MIS 5375

Final Project

*Gym Attendance Data Analysis*

*Elizabeth Elizondo*

*October 5, 2021*

# Introduction

XYZ gym has hired a marketing firm to increase member visits due to low monthly attendance. The firm surveyed 180 people from a local gym to decide how they pursue a marketing strategy to target a young-adult aged audience to increase member visits at XYZ gym. It is believed that there is a correlation to higher fitness levels and higher income levels where age and marital status may have a subsequent correlation. The survey included the participants’ age, income level, personal fitness level they believe to have (scale of 1-5, 5 being the best), and whether they were partnered. The research conducted for this study utilized several software applications to produce a holistic assessment of the data. Microsoft Excel was utilized to clean and prepare the data for analysis and to produce descriptive statistics. The Knime analytics platform was utilized for data visualization to help interpret the variables being analyzed. Finally, the statistics platform WarpPLS was utilized to analyze the data in depth to better understand how the latent variables correlate to one another. Additionally, to better understand and interpret the construct reliability of each variable and determine whether those variables have convergent or discriminant validity, and if collinearity exists. The results revealed that the proposed question of whether higher fitness levels influence higher income from this sample population were proved to have substantial significance, whereas the independent latent variables marital status acted more as a mediator, having some significance, and age had little to no correlation but performing as a moderator variable (5).

# Data summary

The data includes three possible predictor variables which include fitness level, age, and marital status. The data consists of four latent variables: Age, income level, personal fitness level they believe to have (scale of 1-5, 5 being the best), and marital status as partnered or single. The marital status was modified in excel to represent a numerical value for analysis purposes, where 1 is married, and 0 is single. The latent variables included in the set for WarpPLS analysis were input as follows: FitLevel – Self-rated fitness on an 1-to-5 Scale where 1 is poor shape and 5 is excellent shape. Age – Age of participant, Income – Income of participant, MStatus – Marital status of participant. The data was collected by a market research team which surveyed 180 gym participants from XYZ gym.

# Research questions

The questions the market research team is looking to answer is, does fitness level correlate to higher income levels at XYZ gym, and does age or marital status have an impact on those predictors which affect the outcome?

# Descriptive statistics

Figure A1. provides detail for descriptive statistics in categories of age, fitness level, and income. Observing the category *age* has a range between 18-50 years of age, Figure A. reveals the mean age is 29 with a median of 26 and a mode of 25. This tells the marketing group that their target age group is in the mid-twenties, however the standard deviation is 7 which is considered high and may yield too wide of a range of results thereby prohibiting the age category to reflect a direct correlation to the categories of *fitness* or *income* to reveal any type of causation, but, may act as a moderating variable in this analysis.

Graphical user interface

Description automatically generated

**Figure A1. Descriptive Statistics as gathered in Excel for XYZ Gym Analysis**

# Knime Visualizations

The Knime visualization shown in Figure A2. below, provides an immediate snapshot of the surveyed results of marital status compared to income level. In this chart, it can be seen that a majority of participants surveyed were partnered, and those that had a slightly lower income level were single. This could be useful to the marketing firm to target their audience for *couples* in their future advertisements for this gym. Figure A3. demonstrates a stacked area chart which appears to have a direct linear relationship between fitness level and income, resulting in a direct relationship between the two variables. This could be useful to a marketing team, which may choose to use the claim that exercise increases the likelihood of greater income, thereby advertising that claim to increase gym participation.

Chart, sunburst chart

Description automatically generated

**Figure A2. Marital status compared to income level**

Chart, histogram

Description automatically generated

**Figure A3. Fitness level to Income**

# WarpPLS Analysis

# Diagram Description automatically generated

**Figure B1**. **SEM Analysis of XYZ Gym**

Table

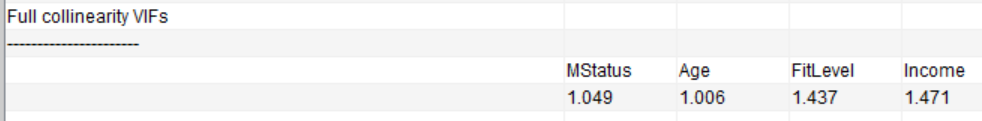
Description automatically generated

**Figure B2. Correlations among latent variables and errors**

Table

Description automatically generated

**Figure B3. Indicator Loadings and Cross Loadings**



**Figure B4. Full collinearity VIF’s**

Chart, scatter chart

Description automatically generated

**Figure B5. Relationship between pair of latent variables Income by Age**

Chart, line chart

Description automatically generated

**Figure B6. Relationship between pair of latent variables Income by Fit Level**

# Results Interpretation

The SEM analysis shown in Figure B1., shows FitLevel to Income has a positive beta coefficient thereby resulting in a direct relationship. It’s P value results as a less than .01 or less than 1% chance the correlation is due to chance. Its R2 value is at 29% chance which identifies the effect the independent variable (FitLevel) has on the dependent variable (Income). On the other hand, income does not seem to have a significant impact on marital status as due to a low beta coefficient and R2 value. Similarly, age appears to have little significance on Income as it too has a low beta coefficient, p value, and R2 value from this representation. The convergent validity observed here, is the data collection about the construct Fitness Level (FitLevel), in that it is a ‘self-rated’ observation of fitness, not based on a physiological test by a physician for example. With the convergent validity known, testing for discriminant validity may be conducted (4). To see good discriminant validity, the top table in Figure B2. should show the numbers on the diagonal to be larger than any numbers on the same column, which is the case shown in Figure B2. top table. Figure B3. Observed reveal the indicator loadings which should demonstrate a value greater than .5 that is a criteria for conversion validity and reveals problems with discriminant validity, the cross loadings (in the horizontal) should be lower than .5, which is the case for most, although FitLevel to Income have a value of .535, thereby pushing the discriminant validity threshold. Another test for discriminant validity is to observe the full collinearity variance inflation factors as shown in Figure B4. The numerical values should be below 3.3 to show proper discriminant validity (6), which is true for that figure shown. Figure B4., also demonstrates the inflation factor to be between 1 and 2, showing essentially no collinearity (2). Figure B5. an B6. support the numerical data, in that the relationship between income and age does not reveal a direct correlation but may act as a moderating variable to income. However, Figure B6. show the relationship between income and fitness level which appear to have a positive direct relationship, where better fitness may be an indicator for a higher household income.

# Practical Business use of analysis

This analysis of data could be of use to a business that wants to measure for a variety of sectors both business and government, such to gain a competitive advantage over another. This data may reveal itself useful to health care analytics, sports analytics, or marketing analytics as demonstrated in this analysis, all of which can be used as descriptive, predictive, and prescriptive analytics.

# Conclusion

In conclusion, it was considered that there was a correlation between higher fitness levels and higher income levels where age and marital status may have had a subsequent correlation. Through the various research conducted for this study utilizing several software applications to analyze the data, the predictor variable (fitness level) showed convergent validity to income level as predicted. The other latent variables age, and marital status revealed discriminant validity and acted as a moderator (age), and a mediator (marital status). No collinearity was shown in the data analyzed. Overall, the analysis provided statistical validation that marketing efforts should target the age range of the mid to late twenties, married, and financially stable in their future advertisements. Having this information would allow the marketing firm to speak directly to their primary customers thereby gaining the attention of their gym members and encouraging present and future attendance.

Resources

1. Kock, N. (2020, August). WarpPLS© User Manual: Version 7.0. https://scriptwarp.com/warppls/UserManual\_v\_7\_0.pdf.
2. Enders, F. B. (n.d.). *Collinearity*. Encyclopedia Britannica. https://www.britannica.com/topic/collinearity-statistics.
3. Escobar, M. R. (2020, February 19). *Structural equation modeling: What is a latent variable?* The Analysis Factor. Retrieved September 10, 2021, from https://www.theanalysisfactor.com/what-is-a-latent-variable/.
4. [scriptwarp]. (2020, Jan. 20). *Conduct a discriminant validity assessment with WarpPLS* [Video File]. Retrieved from https://youtu.be/OZQEEDN6jzM.
5. *Tsang, T. (2015). Mediating and Moderating Variables Explained. Psychology in Action. Retrieved from:*<https://www.psychologyinaction.org/psychology-in-action-1/2015/02/06/mediating-and-moderating-variables-explained>.
6. Shuttleworth, M. (2019, January 27). *What is convergent validity and discriminant validity?* ECWA USA. Retrieved September 19, 2021, from https://www.ecwausa.org/what-is-convergent-validity-and-discriminant-validity/.
7. Anand, S. (2021, August 1). *Cardiogoodfitness - descriptive statistics*. Kaggle. Retrieved October 1, 2021, from https://www.kaggle.com/saurav9786/cardiogoodfitness.