University of Naples Federico II

Political Sciences Department

Master's Degree Program in
Statistical Sciences for Decision Making

Advanced Analytics for Data Science
Erasmus+ Blended Intensive Program (BIP)

Analysis of University Satisfaction through IRT: A Survey into Students' Expectations and Perceptions

Supervisor:
Professor Maria Iannario

Student:
Antonio Cola
Student ID: M10/399

Academic Year 2022/2023

Contents

1 Introduction

This report aims to explore university satisfaction through the application of the **Graded Response Model**, a model within the **Item Response Theory (IRT)** that allows for a detailed analysis of individuals' responses to specific questions. Through this model, we extracted the **latent variable $\boldsymbol{\theta}$**, reflecting each student's position on a continuum of university satisfaction.
Before reaching this stage, we dedicated ourselves to a thorough analysis and careful cleaning of the dataset. This process ensured the integrity and validity of the data, providing a solid foundation for the subsequent analysis. Once the latent variable θ was obtained, we proceeded with a more in-depth analysis through **linear regression**. This phase will allow us to better understand the factors influencing university satisfaction, identifying any significant relationships.
This research not only provides a detailed framework of university satisfaction through an advanced **IRT-based approach** but also aims to pinpoint key factors contributing to such satisfaction. The ultimate goal is to provide valuable insights for the design of targeted academic policies aimed at improving the university experience for students.

2 Exploratory Analysis

2.1 Description of the Dataset

The **dataset** used in this study was collected through a survey conducted via **Google Forms** at a **university in Bangladesh**. The main objective of this survey was to examine the current academic situation of the university and implement improvements. This inquiry was part of an institutional quality assurance program initiated by the **University Grant Commission of Bangladesh** and funded by the **World Bank**.
The context of this initiative falls within a broader effort by the government of Bangladesh to elevate the quality of higher education to global standards, addressing the challenges of globalization. The **Higher Education Quality Enhancement Project (HEQEP)**, assisted by the World Bank, was initiated by the **Ministry of Education** with the aim of enhancing the teaching-learning and research capabilities of tertiary education institutions.
The original dataset comprises **500 observations** and **87 variables**. Subsequently, to make the dataset more manageable and meaningful, several modifications were made, including the removal of variables deemed not useful, handling of missing data (totaling **7971 missing data**), and the combination of certain variables to improve usability and data interpretation.The result is as follows:

- **gender:** It refers to the gender of the interviewee;
- **faculty:** It refers to the faculty to which the interviewee is enrolled;
- **degree:** It refers to the type of degree of the interviewee (bachelor's or master's);
- **ssc_gpa:** It refers to the Secondary School Certificate (10th Class public exam);
- **hsc_gpa:** It refers to Higher Secondary School Certificate (12th Class public exam);
- **tutoring:** It refers to whether the interviewee has ever attended a training center;
- **regular:** It refers to whether the interviewee is in good standing or not with exams;
- **course:** It refers to the type of courses;
- **global_feedback:** It refers to an overall assessment of university students' perceptions on various aspects, including the academic environment, resources, services, admission policies, and other areas involved in the university experience;
- **q1:** It refers to expectations about the university's education quality;
- **q2:** It refers to expectations about the university's faculty quality;
- **q3:** It refers to expectations about the university's resource quality;
- **q4:** It refers to expectations about the university's learning environment quality;
- **q5:** It refers to the extent to which expectations were met;
- **q6:** It refers to the best aspects of the program;
- **first_aspect:** It refers to the students' opinion on the best aspect of the academic program;
- **second_aspect:** It refers to the students' opinion on the next best aspect of the academic program;
- **improvements:** It refers to aspects of the academic program that could be improved;
- **edu_perception:** It refers to whether students feel that the quality of education improved in the EU over the last year;
- **uni_perception:** It refers to whether students feel that the image of the university improved over the last year.

2.2 Variables Distribution

Understanding the **distribution of variables** in a dataset is important because it enables us to grasp the data's structure and assess the presence of **anomalies** or **patterns**. This, in turn, can aid in making decisions regarding the analysis techniques to employ and formulating reliable conclusions about the dataset.
The distributions examined concern the **dependent variables**.

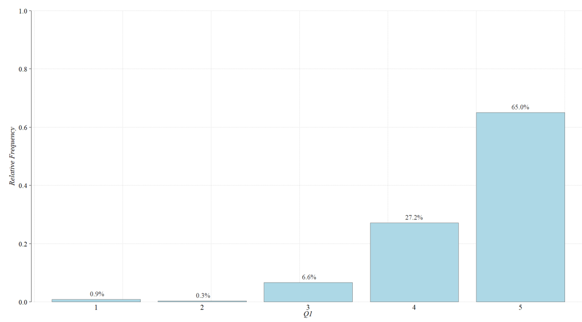

Figure 1: q1 Distribution

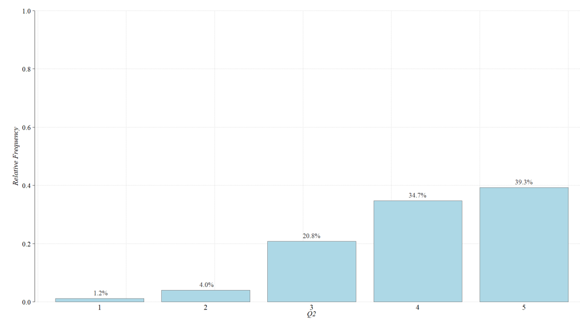

Figure 2: q2 Distribution

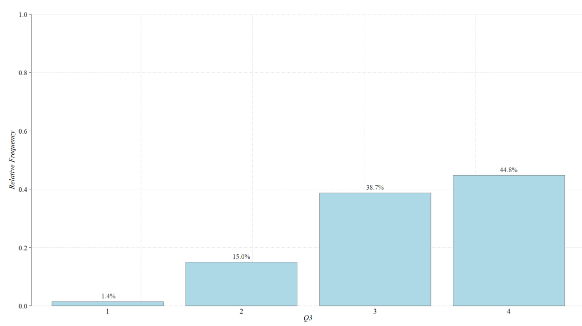

Figure 3: q3 Distribution

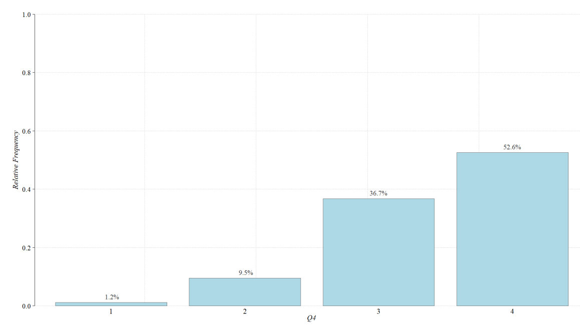

Figure 4: q4 Distribution

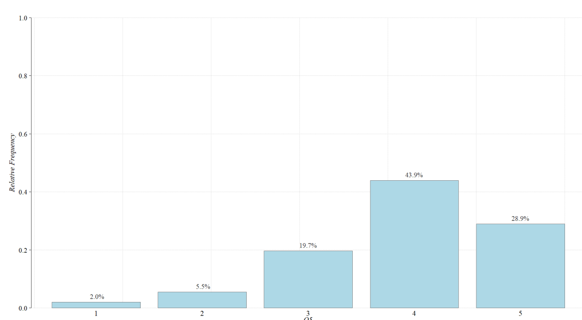

Figure 5: q5 Distribution

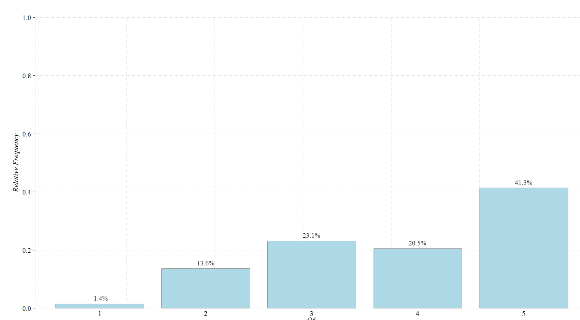

Figure 6: q6 Distribution

The analysis of the bar charts reflects a significant diversity in the expectations and perceptions of university students. While the majority exhibits a positive expectation towards university education, indicating confidence in the quality of faculty, resources, and the learning environment, there are nuances in opinions.
The various percentages of very low or low expectations in some categories suggest the presence of less optimistic voices. However, overall, **the trend is positive**, suggesting that most students perceive the quality of university education favorably. The diversity of perspectives may stem from individual experiences, specific university reputations, or personal expectations. Additionally, specific aspects of the academic program are positively evaluated by the majority of students.

2.3 Correlation

The knowledge of the **correlation** between variables is important because it allows to identify the relationship between two or more quantitative variables. This information is useful for understanding data behavior and can be used to predict the behavior of one variable based on the knowledge of another. In this case, correlations between only the dependent variables were analyzed. The results are reported in the following **correlation matrix**:

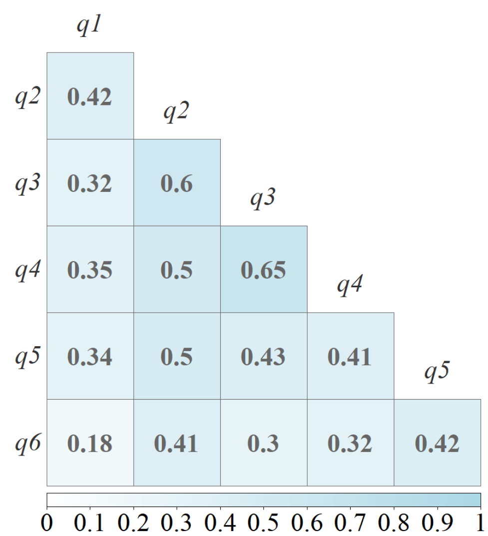

Figure 7: Correlation Matrix

Typically, the correlation matrix is a *symmetric matrix*, meaning it is a square matrix that is equal to its transpose. This implies that element *(i, j)* in the matrix is equal to element *(j, i)* in the same matrix. In this case, to avoid redundant data, the correlation matrix has been reduced to a lower triangular form. Subsequently, data with a unitary value on the main diagonal have been removed. **The correlation values** of the dependent variables **are all positive**, ranging **from a minimum of 0.18 to a maximum of 0.65**. It can be stated that there is a positive linear dependence between all pairs of dependent variables. In other words, one dependent variable increases with the increase of another, to the extent of the correlation coefficient, keeping in mind that the latter is: $-1 < \rho < 1$.
Subsequently, **Cronbach's alpha** was calculated, resulting $\boldsymbol{\alpha = 0.79}$.
The *Cronbach's alpha*, or *Cronbach's alpha coefficient*, is a measure of internal reliability of a *psychometric* measurement tool. It is primarily used in the field of research and psychological assessment to assess the internal consistency of a set of questions or items that constitute a measurement tool, such as a questionnaire or a test.
Cronbach's alpha measures how different questions or items within the tool are correlated with each other, providing an indication of internal cohesion. A higher alpha value indicates greater internal consistency, suggesting that the questions effectively measure the same characteristic or construct.

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} \sigma_{y_i}^2}{\sigma_y^2}\right)$$

Where k represents the number of items, $\sigma_{y_i}^2$ the variance associated with each item i and σ_y^2 the variance associated with the total scores $\left(y = \sum_{i=1}^{k} y_i\right)$.
A Cronbach's alpha value of 0.79 is generally considered quite good, keeping in mind that the latter is: $0 < \alpha < 1$ and indicates a good internal reliability of the measurement instrument used in the research or evaluation. This value suggests that the questions or items included in the instrument are consistent with each other and reliably measure the same characteristic or construct.

3 Models

3.1 Graded Response Model (Item Response Theory)

The **Graded Response Model**, also known as the *Samejima Model* after its developer *Hiroshi Samejima* in 1969, is a statistical model used in the analysis of responses to test items or questionnaires. This model is primarily employed in the field of **Item Response Theory (IRT)**, which is a theory aimed at assessing individuals' abilities or latent traits based on their responses to questions or statements.
The graded response model is a variant within the broader framework of Item Response Models and is specifically designed to handle ordinal or categorical data. This model is often used in situations where individuals' responses to a set of questions can be ordered in a graded manner or into categories. In essence, the graded response model assesses the probability that an individual will give a particular response or fall into a specific category based on their abilities or latent traits. This model takes into account that questions may have more than two response options and that these options can be ordered based on the level of difficulty or ability required to respond correctly.
The model is widely used in the fields of education, psychology, and other areas where it is necessary to assess individuals' abilities or latent traits through questionnaires or tests.

Model formulation:

$$\log \frac{p(Y_{ij} \geq y|\theta_i)}{p(Y_{ij} < y|\theta_i)} = \lambda_j(\theta_i - \beta_{iy}) \quad with \quad j = 1, ..., J \quad and \quad y = 1, ..., I_j - 1$$

Probability of scoring y on item j:

$$p_{iy}(\theta_i) = \frac{e^{\lambda_j(\theta_i - \beta_{jy})}}{1 + e^{\lambda_j(\theta_i - \beta_{jy})}} - \frac{e^{\lambda_j(\theta_i - \beta_{j,y+1})}}{1 + e^{\lambda_j(\theta_i - \beta_{j,y+1})}} = p^*_{jy}(\theta_i) - p^*_{j,y+1}(\theta_i)$$

where:

$$p^*_{jy}(\theta_i) = \frac{e^{\lambda_j(\theta_i - \beta_{jy})}}{1 + e^{\lambda_j(\theta_i - \beta_{jy})}} (2PL\ Model) \qquad p^*_{jI_j}(\theta_i) \equiv 0 \qquad p^*_{j0}(\theta_i) \equiv 1.$$

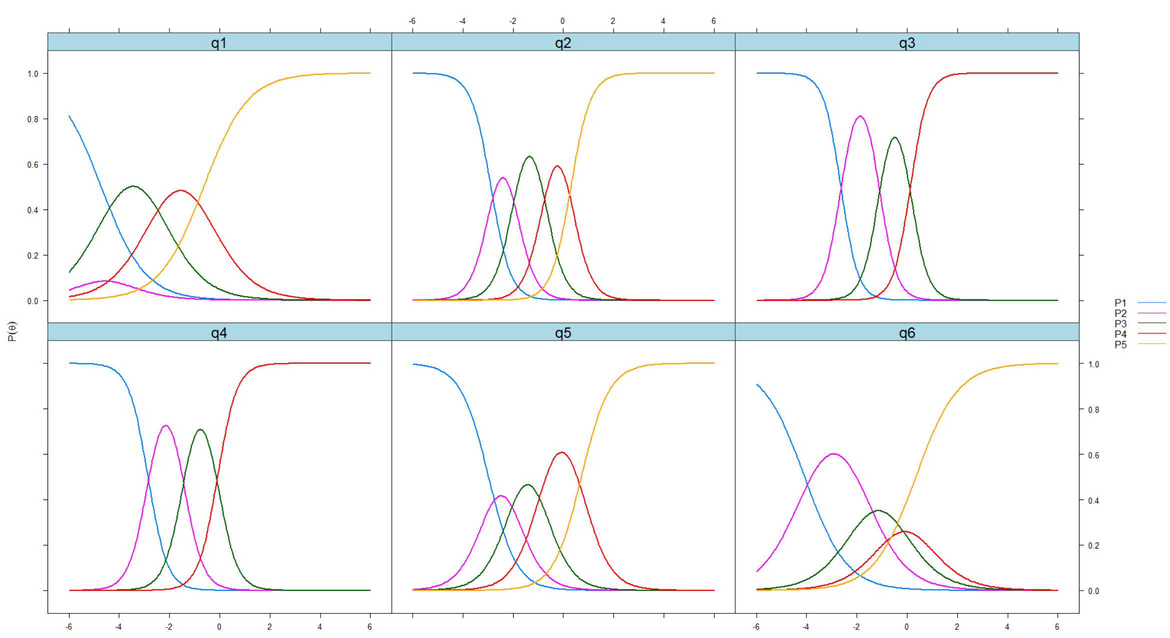

Figure 8: Item Probability Functions

It is interesting to examine the probabilities of responding to specific categories in an item's response scale. These probabilities are graphically displayed in the **category response curves (CRCs)**. Each symmetrical curve represents the probability of endorsing a response category. These curves have a functional relationship with θ, as θ increases, the probability of endorsing a category increases and then decreases as responses transition to the next higher category. The CRCs indicate that the response categories cover a wide range of θ.
To assess the model fit, the **M2 index**, designed specifically for evaluating the fit of item response models for ordinal data, was used. Additionally, the M2-based root mean square error of approximation served as the primary fit index, along with the **standardized root mean square residual (SRMSR)** and the **comparative fit index (CFI)**:

	M2	*df*	*p*	*RMSEA*	*RMSEA_5*	*RMSEA_95*	*SRMSR*	*TLI*	*CFI*
Stats	49.48	9	0	0.11	0.08	0.15	0.06	0.93	0.96

The obtained $\boldsymbol{RMSEA} = \mathbf{0.11}$ (95% CI[0.08, 0.15]) was above a recommended 0.06, $\boldsymbol{SRMSR} = \mathbf{0.06}$ suggest that data fit the model reasonably well using suggested cutoff values of $SRMSR \leq 0.08$ as suggested guidelines for assessing fit. The $\boldsymbol{CFI} = \mathbf{0.96}$ was above a recommended 0.95 threshold.

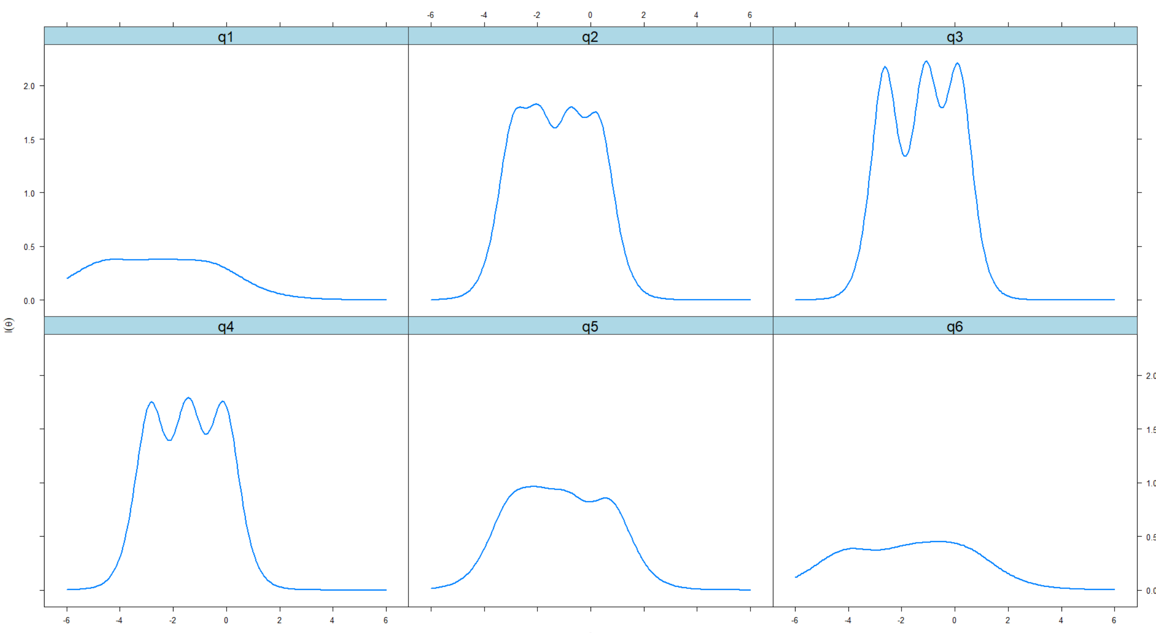

Figure 9: Item Information Curves

In polytomous models, the amount of information an item contributes depends on its **slope parameter**: the larger the parameter, the more information the item provides. Further, the farther apart the **location parameters (b1, b2, b3, b4)**, the more information the item provides. Typically, an optimally informative polytomous item will have a large location and broad category coverage (as indicated by location parameters) over θ.
Information functions are best illustrated by the item information curves for each item as displayed above. These curves show that item information is not a static quantity, rather, it is conditional on levels of θ. The relationship between slopes and information is illustrated here. **Item 1 had the lowest slope** and is, therefore, the least informative item. On the other hand, **Item 3 had the highest slope** and provides the highest amount of statistical information. Items tended to provide the **most information between [-3, 0] $\boldsymbol{\theta}$ range** . The wavy form of the curves reflects the fact that item information is a composite of category information, that is, each category has an information function which is then combined to form the item information function.

	a	$b1$	$b2$	$b3$	$b4$
$q1$	1.14	-4.71	-4.42	-2.49	-0.64
$q2$	2.56	-2.88	-1.94	-0.77	0.29
$q3$	2.93	-2.64	-1.10	0.13	-
$q4$	2.61	-2.85	-1.44	-0.09	-
$q5$	1.80	-2.98	-1.99	-0.86	0.71
$q6$	1.20	-4.08	-1.76	-0.54	0.34

The **latent variable or ability $\boldsymbol{\theta}$** could represent the overall assessment or overall satisfaction of the individual's university experience. It might reflect the program's ability to meet students' expectations and provide a positive educational experience.
Once the latent variable θ was obtained, we proceeded with a more in-depth analysis through linear regression.

3.2 Linear Regression

The **Linear Regression Model** is a statistical method used to examine the relationship between a *dependent variable* (or *response*) and one or more *independent variables* (or *predictors*). The goal is to identify and quantify the linear relationship between the variables so that it can be used to make predictions about the dependent variable based on the values of the independent variables. The regression coefficients provide information about the strengths and directions of these relationships.
The *first step* involved adopting a sequential strategy (*stepwise regression*) for the selection of explanatory variables. In particular, the *backward elimination* strategy was chosen, which is a *step-down* procedure. It starts with the most complete regression model, including all explanatory variables, progressively eliminating variables that are not significant. Then, a new model is estimated at each step with the significant variables from the previous step. Finally, the procedure stops when all the variables present in the model have been found significant at a certain level, meaning that all calculated p-values fall below a predetermined critical threshold ($\boldsymbol{p-value < 0.1}$).

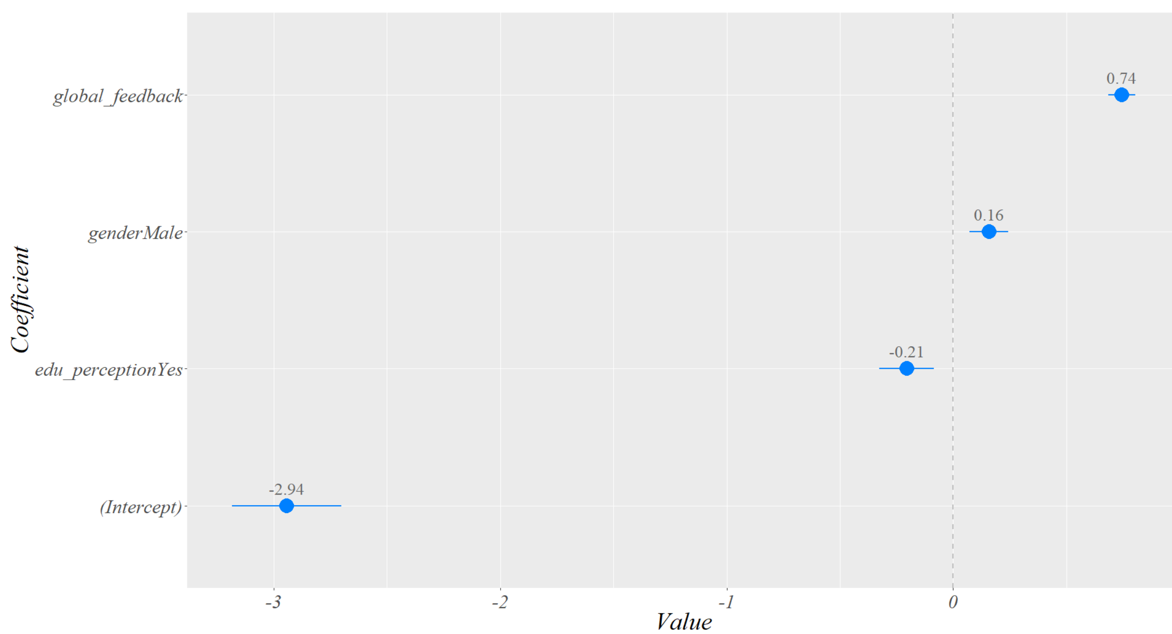

Figure 10: Linear Regression Model

The obtained model takes the following form:

$$ability(\theta) = -2.94 - 0.21\, edu\_percetion Yes + 0.16\, genderMale + 0.74\, global\_feedback$$

The overall assessment or overall satisfaction decreases by 0.21 for individuals who believe that the quality of education has improved at the University in the last year, increases by 0.16 for males, and increases by 0.74 for each unit increase in the overall assessment of university students' perceptions on various aspects. Here are some indices to compare the full model (M_c) with the final model (M_f):

	Full Model	*Final Model*
AIC	832.59	788.71
BIC	1009.52	807.94
logLik	-370.29	-389.36
Adj. R^2	0.31	0.32

Having used the backward elimination method for selecting explanatory variables, we are dealing with nested models. The table presents four evaluation criteria: **Akaike Information Criterion (AIC)**, **Bayesian Information Criterion (BIC)**, **log-likelihood (log-Lik)** and **Adjusted $\mathbf{R^2}$ (Adj. $\mathbf{R^2}$)**. We prefer the model with lower values for AIC, BIC and logLik, and bigger value for Adj. R^2. Furthermore:

$$BIC = AIC + (p+1)[\log n - 2]$$

It will always be:

$$BIC > AIC \quad if \quad n > 8$$

Therefore, the BIC index is more "stringent" in the choice among multiple models. The AIC index tends to overparameterize the models it selects, so BIC is preferred when using a parsimonious model is important. The AIC index is biased, while the BIC index for common models is consistent and as n increases, it specifies the correct model with probability 1. However, the BIC and AIC indices are used to compare non-nested models.
Another method for comparing nested models is the **likelihood ratio test (LRT)**:

$$LRT = -2(\ell(M_f) - \ell(M_c)) = -2(-389.36 - -370.29) = 38.12 \to \chi^2_{g=41} \qquad p - value = 0.6$$

Given that the $\boldsymbol{p - value > 0.05}$, **we don't reject the null hypothesis**. This means that the full model and the final model fit the data equally well. Therefore, we should use the final model because the additional predictor variables in the full model do not offer a significant improvement in fit.

The *second step* was the **model diagnostics**. The diagnosis of a linear regression model is a process aimed at analyzing and evaluating the model's performance. It involves checking the validity of the assumptions of linear regression and identifying any issues or anomalies in the data or the model. Common aspects of diagnostics include assessing **linearity, homoscedasticity, normality of residuals**, and detecting any influential observations or outliers. Diagnostics are essential to ensure that the model provides reliable and valid estimates of the relationship between the involved variables.

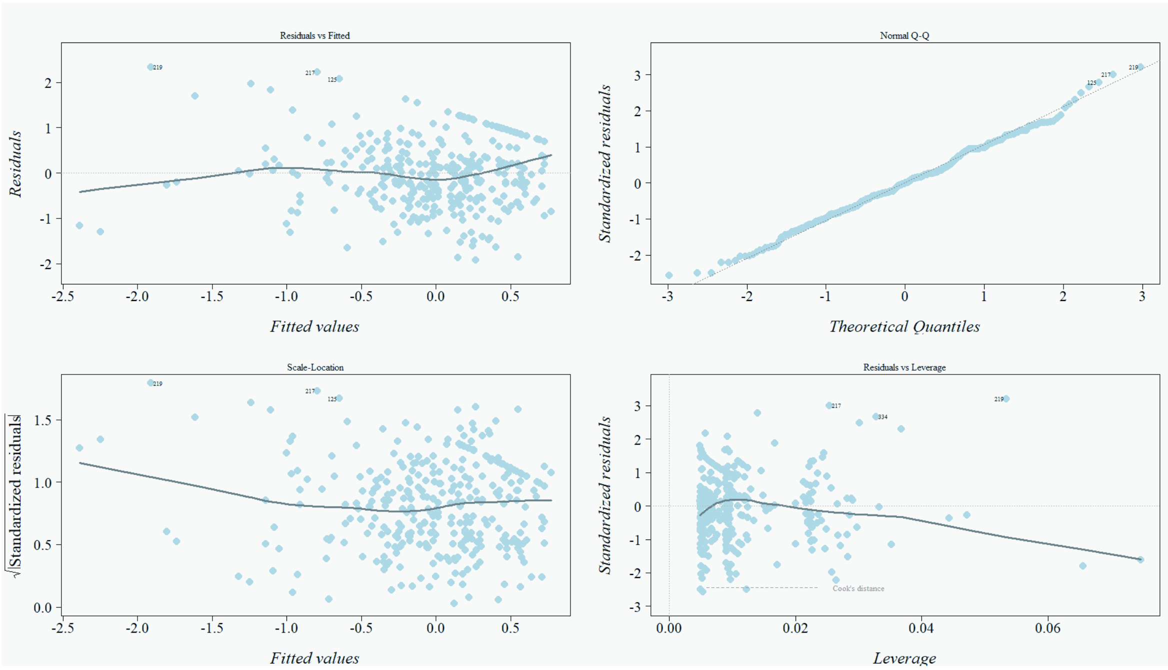

Figure 11: Residuals Distributions

As can be seen from the plotted graphs, **homoscedasticity, linearity, and normality of residuals are verified**. Further confirmation is provided by the **Shapiro-Wilk Test** for normality:

W Statistic	*p-value*
0.99536	0.3918

The results suggest that there is not enough evidence to reject the hypothesis that the residuals follow a normal distribution.

4 Test

4.1 Chi-Squared Test (Independence Test)

The verification of **independence** between two or more phenomena constitutes a preliminary phase prior to the establishment of more explicit and significant relationships, especially in the context of constructing a statistical model.
The procedure of the test can be summarized as follows:

Null Hypothesis	*Alternative Hypothesis*	*Critical Region*
H_0 : X and Y are Indipendent	H_1 : X and Y are not Indipendent	$RC(\alpha) : X^2 > \chi^2_{(\alpha, g=(k-1)(h-1))}$

The test statistic, also known as **Chi-Squared Test**, can be calculated as follow:

$$X^2 = n\left(\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{(n_{ij})^2}{n_{i.}n_{.j}} - 1\right)$$

Below is the table related to the independence tests conducted on the data:

Variable 1	*Variable 2*	*Chi-Squared*	*p-value*	*Signif.*
abilty(θ)	gender	203.6004	2.53×10^{-01}	
abilty(θ)	faculty	324.3752	9.852×10^{-01}	
abilty(θ)	degree	181.7899	6.718×10^{-01}	
abilty(θ)	ssc_gpa	10965.9192	7.737×10^{-01}	
abilty(θ)	hsc_gpa	10804.6671	1	
abilty(θ)	tutoring	184.2729	6.233×10^{-01}	
abilty(θ)	regular	223.4845	5.371×10^{-02}	.
abilty(θ)	course	731.1296	7.985×10^{-01}	
abilty(θ)	global_feedback	31624.6012	1.025×10^{-13}	***
abilty(θ)	first_aspect	671.4013	2.763×10^{-03}	**
abilty(θ)	second_aspect	763.8872	4.943×10^{-01}	
abilty(θ)	improvements	4520.5157	3.399×10^{-04}	***
abilty(θ)	edu_perception	236.9464	1.326×10^{-02}	*
abilty(θ)	uni_perception	254.4025	1.465×10^{-03}	**

Table 1: Chi-Squared Test Table

Below is the table to interpret the level of significance of the tests:

Signif.	*%*
***	0.1
**	1
*	5
.	10
	Not Signif.

References

[1] Agresti A. *Analysus of Ordinal Categorical Data*. Wiley, 2 edition, 2010.

[2] Agresti A. *Categorical Data Analysis*. Wiley, 3 edition, 2013.

[3] Dobson A. *An Introduction to Generalized Linear Model*. Springer, 2 edition, 1990.

[4] Frank B Baker, Seock-Ho Kim, et al. *The basics of item response theory using R*. Springer, 2017.

[5] Li Cai and Mark Hansen. Limited-information goodness-of-fit testing of hierarchical item factor models. *British Journal of Mathematical and Statistical Psychology*, 66(2), 2013.

[6] R Philip Chalmers. mirt: A multidimensional item response theory package for the r environment. *Journal of statistical Software*, 48, 2012.

[7] Lee J Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334, 1951.

[8] Piccolo D. *Statistica*. Il Mulino, 3 edition, 2010.

[9] Murphy K. P. *Machine Learning: A Probabilistic Perspective*. MIT press, 1 edition, 2012.

[10] Fumiko Samejima. Graded response model. In *Handbook of modern item response theory*, pages 85–100. Springer, 1997.