PSYC*6380: Minor Assignment 5 (Categorical Outcomes)

Due: Monday, February 28th, 2022 @ 11:59pm

Part 1:

Background:

After hearing our example today, a consultant became interested in developing a presentation skills coaching program, with the idea that learning to be a better presenter would help people find jobs. After several months of operations, the consultant hired a program evaluator with a background in research psychology to examine whether the training program (called "*Hans' 'Present like a Pro!' Workshop*") was proving successful. Using a sample of 500 job-seekers, the program evaluator gathered data on: 1) each job-seeker's general presentation ability (measured on a scale from 1-10, with higher ratings indicating better presentation skills); and, 2) whether the job-seeker had enrolled in the "*Present like a Pro!*" program. At a three-month follow-up, the program evaluator asked each job-seeker whether or not they were now employed.

Your job is to load and prepare the data, and then look at the relations between: 1) enrollment in the program and job hunt success; and, 2) general presentation abilities and job hunt success.

Filename: "*catOutcomesAssignmentData.csv*"

Structure: Comma-separated values

Variables:

Participant ID Number (*ID*): 1-500

Employed at Follow-up (*Employed*): 0 = No; 1 = Yes

Enrollment in Presenter Training (*HadCoaching*): 1 = Not Coached; 2 = Coached

Expert-Rated Assessment of Presentation Ability (*PresentAbility*): 1-10 Likert-type scale

Missing Data Code(s): N/A (no values are missing in this file)

Your Task:

Use the techniques we covered in today's class (and draw on material from our first class on using *R* to load and clean data) to run a both a chi-square and logistic regression analysis, using recommended corrections for error where appropriate.

Once you've done this, please answer the following questions about the results. Unless otherwise indicated, you do not need to provide full-sentence answers; just the numbers the questions request are fine:

1. Run a chi-square analysis that plots a contingency table containing frequencies for: 1) enrollment in the program; and, 2) employment status at follow-up. Use whatever corrections you feel are most appropriate, given the structure of this contingency table. Please report: 1) the chi-square coefficient (i.e., χ^2); and, 2) the squared Cramer's *V* statistic (i.e., V^2) **(1 mark)**. Based on your interpretation of the V^2 statistic, how would you best characterize the predictive relation that's being demonstrated? Please be as exacting as possible. **(0.5 marks).**

2. To further explore the relation between enrollment in the program and employment status, please report the standardized residual values of employment for participants in each cell of the contingency table. Please report the value of each of these standardized residuals **(0.5 marks)**. Based on your interpretation of these values, how would you best characterize the difference in employment frequency between those who enrolled in the program, versus those who did not? Please be as exacting as possible. **(0.5 marks)**.

3. Run a logistic regression that predicts the likelihood of being employed based on general presentation abilities. Please report: 1) the exponential-transformed slope coefficient for "*PresentAbility*"; and, 2) the 95% confidence intervals around that exponential-transformed slope **(1 mark).** Based on your interpretation of the confidence intervals around the slope, how would you best characterize the predictive relation that's being demonstrated? Please be as exacting as possible when discussing both the slope value and its *CIs*. **(0.5 marks).**

4. Use the techniques we went over in class today to calculate a *pseudo-*R^2 statistic for the logistic regression you ran for q4, and report this value **(0.5 marks)**. Given what you know about how this value was calculated, how you best describe the predictive relation that's being demonstrated in this particular case? Please be as exacting as possible. **(0.5 marks).**

Part 2:

Background:

To help the program evaluator better understand how to analyze data with categorical outcomes (and better interpret the results of your analysis), please answer the following conceptual questions in a short answer (i.e., 1-2 sentences) format:

1. In our class today, Scott said that chi-squares are interpreted in a way that is basically the opposite of how regressions are interpreted. Specifically, he called them "*badness of fit"* tests. What did he mean by this? **(1 mark).**

2. Imagine you were given the following 2x2 contingency table of observed frequencies:

	Hipster	Not Hipster	Total
English Literature	**34**	**11**	**45**
Psychology	**15**	**40**	**55**
Total	**49**	**51**	**100**

   a. What is the expected number of hipsters taking English Literature? **(0.5 marks).**
   b. What is the expected number of hipsters taking Psychology? **(0.5 marks).**

3. Do we need special, non-parametric statistics to run a regression on data with categorical outcomes? Can't we just analyze these kind of data using the linear regression techniques we covered earlier this semester? Please explain why or why not. **(1 mark).**

Please provide your full *R* script with your submission and leave comments in your script (i.e., using "#") explaining what each command you wrote does. **(1 mark for including a full script; 1 mark for including appropriate commenting).**

Good Luck!