DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

EXAM ASSIGNMENT

Study Programme and level	**MSc Business Intelligence**						
Term	Summer 23o						
Course name and exam code(s)	Customer Analytics						460202E007
Exam form and duration	WOA: On-site written exam submitted digitally in WISEflow, use of the internet NOT allowed during the exam, own PC required.						4 hours
Date and time	6 June 2023						09.00-13.00
Supplementary material/aids	All	X	No		Specified		
Hand-in of hand-written material allowed	Yes		No	X			
Hand-in of extra material (appendix) in WISEflow allowed	Yes		No	X			
Anonymous exam	Yes	X	No		Comments: Please do **not** write your name or student ID number anywhere. Use your flow-id number (find this on the cover sheet).		
Other relevant information	**Avoid being suspected of exam cheating.** Remember to state references and use quotation marks, if you copy text from other sources or re-use parts of a previously submitted exam paper (plagiarism and self-plagiarism). Students must answer the exam assignment **individually**. All submitted exam papers are checked for plagiarism, so cheating and collaboration between students will be detected. A dataset is uploaded to WISEflow as appendix.						
Number of pages (incl. front page)	4 pages						

PART A. Lecturer: Ana Alina Tudoran (50 pts in total)

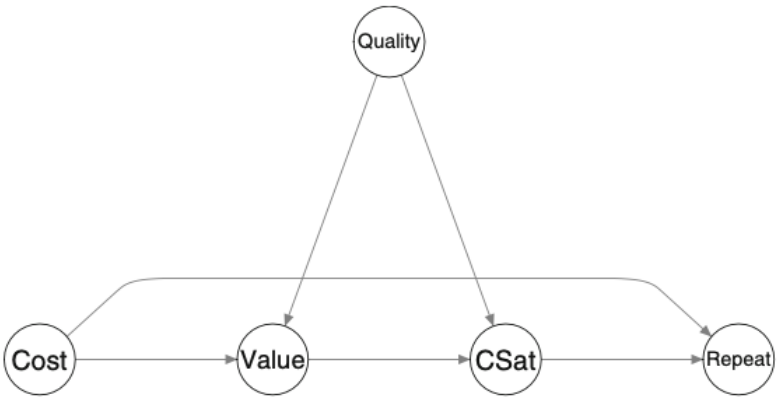
Exercise 1

Table 1 contains a survey of 15 statements based on which customers have provided their assessments of Hewlett-Packard printers. The structure of the survey is such that the first three statements are intended to tap quality, the next three measure cost perceptions, and then the last nine statements can be divided into three concerning customers' perceptions of value, three highlighting satisfaction, and three indicating behavioural intentions (adapted from Iacobucci (2009); Chapman and Feit (2015)).

Table 1
Customer quality survey.

	Strongly disagree						Strongly agree
1. The quality of the HP printer I bought is excellent.	1	2	3	4	5	6	7
2. HP printers are known to be highly reliable.	1	2	3	4	5	6	7
3. I'm sure my HP printer will last a long time.	1	2	3	4	5	6	7
4. The HP printer was reasonably priced.	1	2	3	4	5	6	7
5. HP sets fair prices for its products.	1	2	3	4	5	6	7
6. The HP printers are no more expensive than others.	1	2	3	4	5	6	7
7. I feel like I got good value for this purchase.	1	2	3	4	5	6	7
8. The quality of the printer is worth its cost.	1	2	3	4	5	6	7
9. I could tell my boss this purchase was good value.	1	2	3	4	5	6	7
10. I am very satisfied with my newly purchase HP printer.	1	2	3	4	5	6	7
11. My printer is better than I expected it would be.	1	2	3	4	5	6	7
12. I have no regrets about having bought this printer.	1	2	3	4	5	6	7
13. I would buy another HP if I had to buy another printer.	1	2	3	4	5	6	7
14. I would buy other HP products.	1	2	3	4	5	6	7
15. I would tell my friends and coworkers to buy HPs.	1	2	3	4	5	6	7

Below is a SEM model of customer repeat purchase intent based on this survey. In this model, the cost of a product is associated with both perception of value and intent to repurchase, while perception of quality relates to both perceived value and satisfaction, which is then associated with repurchase.



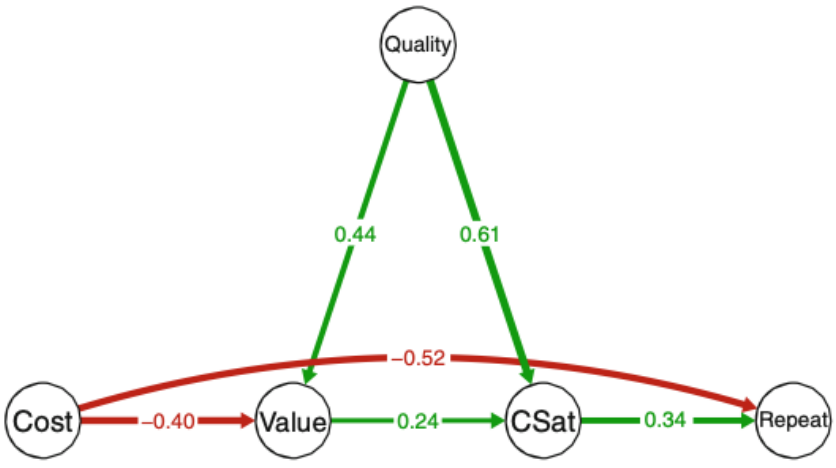
1) Considering the model path diagram above:
   a) Identify the exogenous, endogenous, and mediator variables. Identify a serial, diverging, and V-structure connection, and give an example of conditional independence and its practical implication for explaining customer repeat purchase. (10 pts)
   b) Discuss under which circumstances this model is causal. Justify your answer. (10 pts)

2) Fitting the model, we obtain the following results. Interpret the output below. (5 pts)

```
> sat.fit <- sem(satModel, data= satSimData, std.lv=TRUE)
> summary(sat.fit, fit.measures=TRUE)
lavaan (0.5-17) converged normally after  24 iterations

  Number of observations                           200

  Estimator                                         ML
  Minimum Function Test Statistic               85.454
  Degrees of freedom                                84
  P-value (Chi-square)                           0.435
...
User model versus baseline model:
  Comparative Fit Index (CFI)                    0.998
...
Root Mean Square Error of Approximation:
  RMSEA                                          0.009
  90 Percent Confidence Interval          0.000  0.040
  P-value RMSEA <= 0.05                          0.993

Standardized Root Mean Square Residual:
  SRMR                                           0.052
...
```

3) Based on a model with five latent variables and 15 observed variables, how can one determine if the sample size is sufficient for testing the model using SEM? (5 pts)

4) Based on the coefficient estimates displayed in the path diagram below (all are statistically significant at 5%, and the data was standardized before analysis), estimate the total effect of the cost on repeat purchase. (10 pts)



Exercise 2

Suppose you are conducting an A/B test to determine whether a new webpage design will lead to a higher conversion rate compared to the existing design. Last month, the website received 11,264 visitors out of which there were 780 conversions. The manager will only change the website for a minimum detectable effect of 20%. Assuming a two-tailed hypothesis test with a significance level of 0.05 and a statistical power of 0.8, calculate the total sample size and per group. [Note: you must show your calculations. A number without explicit calculations will not be assessed]. (10 pts)

PART B. Lecturer: Morten Berg Jensen (35 pts in total)

The influencer model is described on pages 70-73 in your curriculum textbook (Hair Jr et al. 2021[1]). An online survey was carried out in 2023, building on this model. Compared to the description of the data collection in the textbook, the respondents in the 2023 online survey only saw the "real" influencer and used a different scale (rate how much you agree with each statement on a scale from 0 = strongly disagree to 100 = strongly agree). However, those responsible for the online survey forgot to include the "sic_7" question when they developed the survey. Thus, only "sic_1" to "sic_6" are available (and they used "global_sic" as the variable name for the global single item for redundancy analysis). See Tables 3.9 and 3.10 in the textbook for an overview of the variables. To solve this assignment, you must use the dataset in the file called "*PLS_data_exam.csv".*

1) Discuss briefly how to treat missing values in PLS analyses in general. (5 pts)

2) Specify and run the influencer model. Next, evaluate the reflective measurement model. Be specific: report your implementation, relevant results, and comment on your output. (15 pts)

Assuming a segmentation of the respondents based on the answers to variables "sic_1" to "sic_6" using cluster analysis:

3) Discuss considerations regarding the use of the three types of cluster analyses: classical cluster analysis, model-based cluster analysis, and latent class analysis. (10 pts)

4) Discuss the suitability of this dataset for the segmentation analysis. [Hint: you may suggest the collection of additional information/variables]. (5 pts)

PART C. Lecturer: Surabhi Verma (15 pts in total)

1) A new online sports store has over 100,000 items of sportswear and sports equipment, but its rating database has only 1,000 ratings. Which of the following would be a better recommendation system?

   a) User-user collaborative filtering
   b) Item-item collaborative filtering
   c) User-item collaborative filtering
   d) Content-based recommendation

   Justify your answer. (7.5 pts)

2) In recommender systems, how can you balance the support and confidence to achieve accurate and relevant recommendations? (7.5 pts)

[1] Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook.