ISDS 415 – Logistic Regression Analysis Project: College Drop-out

<u>Due date</u>: check your course website for due date and time
This project contributes to 10% of the course grade.

Assignment Brief for <u>Postgraduate</u> Students

In this problem, you are working in the Student Support Office at Mont St. Michel, a local community college.

Over the past few years, the percentage of students who leave the college at the end of their first year has increased. Last year, the college started voluntary one-credit (1 hour long) seminars with faculty to help the first-year students to establish an on-campus connection. The college wants to know if it is able to show that the seminars have had a positive effect on the student retention. This could be used to convince the college administrators to continue funding this initiative.

One of the factors that the college administration believe is that first-year students with a lower high school GPAs have a higher probability of leaving the college at the end of their first year. The Student Support Administrator suspects that the high number of units that students are enrolled in in each semester contributes to an overwhelming schedule and discourages them from staying enrolled in the college.

A.   SAMPLE DATA
The college randomly selected 600 students from its last academic year and the data can be found in the file **MontStMichelSpring2022.xlsx**. Each observation consists of a first-year student's high school GPA, whether they enrolled in a seminar, how many units enrolled in a semester, whether they are full-time or part-time students and whether they dropped out and did not return to the college. Demographic information such as gender, age of the students and whether they live at least 20 miles off campus (commuter) were also collected.

B.   PROJECT REQUIREMENTS
You are to conduct appropriate statistical analyses that could be used to assist the college with the student drop-out problem
1.   Explore the given data by using descriptive statistics and charts, investigating any relationships in the data.
2.   Develop, evaluate predictive models based on their predictive performance on the given data set. Recommend a final model and express the model as a mathematical equation relating the output variable to the input variables. What is the implication of the effectiveness of the first-year seminars on retention?
3.   Perform a simple split of the dataset (e.g. 400 observations as training data and 200 observations as validation data) and evaluate your predictive models based on their predictive performance on both the training dataset and the validation dataset.
4.   For the next year's first-year students, what could the college administration do regarding the enrollment in the seminars to better determine whether they have an effect on retention.

Analysis Requirements for Questions 1-3:
Using 10% significant level test, your analysis must have the followings.
a) Descriptive statistics (with appropriate tables/charts) to describe your data.
b) Evaluation of the candidate logistic regression models based on their predictive performance on the probability that students will enroll in the seminar (use **all** given variable **except** the drop-out variable).
c) Evaluation of the candidate logistic regression models based on their predictive performance on the probability that students will drop-out of college after the first year.
d) In both parts b and c, does the students' demographic information play a role in their enrolment in the seminar and their dropping out?
e) Using the predictive **college drop-out model**:
   • estimate the probabilities for the number of units enrolled and their likelihood of dropping out. The number of units enrolled can be grouped as: **3** for 1-3 units, **6** for 4-6 units, **9** for 7-9 units, **12** for 10-12 units, **15** for 13-15 units, and **18** for 16-18 units and seminar as 1 and 0.
   • estimate the probabilities for the high school GPA and their likelihood of dropping out. The GPA can be grouped as suggested in table below (use the Grade Points column) and seminar as 1 and 0.

Grade	Grade Points	
A+, A	4	Outstanding
A-	3.7	
B+	3.3	
B	3	Good
B-	2.7	
C+	2.3	
C	2	Average
C-	1.7	
D+	1.3	
D	1	Poor
D-	0.7	
F	0	

If you believe there are other combination variable(s) from the developed model that play an important role in the classification of students drop-out, you can suggest it here.

The results should provide information to assist you in writing your report to the president of the college, Professor Callaghan.

Notes on Analysis Requirements:
   • Part a: you can use Excel, IBM SPSS or Tableau for this part.
   • Parts b, c, d: you must use IBM SPSS for your logistic regression analyses.
   • Part e: Excel 2-input data table must be used for this part.

5. From your analysis, you are to produce **one report** (1,800 - 2,000 words), addressed to Professor Callaghan, summarizing your evaluations and discussions of the analysis results.
   - Your report must include appropriate tables, charts, figures to support your claim. **Do not reference your claims to the output file(s)**.
   - Your report must have discussions and evaluations of the analyses from questions 1-3. Support your points with appropriate charts/tables. You can insert the tables/charts within text or in the Appendix. Do not refer the reader to your output file(s).
   - Your report must have conclusions and recommendations which are derived from your analyses from questions 1-3.
   - You may suggest any other factors (independent variables) that you think might make important contributions to the retention of the first-year students.
   - The report must be written in a narrative style. The use of bullet points should be minimal. Do not use bullet points throughout to answer the listed questions (points will be deducted if you do so).

   Submission: Using the project submission section on Canvas, submit **3 files**: (a) a copy of your SPSS output file (*.spv) which must include the appropriate data analyses you have carried out, (b) an Excel file that contain the input data tables that you created to estimate the probabilities **and** (c) your report in Microsoft Word.

C. ASSESSMENT MEASURES

Points	Assessments
30	**Correctness:** • Use appropriate statistical analysis • Discussions and evaluations of the statistical results • Interpret the results to make suitable business recommendations
20	**Format and flow:** • The report is well organized, coherent, and looks professional

Note: Be specific and give justification. Commonplaces and clichés should not be used. Give clear and well-illustrated reasons with supportive evidence/analysis results (use charts, graphs, tables and/or figures where necessary to support your evaluation). Arrange your argument/narrative/interpretation logically and coherently. I will evaluate both form and content of your report.

D.   PROJECT SCORE CALCULATION

The total possible point for the project is 75 (project = 50 points; peer evaluation = 25 points)

The following formula will be used to calculate your final project score:

(average score earned from the peer evaluations + original score earned from the project * (average score earned from the peer evaluations/total possible peer evaluation score)) / 75 total possible points * 10 percent of the course grade

For example:
If you earn: 20/25 for your peer evaluation and the project score is 40/50, the score you earn for your project would be: **(20 + 40 * (20/25))/75 * 10 = 6.93/10**

If you earn: 25/25 for your peer evaluation and the project score is 35/50, the score you earn for your project would be: **(25 + 35 * (25/25))/75 * 10 = 8.0/10**

Important Notes on Report:
Report correctness, flow, and format:
Your report should be flawless, that is, error-free and professional looking. Each report must have the following:
- A short introduction paragraph that gives the background to your study and a brief summary of what you have done in the report
- A discussion section covering the different analyses you have carried out and comments/evaluations on the results (you can have multiple paragraphs for this section)
- A conclusion section that wraps up and gives recommendations based on your analyses.

The report, including tables and figures/graphs, **should NOT be more than ten (10) pages** in length. Quantity is no indicator of quality, however. A short but well-focused report is preferable to a long and rambling one.

All pages must be numbered appropriately. Margins need to be 1 inch all around. Use Times New Roman font, size 12. If possible, use headings for better readability. Do not indent paragraphs. Keep them short (no more than 5-6 lines). Single space between lines but double space between paragraphs.

Hint: Plan your report; if necessary outline what you will say. Respond to the task fully. Before writing, analyze the context of the assignment by considering the requirements of the report and address the following questions:
- What is the purpose of the document to be written?
- What result do you hope to achieve by writing it?
- Who are your readers and what do they want from your document?
- What method of organization of your information is most useful?

Sample report
Below is a sample report layout/structure taken from Guffey and Loewy (2018).

RUIZ FINANCIAL SERVICES LLP
MEMORANDUM

Date: October 3, 2019
To: Damian Gorman, Vice President
From: Brenda Tchakerian, Director, Human Resources *B.T.*
Subject: Feasibility of Implementation of Internet Monitoring Program

As you suggested, we hired a consultant to investigate the feasibility of implementing a plan to monitor employee Internet use. The consultant reports that such a plan is workable and could be fully implemented by February 1. This report discusses the background, benefits, problems, costs, and time frame of the plan.

Background: Current Misuse of Internet Privileges. Currently we allow employees Internet access for job-related tasks. Many use social media—specifically, Facebook, Twitter, and LinkedIn—to communicate with clients and the public. However, some employees use this access for personal reasons, resulting in lowered productivity, higher costs, and a strain on the network. Therefore, we hired an outside consultant who suggested an Internet-monitoring program.

Benefits of Plan: Appropriate Use of Social Media and the Internet. The proposed plan calls for installing Internet-monitoring software such as NetGuard or eMonitor. We would fully disclose to employees that this software will be tracking their online activity. We would also teach employees what social media and Internet use is appropriate. In addition to increased productivity, lowered costs, and improved network performance, this software will also help protect our company against loss of intellectual property, trade secrets, and confidential information.

Employee Acceptance. One of the biggest problems will be convincing employees to accept this new policy without feeling that their privacy is being violated. However, proper training will help employees understand the appropriate use of social media and the Internet.

Costs. Implementing the monitoring plan involves two direct costs. The first is the initial software cost of $500 to $1,100, depending on the package we choose. The second cost involves employee training and trainer fees. However, the expenditures are within the project's budget.

Time Frame. Selecting the software package will take about two weeks. Preparing a training program will require another three weeks. Once the program is started, the breaking-in period will take at least three months. By February 1 the Internet-monitoring program will be fully functional resulting in increased productivity, decreased costs, lowered liability, and improved network performance.

Please let me know by October 15 whether you would like additional information about monitoring social media and Internet programs.

Alternative 2: Stop-Smoking Programs Outside the Workplace

Local clinics provide treatment programs in classes at their centers. Here in Houston we have the Smokers' Treatment Center, ACC Motivation Center, and New-Choice Program for Stopping Smoking. These behavior-modification stop-smoking programs are acknowledged to be more effective than literature distribution or incentive programs. However, studies of companies using off-workplace programs show that many employees fail to attend regularly and do not complete the programs.

Cost: $1,200 per employee, three-month individual program
(Your-Choice Program)
$900 per employee, three-month group session

Alternative 3: Stop-Smoking Programs at the Workplace

Many clinics offer workplace programs with counselors meeting employees in company conference rooms. These programs have the advantage of keeping a firm's employees together so that they develop a group spirit and exert pressure on each other to succeed. The most successful programs are on company premises and also on company time. Employees participating in such programs had a 72 percent greater success record than employees attending the same stop-smoking program at an outside clinic (Honda, 2017, p. 35). A disadvantage of this arrangement, of course, is lost work time—amounting to about two hours a week for three months.

Cost: $900 per employee, two hours per week of release time for three months

Conclusions and Recommendation

Smokers require discipline, counseling, and professional assistance to kick the nicotine habit, as explained at the American Cancer Society website ("Guide to Quitting Smoking," 2018). Workplace stop-smoking programs on company time are more effective than literature, incentives, and off-workplace programs. If our goal is to reduce health care costs and lead our employees to healthful lives, we should invest in a workplace stop-smoking program with release time for smokers. Although the program temporarily reduces productivity, we can expect to recapture that loss in lower health care premiums and healthier employees.

Therefore, I recommend that we begin a stop-smoking treatment program on company premises with two hours per week of release time for participants for three months.

References

Arhelger, Z. (2018, November 5). The end of smoking. *The World of Business*, pp. 3–8.

Guide to quitting smoking. (2018, October 17). Retrieved from the American Cancer Society http://www.cancer.org

Honda, E. M. (2017). Managing anti-smoking campaigns: The case for company programs." *Management Quarterly, 32*(2), 29–47. Retrieved from http://search.ebscohost.com/

Mendel, I. A. (2017). *The puff stops here.* Chicago: Science Publications, p. 108.

Rindfleisch, T. (2018, December 4). Smoke-free workplaces can help smokers quit, expert says." *Evening Chronicle*, p. 4.