model discrimination at higher thresholds .However, as thresholds decrease, FPR increases, reducing precision. Overall, the model performs well, especially at higher thresholds, and its effectiveness can be quantified by calculating the Area Under the Curve (AUC), with higher values reflecting better performance.

**A graph of a person with a beard

Description automatically generated with medium confidence**

The histogram shows that the predicted probabilities from the logistic regression model are concentrated between 0.7 and 0.9, indicating the model frequently assigns high confidence to one class. The slight skew toward higher probabilities suggests strong predictions but requires further evaluation to confirm model performance.

**Conclusion:**

In this study, we examined how various demographic factors and personal beliefs influence people’s agreement with the statement: “Science and technology are making our lives healthier, easier, and more comfortable.” To achieve this, we utilized binary logistic regression by recoding responses into two categories: “Agree” or “Disagree.” This allowed us to simplify the analysis and focus on identifying clear trends.Our analysis considered multiple factors, including age, gender, education, and attitudes toward science and faith.

We explored specific questions related to dependence on science versus faith, perceptions of science's impact on morality, and the importance of religion in shaping views about scientific advancements. The model identified significant and non-significant predictors, highlighting the complexity of how individuals perceive science and technology.Older respondents were less likely to agree with the positive impact of science and technology compared to younger individuals.

Those aged 50+ had significantly lower odds of agreement, indicating generational differences in perceptions of scientific progress.Respondents who felt neutral or agreed that science challenges moral standards were more likely to view science and technology as beneficial.This suggests that concerns about morality play a nuanced role in shaping opinions about the benefits of scientific advancements**.**The model showed reasonable classification accuracy, correctly predicting 73.1% of responses.However, the pseudo-R-squared value (0.097) indicated limited explanatory power, and the goodness-of-fit test suggested the model does not fit the data perfectly.The ROC curve analysis showed moderate predictive performance with an area under the curve (AUC) of 0.7.

This report highlights the relationship between demographics, beliefs, and attitudes in shaping perceptions of science and technology. While some significant trends emerged—such as the influence of age and ethical concerns—the model's limited explanatory power suggests a need for further investigation and refinement.The study provides a foundation for understanding public attitudes towards science but underscores the complexity of the factors at play. Future research could benefit from larger datasets, improved model specifications, and a focus on additional variables to capture these dynamics more effectively.