

Advanced Analytics for Data Science (BIP)

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

Antonio Cola

University of Naples Federico II

October 20, 2023



- ① Introduction
- ② Exploratory Analysis
- ③ Models
- ④ Test
- ⑤ Conclusion

1 Introduction

2 Exploratory Analysis

3 Models

4 Test

5 Conclusion

Data Preprocessing

The dataset, available on **Kaggle**, is the result of a survey conducted through **Google Forms** by the **University Grant Commission of Bangladesh**. The original dataset has been cleaned and reduced in terms of both observations and variables for improved usability:

Observations	500
Variables	87
Missing Data	7971

Table 1: Original Dataset



Observations	346
Variables	20
Missing Data	0

Table 2: After Cleaning Dataset

1 Introduction

2 Exploratory Analysis

3 Models

4 Test

5 Conclusion

Distribution of items

The overall trend of the distribution of items is positive.

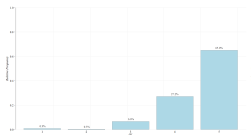


Figure 1: q1

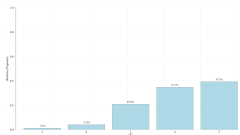


Figure 2: q2

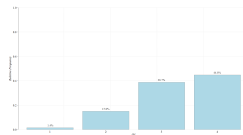


Figure 3: q3

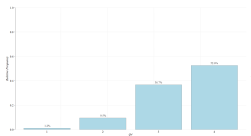


Figure 4: q4

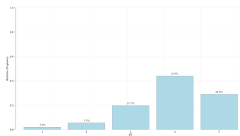


Figure 5: q5

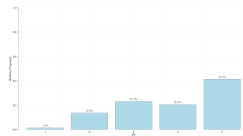


Figure 6: q6

Correlation

The **correlation** values of the items are all positive, ranging from a minimum of 0.18 to a maximum of 0.65.

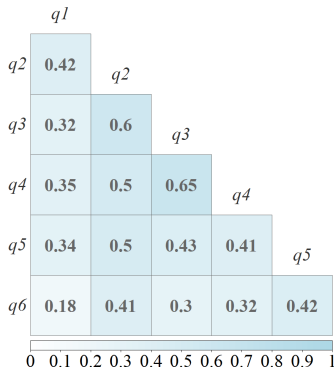


Figure 7: Correlation Matrix

Correlation

Cronbach's Alpha Coefficient measures how different questions or items within the tool are correlated with each other, providing an indication of internal cohesion.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_{y_i}^2}{\sigma_y^2} \right) = 0.79$$

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 $\left(y = \sum_{i=1}^k y_i \right)$.

- 1 Introduction
- 2 Exploratory Analysis
- 3 Models**
- 4 Test
- 5 Conclusion

Graded Response Model (IRT)

Graded Response Model (Hiroshi Samejima, 1969) is specifically designed to handle ordinal or categorical data:

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

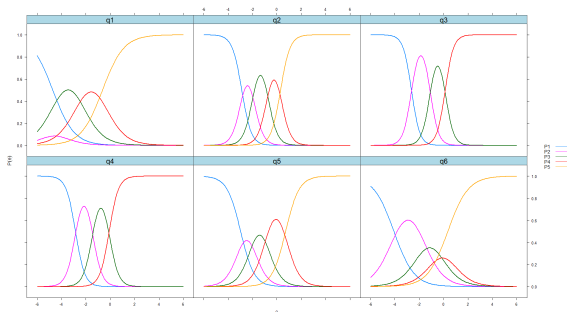


Figure 8: Item Probability Functions

Graded Response Model (IRT)

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 $[-3, 0]$ θ range .

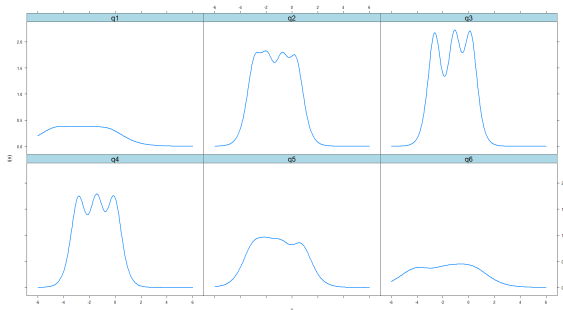


Figure 9: *Item Information Curves*

Linear Regression on Latent Variable θ

The **latent variable** or **ability θ** could represent the overall satisfaction of the individual's university experience. Once the latent variable θ was obtained, we proceeded with a more in-depth analysis through **linear regression**.

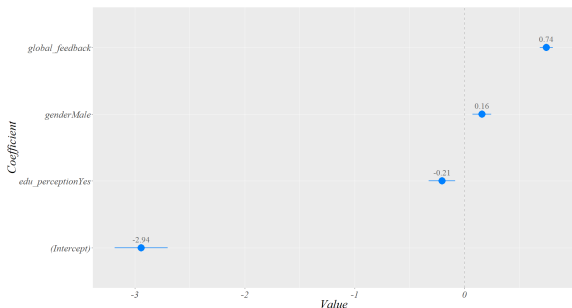


Figure 10: Linear Regression Model

Linear Regression Diagnostic

As can be seen from the plotted graphs, **homoscedasticity, linearity, and normality of residuals are verified.**

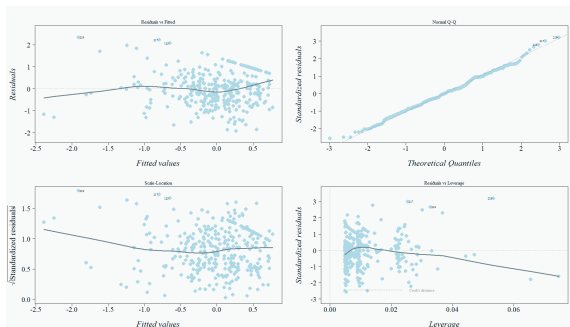


Figure 11: Residuals Distributions

- 1 Introduction
- 2 Exploratory Analysis
- 3 Models
- 4 Test**
- 5 Conclusion

Chi-Squared Test (Indipendence)

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

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

<i>Variable 1</i>	<i>Variable 2</i>	<i>Chi-Squared</i>	<i>p-value</i>	<i>Signif.</i>
abilty(θ)	regular	223.4845	5.371×10^{-02}	.
abilty(θ)	first_aspect	671.4013	2.763×10^{-03}	**
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 3: Chi-Squared Test Table

- ① Introduction
- ② Exploratory Analysis
- ③ Models
- ④ Test
- ⑤ Conclusion**

References

- [1] Agresti A.
Analysis 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.

Thanks!