Harrisburg University of Science and Technology

Information Systems Engineering & Management (ISEM)

**ISEM 706 – Research Methods in ISEM**

**Spring 2022-2023**

**Midterm Examination**

**Due: March 12th, @midnight**

In this exam, you are expected to do multivariate data analysis by using the JMP and R. You will work on two datasets: HBAT\_200 and HATCO. The datasets and the documentation that gives information about the dataset are attached in this examination. Your exam report should include at least the followings:

* Show all your works in the report and add all graphs, charts, tables you used for the analysis.
* You can also print screen the windows you used in JMP to support your answers.
* You are also expected to explain all graphs, charts, tables that you added in your examination report.

**Submission notes:**

* Submit your **examination report (.doc)**. Show all your works in this report. Copy/paste the tables/graphs used in the analysis.
* Submit the **JMP files** including all your analysis. Do not forget to save the scripts to data table (use red triangle - Save script - To Data table).
* Submit **R files** you used to answer your questions.
* Due is March 12th @midnight.

**Question 1. *Examining the data*** *(30 pts)*

Examine the **HBAT-200 dataset** (attached)

1. Stage 1: Graphical representation analysis
   1. Univariate profiling: Examine the shape of the distributions (histograms) and get summary statistics (central tendency, dispersion, skewness, kurtosis, n missing, etc.) for each variable. *(5 pts)*
   2. Bivariate profiling: Examine the relationship between variables by using scatter plot. Consider that X19 is the target variable and X6-18 are independent variables. Include the correlation between the variables and interpret the results. *(5 pts)*
   3. Examine the group differences by using boxplot and check if outliers are present in the groups, and if the groups are different from one another. Please check the differences of X18-Delivery speed for each X1: Customer Type, X2: Industry type, X3: Firm size, X4: Region and X5: Distribution system. *(5 pts)*
2. Stage 2: Assessing assumptions including normality, homoscedasticity, linearity and noncorrelated errors. You are expected to do the following analysis:
   1. Normality: Do statistical test (Shapiro-Wilks) for normality for each variable X6-19. Add also normal plots. Interpret the results. *(5 pts)*
   2. Homoscedasticity: Do statistical tests (Levene test or Barlett’s test) to check whether variances are equal between the variables (X6-X18) and X1 (customer type) (Analyze – Fit Y by X). Interpret the results. *(5 pts)*
   3. Linearity and Non-correlated errors: Fit a model between the predictor variables X6-X18 and the response variable X19 (Analyze – Fit Y by X) Check for residuals by predicted plot and interpret results. *(5 pts)*

**Question 2:** ***Exploratory Factor Analysis (35 pts)***

Perform Exploratory Factor Analysis (EFA) using the Principal Component Analysis and Common Factor Analysis approaches on the HATCO dataset (attached).

You are expected to conduct a factor analysis for the HATCO dataset for the perceptions of HATCO on seven attributes (X1 to X7) using JMP and compile a data analysis report. Note that appropriate assumption checking, and validation steps need to be conducted to support your analysis results. Use the latent root criteria (eigen values>1) for deciding the number of factors to retain.

.

Below are a few questions to be addressed in your data analysis report within the relevant data analysis steps:

1. Upon inspecting the correlation matrix, how many correlations are observed to be significant at the 0.01 level? *(4 pts)*
2. Provide the partial correlation matrix for all variables under consideration. What is meant by partial correlations (in general)? What variables have higher partial correlations? *(4 pts)*
3. Use measure of sampling adequacy (MSA) to decide on the variables that need to be removed? (Use KMOS R file for MSA analysis or JMP) *(4 pts)*
4. For the remainder of the variables for the HATCO dataset, how many factors can be retained using the Scree test criteria? Also, how many factors can be retained using the latent root criterion? *(4 pts)*
5. Considering the latent root criterion, set the number of factors to be retained for the HATCO dataset, and derive the factors using principal component analysis. Provide the communality of the variables. What does the communality of a variable represent? *(5 pts)*
6. Perform an orthogonal rotation (preferably Varimax) and provide the factor loading matrix. Which variables load significantly on which factors? How much variance is explained by each factor? *(5 pts)*
7. Considering the latent root criterion, set the number of factors to be retained for the HATCO dataset, and derive the factors using common factor analysis. *(3 pts/ each = 9 pts)*
   1. Provide the communality of the variables. Interpret the results.
   2. Provide the significancy test results. Is the number of factors you used in the analysis significant? Interpret the test results.
   3. Provide the factor loading matrix. Which variables load significantly on which factors? How much variance is explained by each factor?

**Question 3: *Multiple regression and Logistic regression analysis (35 pts)***

Perform multiple regression analysis and logistic regression analysis on the **HATCO** **dataset**.

1. *(20 pts)* Conduct **multiple** **regression** **analysis** on the HATCO Dataset to predict the product usage levels (X9) of the customers based on their perceptions of HATCO’s performance (X1 to X7) using SAS JMP and compile a data analysis report. Use stepwise regression approach for your analysis.

Given below are a few questions to be addressed in your data analysis report within the relevant data analysis stages.

* 1. Provide scatterplots for individual independent variables against the dependent variable. Do any of them indicate any nonlinearity between the dependent variable and the independent variables? *(3 pts)*
  2. Explain what you understand by the term “heteroscedasticity”. Is it desirable or not? What can you say about the heteroscedasticity of each of the relationships (between the dependent variable and each independent variable) based on the scatter plots obtained in step a) above? *(3 pts)*
  3. Plot the normal probability plots for each of the seven independent variables. What do these plots indicate? Are data transformations for any of the variables required? *(3 pts)*
  4. Considering the original variables without any data transformations, estimate the regression model using stepwise regression and answer the following questions. Assume the p-value threshold for adding new variables to the regression equation to be 0.05, and the p-value threshold for dropping any existing variables from the regression equation to be 0.10. Provide the regression model generated based on this analysis. What can you say about the strength of relationships (in terms of R2) that can be detected with the significance level set at 0.05? What variables are significant in the model? *(10 pts)*
  5. What are the Variance Inflation Factor (VIF) for each of the independent variables in the final regression equation? What do they signify? Interpret the results. *(3 pts)*
  6. What percentage of total variation of X9 is explained by the variate? *(3 pts)*

1. *(15 pts)* Determine whether differences in the perceptions of HATCO’s performance (X1 to X7) exist between the small and large size firms (X8). So, you are expected to perform **logistic regression** on this dataset. For validation purpose, use 60% of the data for training set and 40% for validation set.

Given below are a few questions to be addressed in your data analysis report within the relevant data analysis stages.

* 1. Run the logistic regression model. Is the model significant? Which variables are significant in this model? *(10 pts)*
  2. How much variability of X8 is explained with this model? (3 pts)
  3. What does classification matrix (=confusion matrix = contingency matrix) show? Provide the accuracy of the model. (2 pts)

**Attached files:**

* HBAT\_200.xls
* HBAT documentation.pdf
* HATCO.xls
* HATCO documentation.pdf

**Good luck!...**